DEVELOPING A TWITTER BOT THAT CAN JOIN A DISCUSSION USING STATE-OF-THE-ART ARCHITECTURES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

YUSUF MÜCAHİT ÇETİNKAYA

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
COMPUTER ENGINEERING

AUGUST 2019

Approval of the thesis:

DEVELOPING A TWITTER BOT THAT CAN JOIN A DISCUSSION USING STATE-OF-THE-ART ARCHITECTURES

submitted by **YUSUF MÜCAHİT ÇETİNKAYA** in partial fulfillment of the requirements for the degree of **Master of Science in Computer Engineering Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Mehmet Halit Oğuztüzün
Head of Department, **Computer Engineering** _____

Prof. Dr. İsmail Hakkı Toroslu
Supervisor, **Computer Engineering, METU** _____

Examining Committee Members:

Prof. Dr. Tolga Can
Computer Engineering, METU _____

Prof. Dr. İsmail Hakkı Toroslu
Computer Engineering, METU _____

Prof. Dr. Ahmet Coşar
Computer Engineering, UTAA _____

Date:

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Yusuf Mücahit Çetinkaya

Signature :

ABSTRACT

DEVELOPING A TWITTER BOT THAT CAN JOIN A DISCUSSION USING STATE-OF-THE-ART ARCHITECTURES

Çetinkaya, Yusuf Mücahit
M.S., Department of Computer Engineering
Supervisor: Prof. Dr. İsmail Hakkı Toroslu

August 2019, 118 pages

Twitter is today mostly used for sharing and commenting about news [1]. In this manner, the interaction between Twitter users is inevitable. This interaction sometimes causes people to move daily debates to this social platform. Since being dominant in these debates is crucial, automation of this process becomes highly popular [2]. In this work, we aim to train a bot that classifies posted tweets according to their semantic and generates logical tweets about a popular discussion, namely gun debate of the U.S. for this study. Bots are trained to tweet independently on their side of the debate and also reply to a tweet from opposite view. State-of-art architectures are tested to get more accurate classification. We have applied GloVe embedding model for representing tweets. Instead of using handcrafted features, long-short-term memory neural network is applied to these embeddings to get more informative and equal size feature vectors. This model is trained to encode the tweet by fed as a sequence of embeddings. Encoding is used for both classification and generation tasks. LSTM sequence to sequence model is used to generate tweets and replies to tweets. The attention mechanism is added to the reply model to produce more related tweets. We

propose a new metric for measuring the relatedness of the reply to the target tweet. Additionally, human evaluators measure the quality of generated tweets according to relatedness to the topic and target tweet, which is replied.

Keywords: Natural Language Processing, Tweet Generation, Tweet Classification, Sentiment Analysis, Recurrent Neural Networks, Twitter Bot

ÖZ

SON TEKNOLOJİ MİMARİLER KULLANILARAK BİR TARTIŞMAYA KATILABİLEN TWİTTER BOTU GELİŞTİRİLMESİ

Çetinkaya, Yusuf Mücahit
Yüksek Lisans, Bilgisayar Mühendisliği Bölümü
Tez Yöneticisi: Prof. Dr. İsmail Hakkı Toroslu

Ağustos 2019 , 118 sayfa

Twitter bugün çoğunlukla haberleri paylaşmak ve yorum yapmak için kullanılmaktadır [1]. Bu minvalde, Twitter kullanıcıları arasındaki etkileşim kaçınılmazdır. Bu etkileşim bazen insanların günlük tartışmaları bu sosyal platforma taşımasına neden olmaktadır. Bu tartışmalarda baskın olmak çok önemli olduğundan, bu sürecin otomasyonu oldukça popüler hale gelmektedir [2]. Bu çalışmada, gönderilen tweet'leri anlamına göre sınıflandıran ve popüler bir tartışma hakkında, Birleşik Devletler silah tartışması, mantıklı tweet'ler üreten bir bot geliştirmeyi hedefliyoruz. Botlar, tartışmanın destekledikleri tarafında bağımsız bir şekilde tweet atmak ve aynı zamanda bir tweet'e karşıt bakış açısıyla cevap vermek için eğitildiler. Daha başarılı sınıflandırma sonuçları elde etmek için modern mimariler kullanılmıştır. Tweet'leri temsil etmek için GloVe kelime vektörleri modelini uyguladık. El yapımı özellikleri kullanmak yerine, daha çok bilgi içeren ve eşit boyutta özellik vektörleri elde etmek amacıyla uzun kısa vadeli hafıza sinir ağları bu kelime vektörleri ile beslenmiştir. Bu model, kelime vektörlerini bir dizi şeklinde girdi alarak tweet'in anlamını vektörlerin içine kodlanması için eğitilmiştir. Kodlama hem sınıflandırma hem de üretim işlemleri için kulla-

nılmıştır. Dizi-dizi UKVH sinir ağları, tweet üretmek ve tweet'lere cevap vermek için kullanılmıştır. Daha ilgili tweet'leri üretmek için cevaplama modeline dikkat mekanizması eklenmiştir. Yanıtın hedef tweet'e olan ilişkisini ölçmek için yeni bir metrik tanımlanmıştır. Ek olarak, insan değerlendiricileri, üretilen tweet'lerin kalitesini konuya ve hedef tweet'e olan alakasına göre ölçmüşlerdir.

Anahtar Kelimeler: Doğal Dil İşleme, Tweet Üretme, Tweet Sınıflandırma, Duygu Analizi, Yinelemeli Sinir Ağları, Twitter Botu

To my son, wife and whole family members...

ACKNOWLEDGMENTS

I am very pleased to work with Prof. Dr. İsmail Hakkı Toroslu during my graduate study. He was always positive and gave extreme guidance to me. It would not possible for me to finalize the thesis without benefit from his experiences. Besides, I would like to thank Assoc. Prof. Hasan Davulcu from Arizona State University for his support on this study.

I would also like to thank HAVELSAN Inc. for their support for my Master's degree and for the facilities that they allow me to use for this study.

I sincerely appreciate my friends Emre, Görkem, and Semih for their backing all along the stressful period. It was both fun and helpful to discuss with them about the study. I also value the insights and guidance Dr. Çağrı Toraman provides. Additionally, I am grateful to the friends, Abdullah, Ahmet, Battal, Berkan, Hatice, Mesut, Murat, Rabia for their various contributions.

Most importantly, I would like to thank my wife, my parents, and brother for their compassion and support. Being a father has motivated me a lot throughout the study. I am also grateful to my son, Ali Ömer, for making me have these inexpressible feelings.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ALGORITHMS

ALGORITHMS

LIST OF ABBREVIATIONS

AIML	Artifical Intelligence Mark-up Language
ALICE	Artificial Linguistic Internet Computer Entity
ANN	Artifical Neural Network
BLEU	Bilingual Evaluation Understudy
BoW	Bag of Words
CNN	Convolutional Neural Network
FC	Fully Connected
GloVe	Global Vectors for Word Representations
GOP	Grand Old Party
GPU	Graphics Processing Unit
kn-BLEU	k-nearest BLEU
LSTM	Long-short-term-memory
MC	Markov Chains
NLG	Natural Language Generation
NLP	Natural Language Processing
NLU	Natural Language Understanding
NRA	National Rifle Association
NUT	National Union of Teachers
POS	Part of Speech
RNN	Recurrent Neural Network
TT	Trending Topic
URL	Uniform Resource Locator

CHAPTER 1

INTRODUCTION

Twitter became very popular since it is founded in 2006 by its quick information retrieval property. Even it is not mainly designed for interaction between users [4], just like in other social media platforms, users are willing to comment on tweets that are posted by other users. This interaction causes users to not only share their ideas about shared information but also get into discussions. These discussions have a wide variety of topics, from football to politics, economics to education, etc. Being independent of the theme, any debate can be easily found on Twitter.

By the nature of the debate, people want to persuade the opponent for a specific topic. Social media debates are slightly different than the ones in real life. During this study, we had the chance to examine several tweets. The greater part of the tweets are posted to express the side without any assertion or to discredit the opponents and want them to find a solution to a problem. Instead of convincing the dissidents, they want to show their existence. This approach can work not by persuading but by conformity which is majority influence. Human beings choose the rational and consistent way but also they are prone to choose less effort to much effort. As Rook L. stated, people often choose the simple solution over complex solutions. In debates, the simple solution is to rely on the approach "the majority is always right" [5].

Politicians try to use social media effectively for closing ranks. They arrange campaigns with catchy slogans and want their supporters to share it in their social networks within coordination to seem like the majority. Since organizing this many people is not so easy, automating this process is now widely used [6][7]. The term used for such software is bot. It is a social media account that posts predefined or generated content automatically. A research reveals the activity of bot-driven accounts

during the Second U.S. Presidential Debate [7].

Conscious social media users can easily notice whether an account is a bot or not by checking over previous tweets or interaction with the users. To fool the users and have more impact on social media, bots can be smart and behave like a human. To mimic human attitude on Twitter, a bot should interact with other users. To do so, it has to post logical contents, answer to the questions towards it and, start new conversations. This is only possible if it can understand what other people saying. Otherwise, every word it posts would be a series of meaningless words.

The interaction between computers and humans requires a common language. This common language is achieved by user-interfaces for simple software. However, this approach is not much different than an analog system with an on/off button. Human languages are termed as natural languages in contrast to programmed or artificially developed languages. People expect smart computers to interact with them using natural languages. For decades, researchers have been working on this field called Natural Language Processing. There are several sub-fields of NLP such as natural language understanding, natural language generation, etc.

First studies in the NLP field consist of rule-based systems. Such systems are mainly known as "if-then" systems. All rules are given by a human expert to the system and it decides according to the given case. However, this approach is limited by the knowledge of the user. In addition, some intersecting cases are prone to be lost. Statistical approaches are later highly adopted by the researchers. Such systems determine the rules by statistical interpretation of given texts, instead of using human-crafted rules. It made essential improvements in the field. Support-Vector Machines, Decision Trees, Bayesian Networks, Markov Chains, etc. are popular models that are used for traditional statistical NLP systems.

With the increasing popularity of multi-layer perceptron, researchers applied artificial neural networks to many statistical problems. Various types of ANNs are developed and tried on numerous tasks. For instance, Convolutional Neural Networks has proved its success on the computer vision field and is now widely used [8][9]. Radial basis function Neural Network is another form which is generally applied to regression problems [10]. Similarly, Recurrent Neural Networks have a great capability

of representing sequences of data which is the main structure of the NLP field [11]. Since sentences are series of words, it can be thought of as a sequence. Nowadays, NLP researchers prefer varied types of RNNs.

1.1 Motivation and Problem Definition

In light of the aforementioned information, bots are used extensively in social media for various purposes. Hard-coded bots are programmed for posting predefined tweets from multiple accounts. In social media, posting the same post multiple times is referred to as spamming. It is easy to be noticed because it is not normal for multiple users to post the same tweet. Nevertheless, this type of bots is remarkably used since it is difficult to find original content for each bot. Moreover, there is a list, which is referred to as Trending Topic, that contains the most used keywords within a specific period and is pinned on left of Twitter page. Posting the same tweet from plenty of accounts helps the topic to be in TT.

Bots can still get the intended topic trending while generating unique content. Additionally, they get rid of the risk of being disclosed. The bot account which posts tweets with intelligence would have more influence in social media since it is indistinguishable, and therefore catch other users' attention more. Automatically making original tweets up requires the assistance of NLP.

Since the 1950s, there have been lots of bots that interact with humans. The general term used for them is "chatbot". Most of the models were designed with set of human-crafted rules [12][13][14]. They were trying to behave like a human and give the feeling of understanding what human said. However, there was no such software that can contextualize the sentences inside. Furthermore, they were not able to handle input that is not seen before. Researchers later discovered the notable success of statistical models in NLP and the systems started to discover the language rules by itself. Corresponding bots were more adaptive to unseen data in contrast to hard-coded ones.

With becoming neural networks widespread in machine learning, NLP researchers also have adopted this approach for modeling languages. Outstanding results in the

various tasks made the ANNs like a standard method of the field. It is strictly correlated with the advancement of the technology which reduced the computational cost. Today, deeper neural networks with lots of parameters can be modeled in ordinary computers. GPUs, which make the training process faster, became more accessible with the improvements in their architectures. Consequently, we have focused on deep learning models instead of rule-based or traditional statistical models.

This challenging topic made us excited to study on developing such a smart bot that can generate tweets like humans. We attempt to come up with a smart social media bot that is able to be part of a debate. In this study, we have worked on the tweets of discussions about U.S. gun control. As mentioned before, there are multiple NLP tasks to achieve the objective. In addition to generating tweets, the bot should be aware of other users' posts, catch the meanings and interact with them.

The desired pipeline of the system starts with text classification which is needed to determine the side of the tweet. It is essential for taking a position towards others. Later, a similar model is required for text generation. Differently, it should be trained with the tweets of one party of the debate since each model is targeting to behave like a supporter of one side. Along with these models, another one is needed for getting interacted with other users. Its job is to reveal the meaning and reply it according to the context. As a result, the system will use all models to act like a human and take place on Twitter.

1.2 Proposed Methods and Models

In this thesis, we use deep neural networks, namely RNNs, for addressing the issues mentioned above since they have remarkable success on sequential data. Long short-term memory is a specific type of RNN that is fitting for time series data. It has some special units to remember previous values over predefined time intervals to which makes it able to keep the context until the end [15]. It has variety of application areas, such as classification, generation, neural-machine-translation, etc. [16][17][18]

Since Twitter has no strict rule for texts, many tweets may be grammatically incorrect, contain misspelled words or multifarious smileys. They need to be projected into the

same plane to reduce the probable errors. We have pre-processed retrieved tweets with some rules, such as lowering letters, using tags instead of URLs, mentions, hashtags, multiple characters reduced up to 2, etc. It also reduced the vocabulary size conspicuously which is a good thing according to the curse of dimensionality.

It is necessary for computers to get symbols as numeric values. Since representing each word with a one-hot vector ends up with very huge sparse matrices, we used floating-point vectors with the size of predefined dimensions. This encoding is referred to as word embedding. There are different embedding models for representation of words like GloVe, word2vec, fastText, etc. [3][19][20] In the embedding space, each word has a place related to its context. Therefore, sentences store the meaning as close as they can.

For the first LSTM model that is responsible for predicting the side of the tweet, the input is fed by the tweet as a sequence embedding vectors. Each word is given one by one and the state of the cell is updated according to previous embeddings. At the end of the sequence, the cell encapsulates the representation of the whole tweet that is encoded by the network. It is then connected to the fully-connected layer which classifies the tweet. The model predicts the probabilities of each label that the tweet may belong to.

The tweet generation model has a marginal difference with label prediction. It predicts the next symbol with the given sequence of symbols instead of the label. The symbol may differ for varied tasks, however, we use the word embedding as a symbol. Using character as a symbol makes the model more prone to losing the context. Similar to the prediction task, we fed the network with a series of word embeddings with hiding the last one. The hidden embedding is the targeted output that the network should assert. This technique is called as next-word prediction or language modeling. The model learns the fundamental properties of the language. The model gets a prime text for completing the rest. It may be a single word or multiple.

Only next-word prediction is not enough for getting interacted with other users. The bot should give responses to other users. The responder model should consume the whole text to get the meaning and generate a reply. This is similar to the NLP task called neural machine translation. In such models, LSTM consumes the sequence and

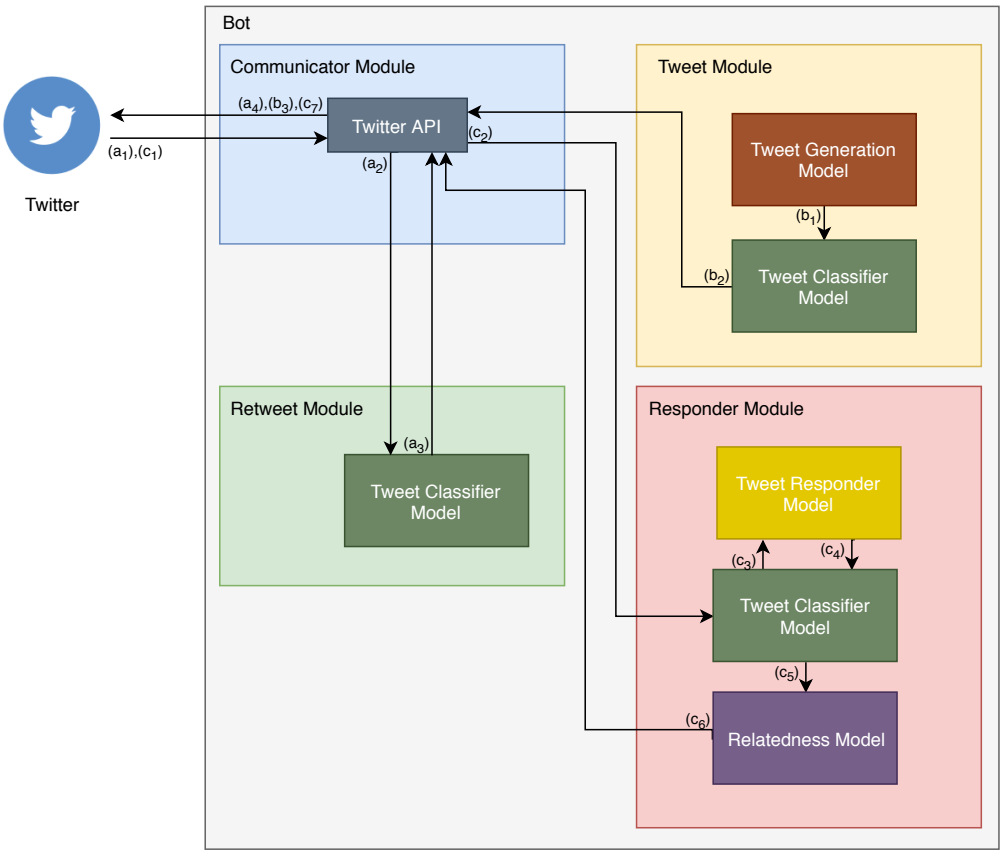

Figure 1.1: Proposed model for the bot that can join a discussion on Twitter.

produces the representation as we do in the tweet classification. This representation is used to bring the replies out. More clearly, the model consists of two separate parts; encoder and decoder. The encoder is analogous to the recurrent part of the classification task. However, instead of connecting the cell to the fully-connected layer, it is connected to the decoder. The decoder is comparable with the one in the generation model. It gets the encoding of the source tweet and generates a response for it. This technique had a good run in automated translations [21].

Generated content needs to be checked concerning the semantic and the relatedness. If the semantic of the bot-produced tweet does not reflect the view correctly, it is not posted. Similarly, the generated reply to the target tweet is required to exceed the threshold of relatedness metric. Filtration is crucial for generation based models since it is possible to produce meaningless or unrelated text.

Figure 1.1 visualizes the flow of how the bot carries out the fundamental operations

on Twitter. The communicator module collects tweets related to the discussion with Twitter API (a_1). The retweet module gets these tweets to classify to retweet (a_2). If the tweet reflects the view of the bot with a certain probability, it is retweeted (a_3)(a_4). The tweet module contains the LSTM model for tweet generation. It generates tweets and sends it to the classifier (b_1). Produced tweet needs to reach the threshold for the semantic. If it exceeds the threshold it is posted (b_2)(b_3). Responder module is responsible for replying the opposite view tweets. The communicator module collects tweet with some specific keywords (c_1) and directs it to the responder module (c_2). If the tweet is dissident to the view of the bot, the LSTM model is fed with the tweet (c_3). The generated reply is checked according to semantic and relatedness respectively (c_4)(c_5). If it passes both filtration mechanisms, it is posted through the communicator module (c_6)(c_7).

1.3 Contributions and Novelties

We propose an end-to-end framework for a Twitter bot to support an idea belonging to discussions. In the Twitter environment, users are capable of posting, replying tweets, mentioning someone else and retweeting other tweets. Our method offers multiple models to work harmoniously to perform these tasks.

In order to select which tweets should be replied, we need a classifier that can predict the side of the tweet in the debate. Using classifier as a scoring mechanism for generated tweets is a new approach to our knowledge. Additionally, the classifier model is very helpful in narrowing the Twitter space during determining which tweets should be replied.

Furthermore, we come up with a new metric that measures the relatedness of the reply generated to the target tweet. This metric can be used on various models to measure the quality of the model on generating relevant content. In our bot components, it is the main part of filtration while generating replies. If the produced reply relatedness score is below the threshold, it is simply ignored and not emitted.

The bot should not repeat the sentences with the same words. It should at least have the ability of paraphrasing. In this manner, we used drop-out also while generating

tweets in our replier model. It produces different sentences within the same context.

1.4 The Outline of the Thesis

In Chapter 2, we provide the information about the chatbots from the very beginning of the literature to nowadays. In addition, several approaches for classifying documents and generating texts are discussed. The supremacy of state-of-the-art architectures over traditional NLP methodologies are explained.

Chapter 3 supplies the formal definitions of the methods that we refer to in this study. Word representation methods, artificial neural networks, and evaluation criteria of generated texts are included. Moreover, the dataset and the pre-processing procedure we applied are explained in detail.

We explain the classification problem of our bot in the theme of the dataset and propose a model to solve it in Chapter 4. The architecture of our classifier is visualized and the performance results are included in the Experiments section.

Chapter 5 provides the details of the procedure we follow while modeling the language and generating tweets. The criteria of which tweets should be posted among all generated tweets are also pointed out. Perplexity values of machine-generated tweets are compared with the human posted ones. Additionally, human evaluators have scored both human and machine-generated tweets.

In Chapter 6, we describe how we design the model to generate replies to target tweets. We provide the details of the new metric to measure the relatedness in this chapter. Generated and real replies are compared in the context of relatedness and quality. The relatedness is analyzed with both human evaluators and kn-BLEU.

Chapter 7 includes the conclusion and the possible improvements on this study. Considered enhancements are listed under the Future Work section.

In the appendices chapter, we attached the multiple-choice and quality tests that are assessed by human-evaluators.

CHAPTER 2

RELATED WORK

There have been various researches on developing software that can mimic human beings in conversations for a long while. In this chapter, we try to cover some of them from the primitive techniques to the state-of-the-art ones. They use different approaches for meeting different requirements. The evolution of chatbots has been accelerated with the advances in technology. The improvements in the storage and computational capabilities give us the ability to keep and working on very large corpora.

The main issue for a chatbot is extracting the meaning of human input which is a very challenging task. Different methodologies have been developed and tested to pass the Turing test, which is a test proposed by Alan Turing to distinguish if the responder is machine or not [22]. Early developments include predefined outputs that fit best to the human input. However, it is bounded by the set of given sentences. Therefore, database retrieval, statistical, and hybrid systems are developed.

There is an annual competition for chatbots that is started by Hugh Loebner in 1991 [23]. He offers $ 100,000 for the first computer program that can pass the Turing test. Up to the present, any of the participants could not pass it but the most human-like bot among them is selected as the winner for that year. Examining the Loebner Prize year by year displays the progression in the domain.

Researches in this domain start in the 1960s with a project called ELIZA which is a text-based chatbot developed in MIT AI Laboratory [12]. It imitates a Rogerian psychotherapist by specifying rules called pattern matching. The model does not include any component for capturing the context. It looks for specific keywords in the input

and gives weights, called rank, to them. Independent from the context, it produces the same output if the keyword exists in the sentence. In 1972, PARRY is released at Stanford University which is a text-based chatbot [13]. It also contains handcrafted rules for generating responses. It behaves like a person with paranoid schizophrenia. Until the mid-90s, these hard-coded chatbots were popular. PC Therapist, Professor, and Politician are the other bots that use similar pattern matching and text parsing in their algorithms. [24].

Adding database retrieval helps bots to not repeat themselves, which is the big portion of the problems in the earlier systems. After that point, statistical models became popular. Markov chain models are used for language modeling. It extracts the probability of occurrences with the given sequence. These probabilities do not depend on the property of the tokens but only the token itself. HeX is the winner of the 1996 Loebner Prize competition which uses MC in addition to pattern matching [24]. Semantic nets made the bots able to get the interconnected context related to the given input. Together with semantic nets, CONVERSE won the Loebner Prize competition in 1997 [25]. This way, CONVERSE could imply new statements by reasoning based on its knowledge and sometimes lead the conversation.

An open-source project AIML, Artifical Intelligence Mark-up Language, is developed by Richard Wallace and community after 1995. It is a standardized way of programming the bot with predefined outputs. It gives the programmer the ability to add more patterns in a generic way easily. Nevertheless, A.L.I.C.E got the first place in the 2000 Loebner Prize [26]. It searches for the best pattern it matches and returns the hard-coded output.

AIML has some weaknesses on pattern matching algorithms and ChatScript is developed as a successor. It has a hierarchical structure for determining the best topic that matches user requests. ChatScript is case-sensitive in contrast to AIML. Suzette is the first chatbot that is developed with ChatScript engine [27].

Due to advancements in deep learning, researches on generation-based approaches are accelerated. Recurrent Neural Networks, especially Long-Short Term Memory RNNs, are used in this approach [18]. There are two components in these architectures; encoder and decoder. This model is generally used for neural machine trans-

lation tasks [28]. However, it can be easily applied for generation-based chatbots where source sentences can be considered as user requests and the translation as the response.

In the past decade, big companies have made investments on intelligent personal assistants (IPAs), such as Apple Siri, Microsoft's Cortana, Google Assistant, Facebook M, and Amazon's Alexa [29]. They can be considered as a type of chatbot. They understand what user requests and response appropriate answers accordingly. Additionally, IPAs proactively anticipate user needs and provide in-time assistance such as reminding of an upcoming event or recommending a useful service without receiving explicit requests from the user[30].

Today's social media bots are developed not only respond to the users' requests but also to establish an emotional connection with them. Microsoft's chatbot XiaoIce has been designed as a 19-year-old female persona, with strong language ability, visual awareness, and over 180 skills [29]. Microsoft released Tay in 2016. It behaves like an 18 - 24-year-old American woman and has conversations over Twitter. XiaoIce and Tay have the ability to expand their knowledge during the conversations. This skill made Tay learn offensive language and content that causes Microsoft to take the chatbot down [31].

CHAPTER 3

PRELIMINARIES

In this thesis, we have focused on both text classification and text generation tasks. Since the theme of this study mainly depends on NLP and machine learning, this chapter includes the fundamental overview of the concepts, such as feature representation, the label of a tweet, training process, etc. We used deep neural networks rather than building hand-crafted features. Therefore, the basic architecture of RNNs is explained. In addition, we describe the details of the data set to complete the content here.

3.1 Word Embeddings

In order to symbolize documents for the computer to differentiate, we need to represent words in a logical way. Storing the relation between words is crucial for NLP tasks since the context of a sentence strictly depends on it. Besides, dense vectors are the better choice for machine learning tasks in terms of computation and storage feasibility [32]. Word embedding is a highly popular document representation technique that meets the mentioned requirements.

To realize the significance, first, demonstrate two primitive document representation approaches, namely the sequence of one-hot vectors and bag-of-words. For the first one, documents are encoded by multiple vectors. Every vector has the length of vocabulary size and a single value is one where the others are zero. Obviously, each document has a very sparse matrix representation. On the other hand, in the BoW technique, even vectors have higher density, they are still sparse. In this method, each document is represented by a single vector where each index is matched with a word

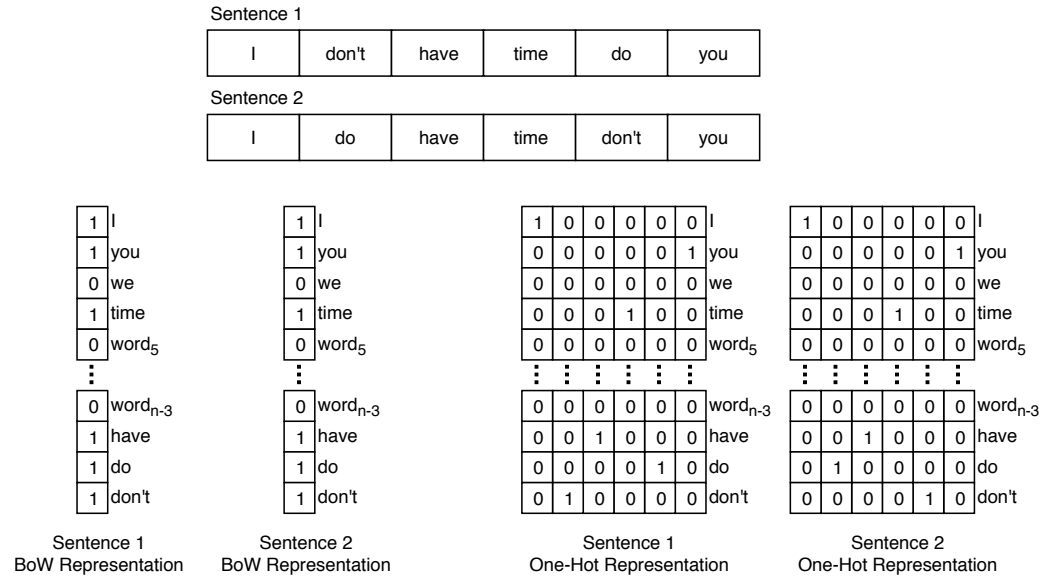

Figure 3.1: The representations of documents in primitive techniques.

in the vocabulary. Moreover, there is a lack of capability to keep the meaning. For example, let's consider the following sentences; "I don't have time, do you?" and "I do have time, don't you?". Both documents have the same BoW representations while having totally opposite meanings.

In addition to the sparsity, both approaches ignore the interrelationship between words. Models that use them can not associate the words according to their semantics. In those techniques, words are represented in a discrete space. Therefore, it is not able to find a relationship between the words, such as "parent" and "mother". Noticeably, modeling a language with disregarding the tie between the phrases is unimaginable.

Word embeddings are computed in various ways using unsupervised learning techniques. They correlate the words statistically using corpus. In other words, similar words are more likely to have close embedding vectors. The dimension of the representative vectors can be defined in any size. We will cover GloVe in this chapter.

3.1.1 GloVe

GloVe, global vectors for word representation, is an unsupervised model for representing the words using co-occurrence probabilities retrieved from large corpora.

Probability and Ratio	$k = solid$	$k = gas$	$k = water$	$k = fashion$
$P(k\|ice)$	1.9×10^{-4}	6.6×10^{-5}	3.0×10^{-3}	1.7×10^{-5}
$P(k\|steam)$	2.2×10^{-5}	7.8×10^{-4}	2.2×10^{-3}	1.8×10^{-5}
$P(k\|ice)/P(k\|steam)$	8.9	8.5×10^{-2}	1.36	0.96

Table 3.1: Co-occurrence probabilities extracted from a large corpus containing 6 billion tokens retrieved from GloVe paper [3]. Ratio of probabilities of ice and steam with non-discriminative words like *water* and *fashion* are close to 1 since both are unrelated. Context word *solid* is more relevant with *ice* than *steam*. It is vice versa for *gas*.

The statistics of word occurrences in a corpus is the primary source of information available to all unsupervised methods for learning word representations [3]. The inter-relationship between words is discovered using this co-occurrence data. Pennington et al. focused on generating the meaning using this information and representing it with a vector.

To explain how co-occurrence information is useful, the paper starts with a simple example. It clarifies how to extract meaning directly from these word-word co-occurrence probabilities. The relatedness between words, i and j, are inspected by examining the ratio of their co-occurrence probabilities with some context words, k. The notation of the co-occurrence matrix is X where X_{ij} stands for the entry that implies the number of j occurs in the context of i. The probability of i to occur with the probe word k is denoted by $P(k|i)$. The expectation is that P_{ik}/P_{jk} should be large for contexts relevant to i but not j. If the situation is the opposite, it should be small. For the cases, if both are related or unrelated, the ratio should be close to one. Table 3.1 demonstrates it with the concept of thermodynamic phase. Probe word *solid* is more likely to co-occur with the word *ice* than it does with *steam*. In contrast, for the *gas* context word, *steam* has more probability. Since the word *water* is related and *fashions* is unrelated with both, the ratio is close to one.

Since the ratio gives more information about relatedness, the model depends on it instead of probabilities themselves. The research indicates that depending on the vocabulary and the corpus zero entries consist of 75-95% of the data. Therefore,

the proposed model uses a weighted least-squares regression for getting rid of co-occurrences that happen rarely or never.

$$\sum_{i,j=1}^{V} f(X_{ij})(w_i^T \tilde{w}_j + b_i + \tilde{b_j} - logX_{ij})^2 \tag{3.1}$$

The model is designed with a linear structure to keep the linear relationship between words. The total number of occurrences of a word i in the corpus can be calculated from the X matrix by summing entries with all probe words up, which is $X_i = \sum_{\forall k \in V} X_{ik}$. Desired word embedding and the probe word embedding are denoted as w and $\tilde{w}$. The weighting function $f(X_{ij})$ has the following properties for the integrity:1) $f(0) = 0$, 2) $f(x)$ should be non-decreasing so that rare co-occurrences are not over-weighted, 3) $f(x)$ should be relatively small for large values of x, so that frequent co-occurrences are not over-weighted. The chosen function that obeys these rules is;

$$f(x) = \begin{cases} (x/x_{\text{max}})^{\alpha} & \text{if } x < x_{\text{max}} \\ 1 & \text{otherwise .} \end{cases} \tag{3.2}$$

x_{max} is used as a cutoff variable. It is fixed to 100 for all experiments in the study since the model depends weakly on it. For the α, they have found that 3/4 value gives a modest improvement over a linear version with α = 1.

In consideration of the linear structure of the model, the closest words can be found with linear distance metrics, such as Euclidean distance. There are pre-trained models shared on the study web-page. The one we have used in this thesis is trained with Twitter data containing 2 billion tweets with 1.2 million unique words of 27 billion tokens. Table 3.2 shows the closest 7 words to the word *rifle* in that model.

	Closest Embeddings	Distance
1	rifles	5.12
2	airsoft	5.25
3	handgun	5.64
4	sniper	5.65
5	paintball	5.71
6	assault	5.77
7	shotgun	5.90

Table 3.2: Closest words to the word *rifle* according to Euclidean distance in the 200-dimensional vector space which is trained with Twitter data.

3.2 Artificial Neural Networks

Inspiring from biological neural networks, artificial neural networks are the architectures for solving various statistical problems, such as classification, regression, clustering, etc. The model is based on a processing unit called perceptron which is invented by Frank Rosenblatt in 1958 [33]. Similar to biological neural cells, a perceptron collects multiple input and weights each. It emits the summation of weighted inputs as a signal. $o = \sum_i^N w_i x_i$. To add non-linearity capabilities to the cell, an activation function might be applied to the summation. Figure 3.3 visualizes the operation of a perceptron. A single neuron stands for a single function that takes multiple inputs and produces one output. Therefore, it might not be able to solve some problems. This processing unit can be combined in several ways to deal with more complex problems.

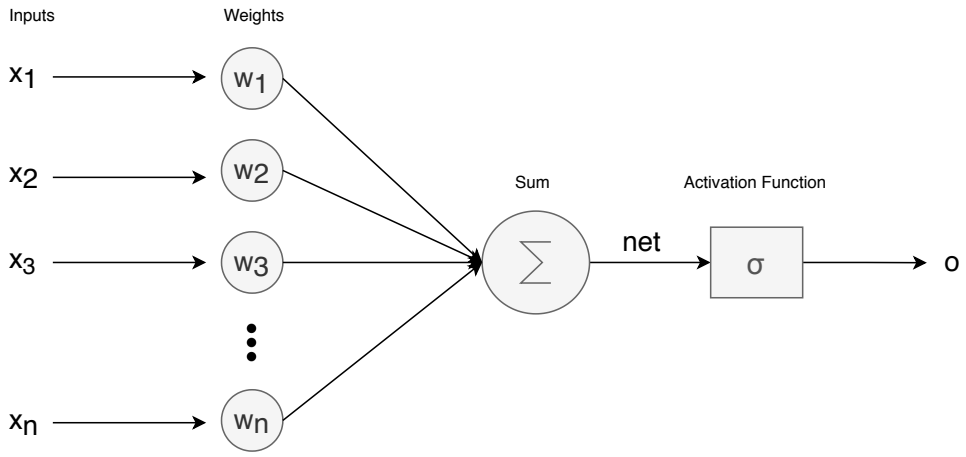

Figure 3.2: A single neuron. It takes n inputs and weights each. The summation of these weighted inputs is sent to the non-linear function.

Multilayer Perceptron is a popular type that consists of multiple perceptrons that are stacked on each other. It has been used for supervised learning. The architecture may vary with number of layers, number of neurons, activation function applied to the neurons, etc. Layers are called with special names. Namely, the first layer and the last layer are called input and output layer respectively. Other layers between these two are called hidden layers. If there is more than one layer, the architecture is referred to as a deep neural network.

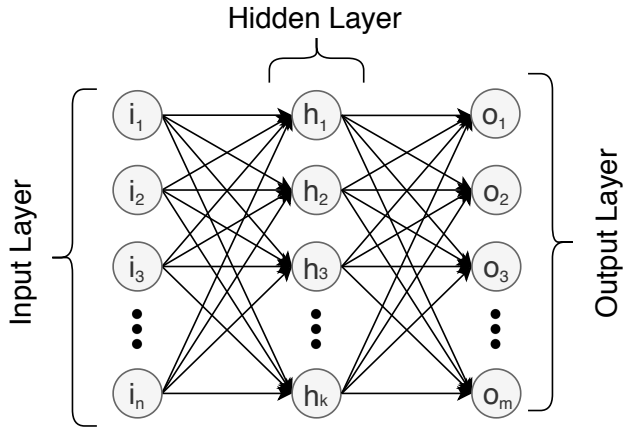

Figure 3.3: A MLP with one hidden layer, n inputs and m outputs.

Although the architectures of ANNs differ, the main structure of the training is the

same. It consists of two flows: forward and backward. The first one is responsible for prediction. It takes the inputs over the input layer and weights emits the signals to the next layer up to the output layer. Forward propagation overs after the last layer produces the outputs. Backward propagation starts with the calculation of error rate by comparing the prediction and the targeted outputs. The weights inside the model are adjusted with respect to the error with the back-propagating algorithm [34]. Forward and backward propagation continues until the model converges to a point or for several epochs.

There have been several architectures that are derived from the perceptron model. Recurrent neural networks, radial basis function networks, convolutional neural networks, autoencoders, etc. are some of them and specialized for solving different problems. For instance, convolutional neural networks work very well on spatial data like computer vision problems [9]. Autoencoders are unsupervised learning models that are to produce the output as same with the input to ignore the noise in the input and get a good representation of the data [35]. Similarly, recurrent neural networks are good at modeling the sequences and widely used in NLP nowadays [36].

3.2.1 Long-Short Term Memory Network

Traditional RNNs have some problems with training while optimizing the network. In order to deal with exploding and vanishing gradient problems, Sepp Hochreiter and Jürgen Schmidhuber proposed a new model called long-short term memory network in 1997 [15]. It is successful with sequential data such as video, speech, language models, etc. In addition, LSTM is able to bridge very long time lags [37].

Neurons inside the LSTM model is called memory cells. The architecture of LSTM units can vary but the proposed model in the papers [37] consists of a cell and gates. There is a linear unit with a fixed-weight self-connection to prevent the gradient from exploding or vanishing during the flow within the cell. The gates are responsible for the flow of information. The gates have different aims. The input gate takes care of the perturbation of the error flow within the memory cell from by irrelevant inputs. Output gate is there for protecting other units from perturbation by currently irrelevant memory contents, or states, kept in the memory cell. Forget gate is proposed by Felix

Gers et. al. in 1999 which is in charge of preventing the cell to be saturated [38]. It learns to reset memory state once their contents are out of date and hence useless.

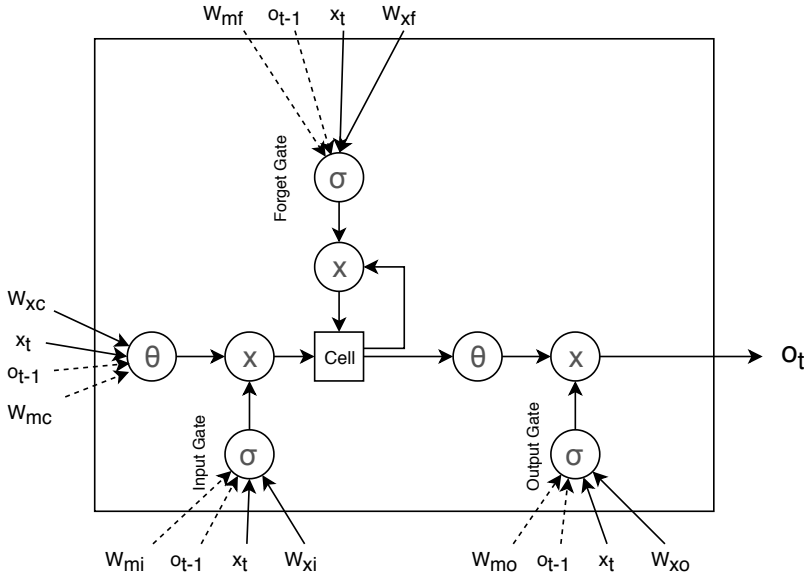

Figure 3.4: LSTM unit with three gates; input gate, output gate, and forget gate.

Denoting the connection from unit v to unit u as w_{uv}, net input and the value of u after activation at time t are $net_u(t)$, $y^u(t)$ respectively. Namely, $y^u(t) = f_u(net_u(t))$ where f_u is the activation function and $net_u(t) = \sum_v w_{uv} y^v(t-1)$. Each memory cell, c_j, is structured over a central linear unit with a fixed-weight self-connection and identity function as activation function. In addition to activated value from timestamp $t-1$, c_j gets input from special units input and output gates, in_j and out_j. Similarly, they are denoted as $y^{in_j} = f_{in_j}(net_{in_j}(t))$ and $y^{out_j} = f_{out_j}(net_{out_j}(t))$ where net values are defined recursively as $net_{in_j}(t) = \sum_u w_{in_j u} y^u(t-1)$ and $net_{out_j}(t) = \sum_u w_{out_j u} y^u(t-1)$. The summation indices u may refer input units, gate units, memory cells, or hidden units. At time t, the output of c_j is computed as $y^{c_j}(t) = y^{out_j}(t) h(s_{c_j}(t))$. The internal state of the cell,s_{c_j} , can be defined recursively as;

$$s_{c_j}(0) = 0, s_{c_j}(t) = s_{c_j}(t-1) + y^{in_j}(t) g(net_{c_j}(t)) \text{ for } t > 0 \tag{3.3}$$

The differentiable functions g and h scale net_{c_j} and memory cell outputs respectively. The weights are adjusted with the back-propagation algorithm similar to MLPs.

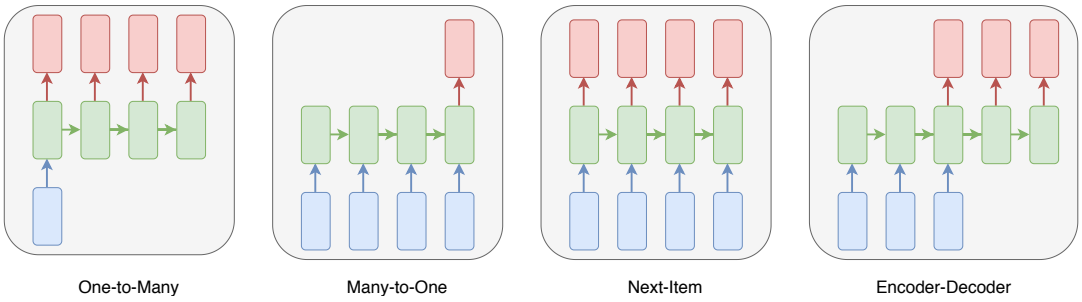

Figure 3.5: Different RNN configurations for solving different problems.

LSTM units can be arranged with various configurations to solve different problems. Sample problems for each configuration that is visualized in Figure 3.5 are as follows;

- **One-to-Many :** The model does not consume a sequence but produces a sequence. Generating memes using deep neural networks study uses this approach [39]. It takes an image as input and generates a sentence.
- **Many-to-One :** The model is fed by a sequence input and returns a single output. Wang et. al. feeds the model with sentences and try to predict the sentiment in the study [40].
- **Next-Item :** The model produces an output synchronously while getting input. Language modeling with LSTM uses this approach [41].
- **Encoder-Decoder :** The model consists of two LSTM parts; encoder and decoder. In Neural Machine Translation (NMT), the model takes a sequence of words from a language and produces a sequence from another language [18].

3.3 Beam Search

While searching in a large space, memory requirements might limit the algorithm. Therefore some heuristics could be used to reduce the memory requirements. Beam search is a heuristic search algorithm that expands the most promising n branches. It does not guarantee the best solution in the space but the local optimum. Therefore, it is a type of a greedy algorithm. It is an enhancement on the best-first search.

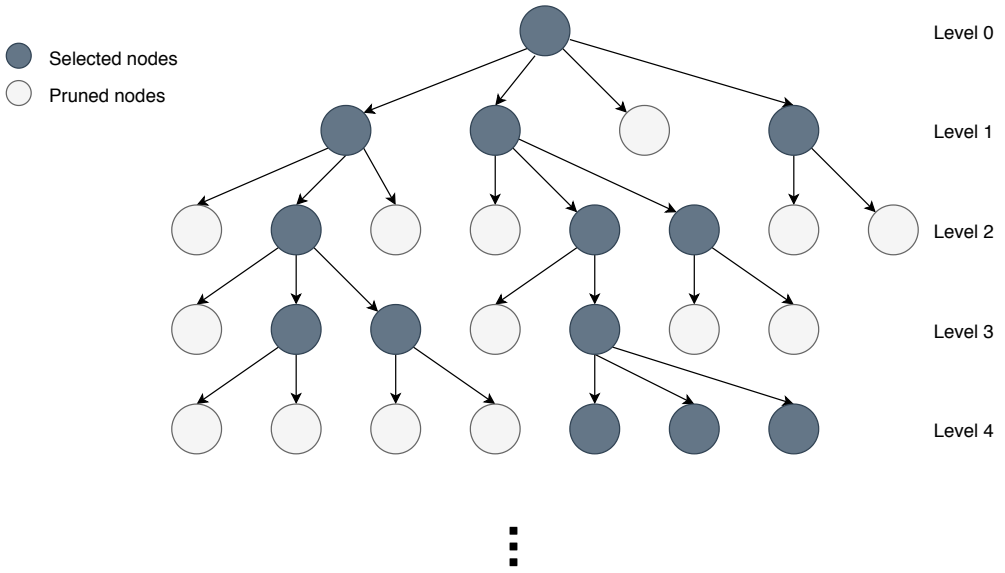

Figure 3.6: Beam search on a tree with a beamwidth of 3.

The algorithm applies the breadth-first search to find the best node in the tree. However, instead of expanding the tree up to the leaves, at each level it sorts the nodes according to the score. Only the top n nodes are expanded. This hyper-parameter is called beam-width. If the beam-width is selected as infinite, at each node is expanded in each level and it becomes the breadth-first search. It is first used in Harpy Speech Recognition System [42].

The term *best* may refer different definitions for different tasks. For instance, in language modeling, choosing always the most promising next word may end up with a bad sentence. Selecting the less probable word at a timestamp may form better sentences. Namely, if we select always the most promising next word, the sentence becomes $P(S_1) = P(x_1^1) \cdot P(x_2^1|x_1^1) \cdot P(x_3^1|x_1^1, x_2^1) \cdot \ldots P(x_m^1|x_1^1, x_2^1, x_3^1, \ldots, x_{m-1}^1)\cdot$, where x_u^v denotes the u^{th} word in the sentence with length m and v^{th} most probable word for that place. Using beam search, produced sentence may be $P(S_{best}) = P(x_1^i) \cdot P(x_2^j|x_1^i) \cdot P(x_3^k|x_1^i, x_2^j) \cdot \ldots P(x_m^t|x_1^i, x_2^j, x_3^k, \ldots, x_{m-1}^l)\cdot$ where each word may be coming from any node in the top n branches. Since $P(S_1)$ is identical to the beam search with beam width 1, $n = 1$, it is guaranteed that $P(S_1) \leq P(S_{best})$.

3.4 Quality Evaluation for Generated Texts

Natural Language Generation models produce texts with different techniques. In order to compare these models, we need some metrics. The quality of generated texts can be examined in various textual features. It may refer to grammatically correctness, being meaningful, fluency, easy to read, etc. Some of these evaluations are automated and some are subject-based human evaluation. There are numerous metrics to measure different properties of the text. Generated texts might be produced for several tasks. Namely, it can be a translation, text-summarization, question-answering, language modeling, etc. For instance, BLEU (Bilingual Evaluation Understudy) is the most popular metric for evaluating machine-translated texts [43] and ROUGE (Recall Oriented Understudy for Gisting Evaluation) is mostly used for evaluating the text-summarizations [44].

3.4.1 Perplexity

Language models with statistical approaches can be thought of as a probability distribution over all possible sentences [45]. The meaning of the sentence is not taken into account. The distribution of words is the main property for these models. The quality of a corpus is measured by comparing with the probability distribution that is extracted from a larger corpus of the language. Denoting M as the language model that is constructed from the larger corpus, the likelihood of new data can be used to assess the quality. Equation 3.4 formulates the average log-likelihood of new dataset $D = \{D_1, D_2, D_3, \ldots, D_n\}$.

$$\text{Average-Log-Likelihood}(D|M) = \frac{1}{n} \sum_i \log P_M(D_i) \tag{3.4}$$

The perplexity metric is based on the cross-entropy of the true data distribution P concerning the model distribution P_M. The definition of perplexity is 2^H where H stands for the entropy of the model over the sample. The language model can be generated for different n-grams. If the bigram model is chosen, the probability distribution gives the probability of a word followed by another word. If it is a unigram model, it simply calculates the occurrence probability of each word. For bigram models probability is

computed as $P(X_2 = x_2|X_1 = x_1)$.

$$H = -\sum_{D} P(D) \cdot \log P_M(D) \tag{3.5}$$

$$\text{perplexity} = 2^H \tag{3.6}$$

Equation 3.6 shows that if the entropy is lowered, the perplexity is also lowered. Since entropy refers to uncertainty, the metric basically measures how a given data set is similar to the language model in the context of word arrangements.

3.4.2 BLEU Score

Determining the quality of the machine translation models with human evaluation is expensive and takes months to finish. Hence, Papineni et. al. propose a new automated method for evaluation of machine translation called Bilingual Evaluation Understudy (BLEU) [43]. The proposed method is quick, inexpensive and independent from language.

The main measure of the metric is precision. The number of words that occur in both candidate translation and any of the reference translation is counted and it is divided by the total number of words in the candidate translation. Using unigrams for the metric computes the adequacy of the words in the translation where the longer n-grams matches account for fluency.

The length of the candidate translation is also important. It should not be too long or too short. Furthermore, if a word occurs more frequently in the candidate translation than its maximum reference count, modified precision is penalized. Another penalty is brevity penalty which expects a high-scoring candidate translation match the reference translations in length, in word choice and word order. However, this approach penalizes length deviations in short sentences harshly. First, test corpus' effective reference length, r, is computed by summing the best match lengths for each candidate sentence in the corpus. Later, brevity factor is set to decaying exponential in r/c, where c is the total length of the candidate translation corpus.

The geometric mean of the test corpus' modified precision scores are calculated and it is multiplied by brevity penalty factor. The geometric average of modified precisions for each n-gram is computed where n is up to length N. Summation of positive weights for each n-gram, w_n, is one.

$$\text{BP} = \begin{cases} 1 & \text{if } c > r \\ e^{1-r/c} & \text{otherwise .} \end{cases} \tag{3.7}$$

$$\text{BLEU} = \text{BP} \cdot \exp\left(\sum_{n=1}^{N} w_n \log p_n \right) \tag{3.8}$$

In the paper [43], n-grams are computed up to 4 with uniform weights. Namely, $N = 4$ and $w_n = 1/N$. In the log domain BLEU score becomes into 3.9. The BLEU score is in the range of $[0, 1]$. However, generally scores are multiplied by 100.

$$\log(\text{BLEU}) = \min(1 - \frac{r}{c}, 0) + \sum_{n=1}^{N} w_n \log p_n \tag{3.9}$$

3.5 Dataset

The dataset that we work on in this study contains nearly 24 million tweets in English about gun control in the U.S. 2,662 of the tweets contain geographical information. 13 million of the tweets have predicted location information. There are two sides of the discussion, pro-gun and anti-gun. Tweets are posted by 3,282,592 unique accounts. Users are labeled instead of tweets. The label of the tweet is accepted the same with the label of the owner. 293,046 of the users are right-winger while others are left-winger. Right-wingers are considered as pro-gun accounts. Almost 5 million tweets belong to right-wingers. Although the dataset is skewed, we sample equally from each side during our experiments. The training and test set size information can be found under *Experiments* sections of each chapter. Tweets are not pre-processed and contain unicode characters.

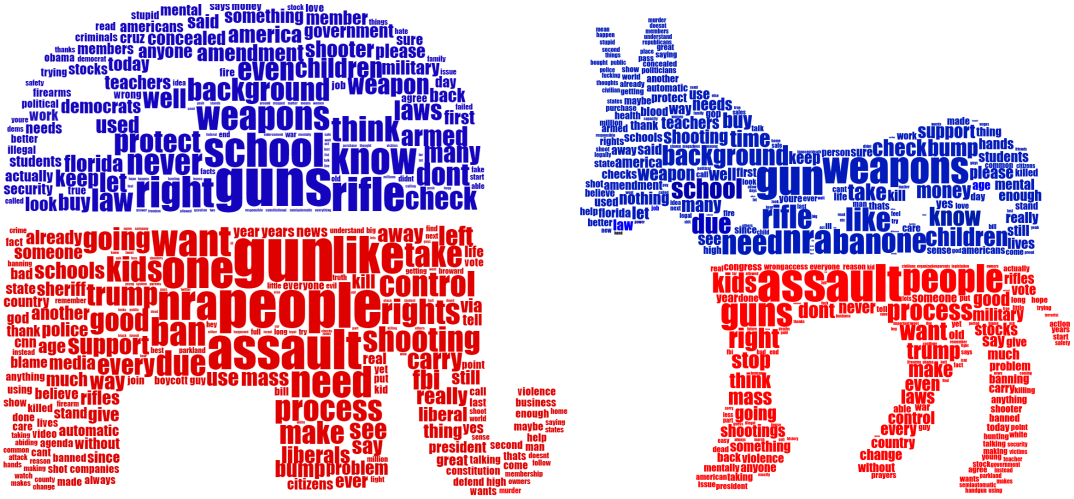

Figure 3.7: Word cloud that represents the top 350 words used in the tweets of both sides of the debate separately.

3.5.1 Preprocessing

Since Twitter limits the length of tweets, people use abbreviations, slangs, etc. This situation makes processing tweets harder. Additionally, emojis do not have standards. Any sequence of characters that looks like an entity is used as emoji. Therefore, we have to process tweets before using them in NLP tasks. All letters are converted to lowercase to get rid of the variation of the same words. Our bot does not have a module that can crawl a webpage and analyze the sentiment. Hence, the hyperlinks are not necessary and URLs are replaced with <*url*> tag. Similarly, thoughts are independent of the user information and <*user*> tag is inserted for any user mention inside the tweets. Hashtags might be used in the tweets while forming the meaning but they are used for supporting the idea mostly. <*hashtag*> token is used for any hashtag. We have replaced all numbers with <*number*> tag because behaving each number as a separate word is not applicable. Punctuations other than $\{!, ., ?, \%\}$ are simply removed. Simple emojis are changed with the definitive tags, such as <*smile*>, <*lolface*>, <*neutralface*>, <*sadface*>, etc. Emoticons which are emojis with unicode characters are purged. Lastly, we have replaced characters that occur more than 2 with up to the length of 2. After preprocessing is done, tweets are filtered according to left words. The ratio of non-tag words to the sentence needs to exceed some

threshold.

Algorithm 3.1: Algorithm for preprocessing tweets.

Process `PreprocessTweet(`*tweet*`)`**:**

```
tweet ← lowercase(tweet)
tweet ← replaceUrls(tweet, "<url>")
tweet ← replaceMentions(tweet, "<user>")
tweet ← replaceHashtags(tweet, "<hashtag>")
tweet ← replaceNumbers(tweet, "<number>")
tweet ← replaceEmojis(tweet)
tweet ← removeEmoticons(tweet)
tweet ← removePunctuations(tweet, except=!, ., ?, %)
return tweet ;
```

Table 3.3: Sample tweets after preprocessing is applied.

@RealDavidTafoya @charssocialism @shanragirl @moledwh79 @theogeo14 @KyleKashuv @kjcolvin611@davidhogg111 1.) The NRA does more than abortions. 2.) Stay out of women's health. urbodies, our choice.
<user> <user> <user> <user> <user> <user> <user> <user> <number>. the nra does more than abortions . <number>. stay out of womens health . ur bodies our choice .
@sweetestswim @LynseyDee2 @denise_kandra @TobyasStorm @gilbert87 @davidhogg111 @NRA And theClintons received hundreds of millions of directly or through their foundation . hat does that make her and the Dems?
<user> <user> <user> <user> <user> <user> <user> and the clintons received hundreds of millions of directly or through their foundation . hat does that make her and the dems ?
83% of Democrats and 54% of Republicans support raising the minimum gun purchase age to 21.So why isn't anything being done? The @NRA is bribing politicians, apparently sometimes even with Russian money.WAKE UP AMERICA!
<number> % of democrats and <number> % of republicans support raising the minimum gun purchase age to <number>. so why is not anything being done ? the <user> is bribing politicians apparently sometimes even with russian money . wake up america !

CHAPTER 4

POLITICAL VIEW ESTIMATION OF THE TWEET

In this chapter, we point out the first task and the corresponding model of our bot, namely tweet classification. The main component of the proposed model is LSTM which is successful for encoding sequences. We represent the tweets with GloVe embeddings. The encoding of the tweet is then connected to a fully-connected layer, FC, which is responsible for classification.

Experiments showed that the trained model learned the semantic of tweets about gun control well. While the models' prediction probabilities are getting higher, accuracy is also increasing. This implies that the model has the ability of a good generalization of tweets. Even it can not classify all tweets correctly, it is sufficient for a Twitter bot to accurately interpret a substantial number of tweets. Therefore, it is a good approach for relying on the tweets where the bot's prediction probability is above a threshold.

4.1 Classification Problem

As described in Chapter 3, the data set consists of two sides of a debate; left and right. Right-wingers think that keeping and bearing arms is an important right for citizens. Others are opposed to this idea and think that it is the main reason behind most of the deaths. In reality, it is difficult to draw a strict line between these opinions. There are various point-of-views, such as supporting owning only handguns but not rifles, wanting that all guns should be prohibited, making mandatory to do a detailed background check before getting an arm, etc. Since grouping each thought is eventually ended up with the number as many as the individuals exist, we simply chose the labels as right

and left.

4.2 Encoding the Tweets

In this section, we explain how we feed the LSTM to get the related encoding. The main aim of the network is moving sentences with natural language into the computer encoded space where semantically close tweets are close to each other. The data set is represented in this chapter as $D = (S, L)$, where $S = \{s_1, s_2, s_3, \ldots, s_N\}$ and $L = \{l_1, l_2, l_3, \ldots, l_N\}$ stand for the tweets and the labels. Tweets are tokenized to get sequences of words. Tweets, s_i, comprise of words denoted as $\{t_1, t_2, t_3, \ldots, t_m\}$ where m is the length of the tweet. Each word is represented with a GloVe embedding vector. Embedding vector representation of the tweet, s_i, is expressed as $\{v_1, v_2, v_3, \ldots, v_m\}$. Computed encoding of the tweet, s_i, is denoted as e_i.

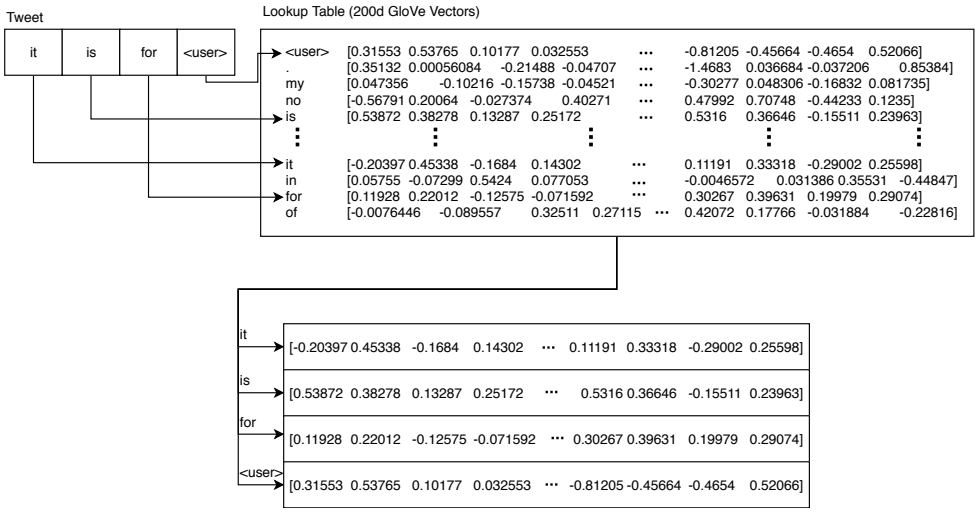

Figure 4.1: The lookup method for generating tweet representations.

The model is fed with v_i vectors. Since pre-trained GloVe vectors are used, the context is preserved in the representation. The model is not responsible for extracting the relationship between words. Instead, it captures the meaning according to the arrangement of these words. LSTM cells keep a state inside which is calculated with given sequences. The hidden state is updated when every vector, v_i, is sent to the LSTM cell. The state depends on the previous vectors given, $\{v_1, v_2, v_3, \ldots, v_i\}$. In

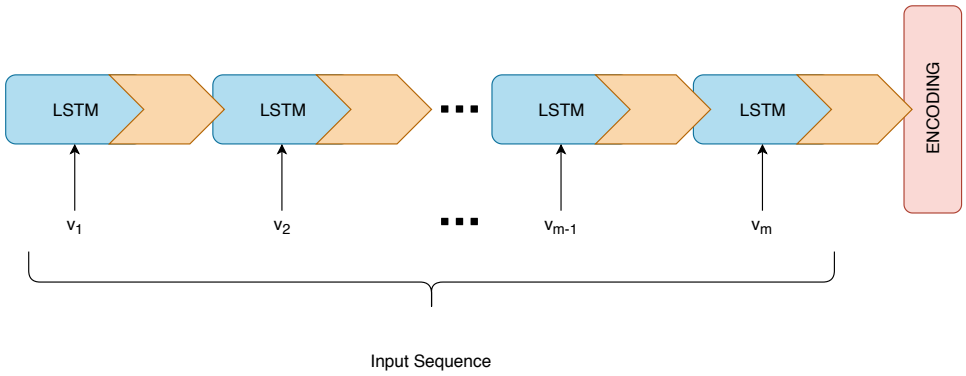

Figure 4.2: LSTM model for calculating the encoding of given tweet.

the end, the state, e_i, is the feature itself that is used for classification. The resulting state is important since the meaning can be captured only when the sentence is completed. Therefore, we are not interested in intermediate states.

We want the model to be more robust to unseen data. Therefore, the dropout technique is used in LSTM cells. It randomly chooses nodes and removes them for the current iteration. At each iteration, different neurons might be chosen. A hyper-parameter is used for detecting how many nodes to drop out. This method allows nodes to learn its weights independently. Otherwise, they might work together for representing the input during training which makes the model to memorize training data. Each neuron becomes more talented in extracting deep features with the coordination of random neurons.

4.3 Classifying the Encoding

The output of the LSTM, e, is connected to the FC layer for classification. There are output nodes as many as unique labels which are two in our data set. The size of the encoding may vary according to the model and it is denoted as D in this chapter. Every dimension of the encoding, e_j, is connected to these outputs with trainable weights, w_{jk}, where k stands for the label node. An additional bias term is added to the multiplication. Output of the label node o_k can be formulated as $\sum_j^D w_{jk}e_j + b$. An activation function is applied then to o_k for adding non-linearity. Final value of the node becomes $\sigma(o_k)$.

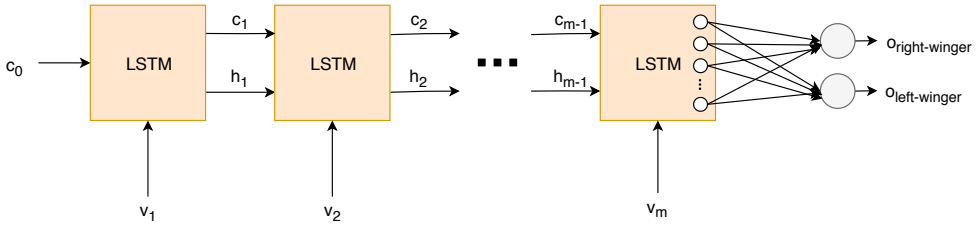

Figure 4.3: Fully-connected layer used for classifying the encoding produced by the LSTM.

The values of label nodes do not show the probability itself. They are raw scores, usually called as logits, and may differ between $-\infty$ and ∞. We need to turn these values into probabilities for each label. For this purpose, the softmax function is used as an activation function. It takes logits as input and returns a list of probabilities that sum to one, $\sigma(o_k) = \frac{e^{o_k}}{\sum_k e^{o_k}}$. The node with the highest probability is the prediction of the model.

While making wrong predictions, the network evolves and changes the way of computing the encoding by updating its weights. This training procedure is explained in the next section.

4.4 Overall Architecture

The training is done in two phases; forward propagation and backward propagation. First, we feed the system with data, starting from the LSTM. It calculates the encoding by the weights inside the LSTM cell. The final state is given to the FC layer for predicting the label. The process up to prediction is referred to as forward propagation. The term learning is basically learning the proper weights by comparing the predicted output and the targeted output. The model compares them for measuring the distance between the prediction and the true label which is called loss. It reveals how much the network is close to the truth. This loss is propagated backwardly for adjusting the weights to make correct predictions.

We used cross-entropy loss since the output of label nodes are the probabilities where tweet may belong to. It calculates the distance between two distributions, $-\sum_i y_i log(\tilde{y}_i)$.

In our network, these distributions are the prediction and the label. Hence, the loss can be denoted as $Loss = -\sum_k l_{ik} log(\sigma(o_k))$. In addition to cross-entropy loss, we added L2 regularization for avoiding the over-fitting problem. Adding a regularization term causes the weights' values to be decreased which is accepted as simple models. Therefore, it will prevent the model to memorize training data. L2 regularization is also known as weight decay since it adjusts the weights close to zero. An additional parameter, λ, is used for indicating the magnitude of the regularization. Finally, our loss function turns into $Loss + \lambda \times \sum ||w||^2$.

The targeted output is a one-hot vector where the correct label is one. The effect of each weight on the wrong prediction is calculated by taking derivatives of the loss function accordingly and they are updated. The backward propagation phase does not update only FC weights but also LSTM weights. Thus, LSTM learns locating tweets with the same label to close places in the encoding space. This procedure is repeated for several epochs until the model converges.

4.5 Experiments

In the experiments, we used 200-dimensional GloVe vectors which are pre-trained with Twitter data. 150,000 tweets are used from both sides. If a word is absent in this data, we used uniformly random vector with the values in range $(-0.25, 0.25)$. The size of the LSTM cells is 128. Drop-out probability and L2 regularization coefficient λ are 0.5 and 3 respectively. The learning rate for updating the weights according to the gradient is chosen as 10^{-4}. L2 regularization coefficient λ is determined as $\lambda = 3$. We have applied k-fold cross-validation to estimate the generalization performance of our model in a consistent way. The number of folds, k, is decided as 7. The model uses ADAM optimizer for adjusting weights.

Table 4.1: Classifier performance with 7 fold cross-validation on different filter threshold values.

	Accuracy		$\text{Precision}_{\text{left}}$		$\text{Recall}_{\text{left}}$		$\text{Precision}_{\text{right}}$		$\text{Recall}_{\text{right}}$	
τ	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
0.50	74.07	0.003	73.65	0.019	74.29	0.007	74.48	0.016	73.90	0.009
0.55	75.75	0.004	75.33	0.020	75.99	0.008	76.15	0.016	75.56	0.010
0.60	77.30	0.006	76.97	0.019	77.47	0.008	77.59	0.016	77.17	0.010
0.65	78.96	0.007	78.67	0.018	79.19	0.010	79.22	0.018	78.78	0.010
0.70	80.64	0.008	80.46	0.018	80.85	0.012	80.78	0.019	80.48	0.010
0.75	82.55	0.009	82.63	0.016	82.73	0.013	82.43	0.020	82.42	0.010
0.80	85.06	0.011	85.59	0.014	85.16	0.015	84.47	0.021	85.00	0.010
0.85	**89.99**	0.016	**91.99**	0.020	**89.71**	0.020	**87.40**	0.032	**90.48**	0.015

The accuracy we obtain is 74.07% on average of 7 folds. It can be thought of as good performance for an NLP task. However, for making our bot consistent in the context of generated tweets, it is better to be higher. In addition, just like real Twitter users, our bot should not interact with all accounts concerning casting doubt on it. Therefore, we come up with the idea of filtering these tweets to detect which tweets should be taken into account or not. This filtration is done with a threshold of prediction probability. We examine the performance of our classifier on these filtered tweets. It showed that our model predictions are more accurate on these tweets.

Table 4.2: The ratio of tweets that classifier has prediction probability above different filter threshold values to the whole test set with cross-validation with 7 fold.

	Tweet Ratio	
Threshold	Mean	Std
0.50	100.0	0
0.55	93.17	0.005
0.60	86.47	0.011
0.65	79.50	0.016
0.70	72.06	0.022
0.75	63.34	0.027
0.80	51.48	0.036
0.85	27.91	0.062

As shown in the Tables 4.1 and 4.2, the tweets that our model predicts the label with the probability above 80%, covers 51.48% of all test set in average. However, it dramatically boosts the performance up to 85.06%. Since using half of the tweets for replying is still a big portion, we filtered them with a threshold of 85% which narrows the tweets down to 27.91%. It ends up with a reasonable number of tweets and the accuracy of 89.99%.

Believing that adding more features may help the model to capture the meaning, we add part-of-speech tags along with GloVe vectors and do the same experiment on it. We used Stanford POS tagger to tag tokens inside the tweets [46]. We convert these tags into one-hot vectors since there is not a hierarchical relationship between them. These vectors are concatenated with GloVe vectors. Other hyper-parameters are kept the same. Results are surprisingly worse than the one without POS tags. We observe that adding this feature lowers each score 1-1.5 point. Therefore, we do not use them in other chapters of this thesis.

4.6 Chapter Discussion

In this chapter, we used LSTM for detecting the side of a tweet in a debate. It has great power on extracting the semantic by looking at the arrangement of words. We observed that the state of the cell, which is the encoding of the tweet, correctly represents the place of the tweet in the encoding space conforming to the label. This model is used for while retweeting and determining which tweet should be replied.

The words used in both sides of the debate are highly overlapped as can be seen in Figure 3.7. This situation dissuaded us from using the technique that looks only the tokens inside the texts. Since the meaning is constructed by the composition of the words, the representation should be extracted from it. As mentioned in Chapter 2, LSTM uses this information to encode the given input. Hence, we went over our research using this model.

More descriptive features may be used alongside with the embedding vectors. We tried part-of-speech tags since it encloses fundamental information on the language. It did not work in our data-set suitably. However, it may because of the nature of Twitter contents. They involve lots of grammatical mistakes, misspelled words, non-standard abbreviations. It might be a successful auxiliary feature for data sets containing proper texts, such as books, Wikipedia, etc.

In addition, we wonder which tweets are predicted with low probabilities. When we examine them, we saw that there are no clues in the tweet itself. Their content mostly does not show an opinion specific to the side. Some of them are listed in Table 4.3. Distinctive properties may be found inside the URLs they shared or on the previous replies. Crawling the web-pages and classifying their content may increase the performance.

Table 4.3: Sample tweets that the model predicts the label with low probabilities.

Tweet	P_{left}	P_{right}	True Label
next time someone wants to talk about <hashtag> share this about <hashtag> <hashtag> <url>	49.87	50.13	Right
this is peak crazy america <url>	51.87	48.13	Left
<user> a background check on potential political candidates is not a bad idea either	48.07	51.93	Left
warning you can not unwatch this . <url>	52.23	47.77	Left
<user> first day back was kinda rough but there was a lot of support and plenty of good doggos many people were asking how	46.88	53.12	Left
<user> i did yes . and ill stay quiet until ive read enough to have a useful opinion . the shouting match we see in the immed	53.64	46.36	Right
cant do anything about social posts sry this is america still . <url>	46.21	53.79	Right
this whole video is a joke these people are so ignorant <url>	54.63	45.37	Left

The main focus of this thesis is generating a bot that can comprehend the environment on Twitter and to interact with other users just like human-beings. Therefore, it should not intervene in every discussion. For this reason, filtering the tweets that the model is sure above a predefined percentage is functional for our purpose. Given performance statistics show that the model works very well on a sufficient amount of tweets. Replying 27% of the posted tweets is still suspicious with respect to real accounts. However, in order to not being disclosed, generated tweets should look-alike human posted tweets. This requirement is addressed in other chapters.

4.7 Chapter Conclusion

In this chapter, we applied an LSTM classifier over tweets to detect whether the post belongs to Right-winger or Left-winger account. It detects the side of the tweet with high performance. The model's consistency is approved by narrowing them respect to the prediction probability. While increasing the threshold, accuracy is also increased persistently. It is enough for our bot to go on with a sufficient amount of tweets.

The achievement of LSTM on encoding texts convinced us to use it in other tasks for our bot, namely, text generation and reply generation. We used it as the main model for these tasks with some modifications.

CHAPTER 5

TWEET GENERATION IN THE CONTEXT OF A DEBATE

Posting predefined tweets from multiple accounts is a bad approach since it is easy to be disclosed by other users. Therefore, generating random but logical texts is a must for a smart bot. As mentioned in Chapter 4, the arrangement of the words is the main feature of forming the semantic. While generating text from scratch, the computer starts with a random word and adds a new one that is more probable to follow it until the end.

In this chapter, we use a quite different LSTM model than the one adopted in the classifying task. Instead of predicting the label of the whole tweet, we need to train the model to predict the next word in accordance with the current state depending on preceding words up to the point. During the model determines each word, the state is updated and the meaning appears. GloVe embeddings are used for representing the words. There are two models that are imitating two sides of the debate. One is trained with the tweets of left-wingers and the other is with right-wingers. Generated tweets seem admissible in pursuant of perplexity metric and human evaluators.

5.1 Designing Model for Next-Word Prediction

To generate random tweets, first, we need to model the language properly. Instead of using hand-crafted features, we take advantage of LSTM to reveal the hidden features that are correlated with the relationship between words. The model extracts the statistics of word arrangements similar to traditional n-gram models. However, natural languages are moreover than predefined phrases.

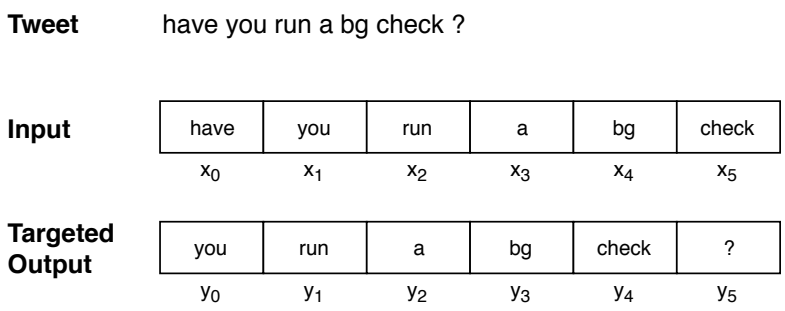

Figure 5.1: Training set prepared for next-word prediction task.

We bring each tweet, $s_i = t_{i1}, t_{i2}, t_{i3}, \ldots, t_{im},$, to same length, M, by padding them with special tag denoting the end, <*end*>. In this manner, the model learns when and how to conclude the tweet. LSTM decides to put an ending token when it completes the tweet. Otherwise, it gives random next-words up to predefined sequence length without a stop. It is impossible to determine a fixed-length where the tweet is ended with a logical meaning. In the first epochs, the model understands that when <*end*> token appears, it should be continuously followed with other <*end*> tokens.

The output layer consists of multiple nodes, each representing one word in the vocabulary with the size of N. The vocabulary is generated from the training set. Similar to label prediction, the model computes the probability of each word. Since it limits us to only seen words, we tried to predict the GloVe representation of the following word by converting the task into a regression problem. However, we observed that the model does not converge to a point due to the difficulty of predicting 200 floating-point values correctly at the same time. Consequently, we proceed with predicting the label considering the generated tweets are still genuine.

Since we aim to construct a model that can form a sentence by predicting the next word, targeted outputs are prepared with the inputs itself as visualized in Figure 5.1. Tweets are one word shifted to obtain the inputs and the labels. For instance, if the tweet, s_i, is *"the truth will set you free"*, then input and output are respectively; $x_i =$ [*"the"*, *"truth"*, *"will"*, *"set"*, *"you"*] and $y_i =$ [*"truth"*, *"will"*, *"set"*, *"you"*, *"free"*]. While moving on the sequence in time, the model slides over the input and the true output lists. The essential part is that the values in x_i list are GloVe vectors. However, the ones in the y_i are one-hot vectors where the cell indexed with the corresponding

word is one.

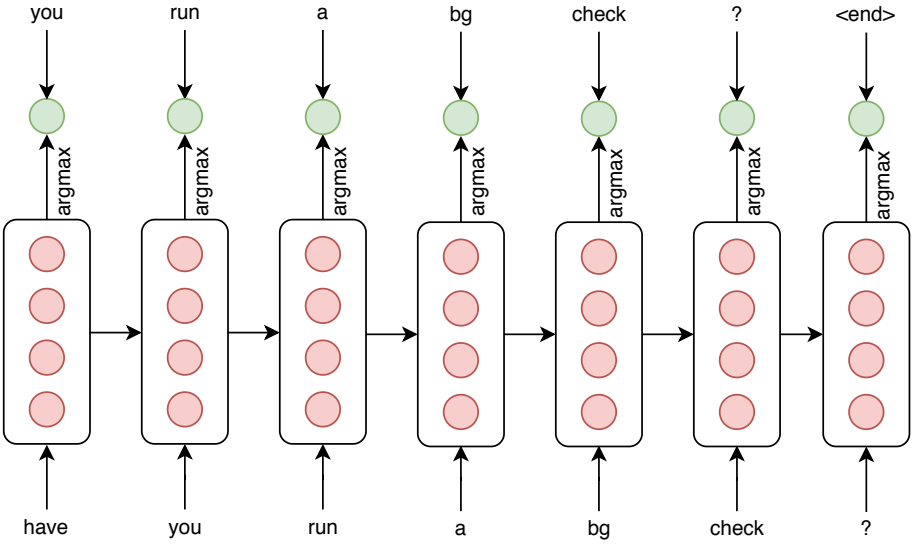

Figure 5.2: The architecture of language modeling.

Word prediction is done similar to the label prediction in Chapter 4. However, this time intermediate states, h_j, represent the intermediate words. The state is connected to an FC layer for estimating the most probable word. Figure 5.2 demonstrates the training procedure. The resulting formula for output becomes $o_k = \sum_j^D w_{jk} h_j + b$. For making each value to refer a probability, softmax function is applied to these logits. For optimizing the model, we use sequence loss which is a cross-entropy loss. The difference is that it is calculated over a sequence with given weights. The mistakes done at the beginning of the sentence are as important as the ones done at the end. Therefore, we set all the weights equal. In total, it sums up the cross-entropy loss of overall tokens.

5.2 Constructing Tweets Using the Language Model

The sampling phase is done after training is completed. The methodology followed for sampling is as important as the language model. Generated tweets may be boring if they always trace the same path. In this section, we cover how to select words to arrange a sentence that reflects the opinion using the trained model. The aim is to generate unique tweets that look like a real human post. Hence, there is no absolute

correctness for the outputs.

The model produces samples with a given prime word sequence. Results are more promising when the prime is selected with the beginning of tweets in the training set. However, unseen primes also bring logical sentences out. Prime can be imagined as the preparation of LSTM. The state inside the cell needs to be initiated with a value to be canalized to a point. The model is incapable of forming a sequence with an empty state. When the state is ready for prediction, it completes the rest of the tweet.

We designed the model to estimate the most probable following word, up to the end. FC layer computes the probability of each word to appear next with the current state of the cell. The word with the highest probability is accepted and appended to the sentence. We feed the model with the embedding of the chosen word to determine the succeeding state as shown in Figure 5.3. This process repeats until the model outputs the <*end*> tag. Since we arrange the training set accordingly, it can form the semantic and conclude the generation when the meaning appears. Generated tweets seem very similar to the real ones. However, this approach always comes up with the same result for the same prime. Additionally, choosing the most probable word at each step may end up with a local minimum. Specifically, choosing the best word for a step might congest the meaning for the whole sentence.

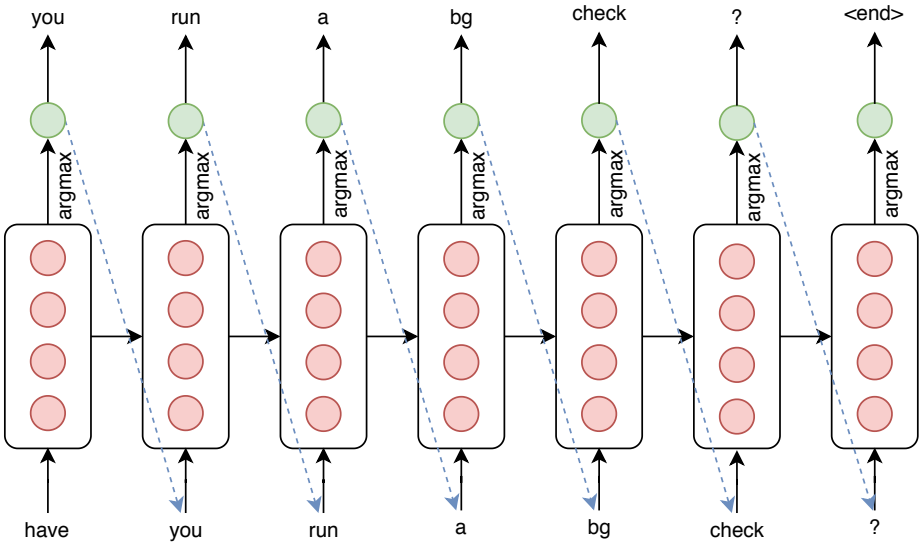

Figure 5.3: Tweet generation using LSTM model with next-word prediction.

We decided to add some randomness to choosing the word to get more characteristic tweets. However, selecting the next word in a fully random manner disrupts language modeling. Therefore, we sort the words by their probabilities and choose one of them randomly where higher probability words have more chance. This method increases the originality of the generated tweets. Besides, the sequence lasts longer because we force the model to not signal <*end*> tag. Nonetheless, this situation leads the model to produce tweets with nonsense from time to time.

As a result, we go on with a heuristic for choosing the next word. The beam search is applied to the prediction probabilities. It expands given the number of possible branches for estimating the most promising sample. At each time step, non-promising alternatives are pruned from the hypotheses space. Sum of negative logarithm of all chosen words' probabilities in the sentence is used as scoring value, $\sum_{j}^{D} -log(\tilde{y_j})$. Since the addition of logarithm gives the multiplication of the values inside, the score represents the multiplication of probabilities of each prediction. Therefore, predicting a single word with high probability is not enough itself to accept a generated tweet to have a high score. It prevents the tweet to go in the wrong direction by doing one wrong prediction. We observed that this method produces more stable sentences than other approaches.

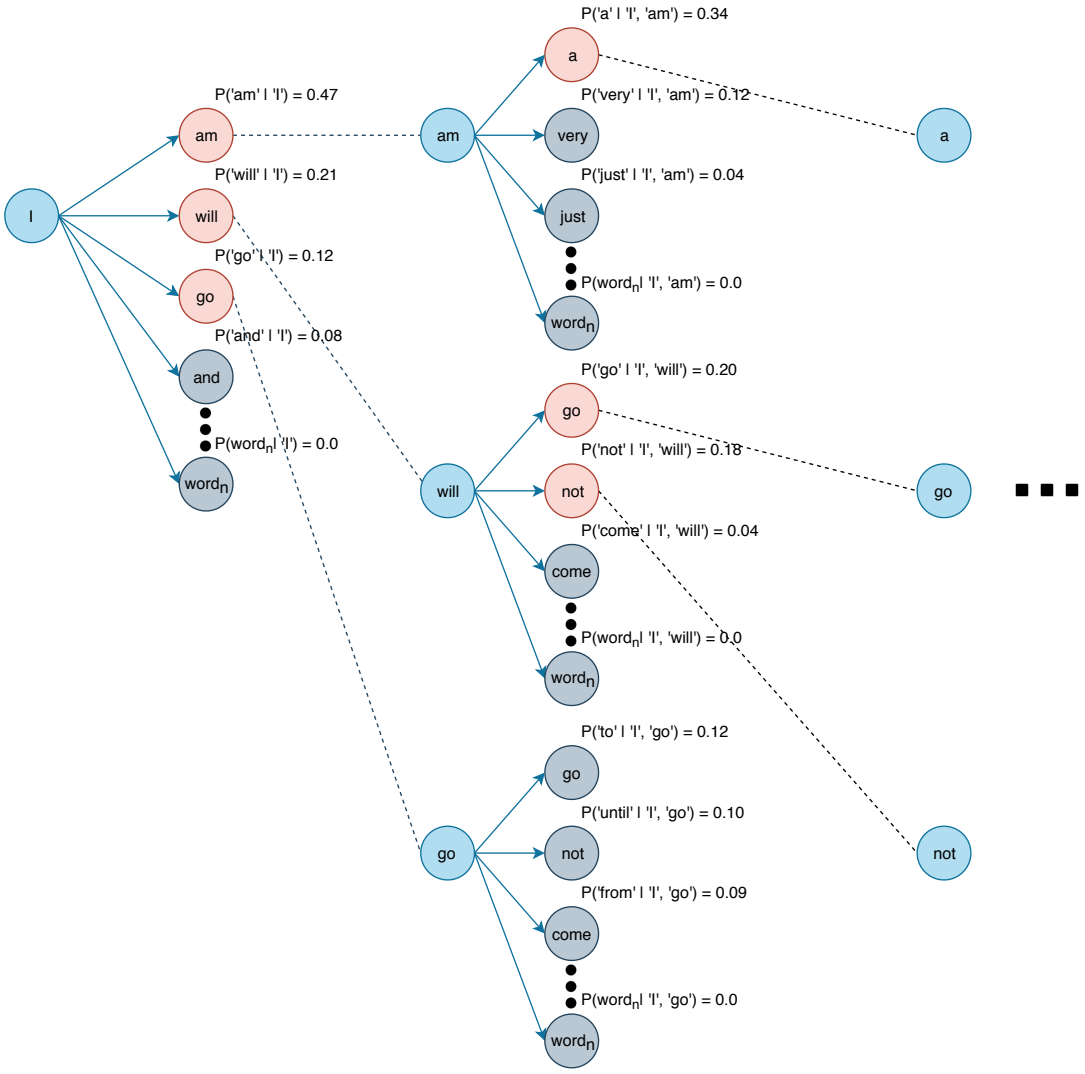

Figure 5.4: Choosing word with beam search.

5.3 Experiments

For representing the words, 200-dimensional GloVe vectors are used. The words do not exist in the pre-trained GloVe model are represented with special tag <*unk*>. There are 2 layers in LSTM. The size of the cells is 256. The decaying learning rate is used for the model starting with the value of 10^{-3}. At each epoch, it is decreased with 3%. ADAM optimizer is used for applying the gradients according to the loss. The drop-out technique is not used for this model. The training lasted for 32 epochs. Two different models are trained with the same configurations; one for left-wingers and

one for the other side. For both models, training sets are the equal size of 750,000 tweets. For comparing the quality of generated and real tweets, 65,000 tweets are kept as a test set. Generated tweets are sent to the classifier explained in Chapter 4. The tweets with the highest probabilities are used from both real and generated tweets during the evaluation.

Generated tweets are evaluated by human evaluators in two criteria. Namely, we want from our participants to score the tweets according to *Meaning* and *Quality* between the range of (1-5) where 5 refers the best. *Meaning* score denotes whether the tweet reflects the view of the discussion side. *Quality* of the tweet is whether it looks like a human-posted tweet. Each tweet is assessed by 3 human evaluators and 128 tweets are used in the experiment. Tweets used in the evaluation can be seen in Table A.1. In order to compare with the original tweets, half of them are real tweets from the corpus.

Table 5.1 shows that generated tweets are scored as high as human posted tweets. The *Quality* score of the tweet is very important for the bot to deceive other users easily. Fake tweets have also good *Meaning* score relative to the real ones. It means that the model is able to produce tweets that reflect the side of the bot in the debate. Since we aim to develop a bot that supports a side in this thesis, together with *Quality*, *Meaning* occupies an important place.

Table 5.1: Human evaluation results for generated and real tweets.

	Meaning	Quality
Real Right-Winger Tweets	3.73	3.73
Generated Right-Winger Tweets	3.55	3.56
Real Left-Winger Tweets	3.66	3.77
Generated Left-Winger Tweets	3.62	3.52

Additionally, we prepared a multiple-choice test where containing 3 real user tweets and 1 generated tweet as shown in Table 5.2 to test the indistinguishability of our tweets. Obviously, the model is able to generate tweets that are highly similar to the real ones. The evaluator is expected to pick out the fake tweet. There are 60 questions

where each one is answered by 7 participants. Questions are listed in Table A.3.

Table 5.2: Sample questions on multiple choice test for finding the generated tweet from among the real ones. Generated tweets are marked in bold.

Choices	Wing
(a) im so tired of this drama king ! <url> **(b) im so tired of the left s face of bigotry and hate of the nra . <hashtag>** (c) im so tired of hearing the term assault weapons . (d) im so tired of hearing about <hashtag> it didnt work today and it wont work tomorrow	Right
(a) this is why people need the right to protect themselves from the liberal left . <url> (b) this is why people need to arm themselves . period ! <url> (c) this is why people need to prosecuted for <hashtag> <hashtag> **(d) this is why people need to be heard ! <url>**	Right
(a) unlike the <hashtag> started pointing out what to form the constitution . (b) unlike the <hashtag> crowd who digs in and gets it done on behalf of everyone s <hashtag> <url> (c) unlike the <hashtag> these heroic cops ran toward the florida school shooting . <hashtag> <hashtag> <url> (d) unlike the <hashtag> hollywood does promote criminal <hashtag> and <hashtag>	Right
(a) <user> you mean like the cop that was at <hashtag> (b) <user> you mean like those poor kids that were slaughtered in <hashtag> . <url> **(c) <user> you mean that you are not a gun owner . <hashtag> <hashtag>** (d) <user> you mean warriors of the left ? . <hashtag> <hashtag>	Left
(a) this is about a hard decision . god bless our kids and teachers <hashtag> <hashtag>	

(continued)

Table 5.2 *(continued)*

Choices	Wing
(b) this is about guns . guns . if you cant see it you dont want to see it . <hashtag> <hashtag> <url>	
(c) this is about <hashtag> for our <hashtag> reveal plan to destroy <hashtag> in bed with <hashtag>	
(d) this is about capturing a ground swell of emotion against the nra and for stricter gun control laws . <user> on <hashtag> <url>	Left
(a) children are our future has never been more true . humbled and grateful for the <hashtag> students speaking truth to power . <hashtag> <url>	
(b) children are our future ! you stand the <hashtag> and stop taking money from the nra . do the right thing or <hashtag>	
(c) children are our future ! please support them ! ! ! <hashtag> <hashtag> <url>	
(d) children are our future . but how can they have a future if our gun laws dont change ? <hashtag> <hashtag> <url>	Left

Table 5.3: Human evaluation results on multiple-choice test for finding the fake tweets among the real ones.

	Nondisclosed Tweets	
Correct Participants	Right-Winger	Left-Winger
2	26.5%	16.5%
4	70.0%	53.5%
5	100.0%	93.5%
7	100.0%	100.0%

In Table 5.3, the ratio of tweets that participants can not find the generated one on multiple-choice tests is listed. The correct participants column shows that if this number of the evaluators finds the generated tweet for a question, then we count it as revealed. Additionally, if we handle the evaluations for each question separately, we get the probability of a generated tweet to be not disclosed as 77.5 % and 60.5 % for

right-winger and left-winger tweets respectively. The tweets generated by our model seem highly similar to the real ones.

Furthermore, we measure the quality of the generated tweets with respect to the statistical method. Perplexity uses the probability distribution. Low perplexity means that the output is as good as the training set. To compare our tweets with real tweets, we used the separated test set with the size of 65,000 which is not used in training. 65,000 random generated tweets are used for both sides. The language model is generated from a 1.5 million size tweet data set. For each wing, we used different corpus. Table 5.4 indicates that our generated tweets have also close bigram perplexity to the real ones.

Table 5.4: Perplexity values of generated and real tweets.

	Perplexity
Real Right-Winger Tweets	96.71
Generated Right-Winger Tweets	109.71
Real Left-Winger Tweets	118.65
Generated Left-Winger Tweets	130.13

5.4 Chapter Discussion

In this chapter, we aim to generate high-quality logical tweets. The results are very promising as reported by human evaluators. Different sampling approaches are tried to produce both unique, diversified and meaningful tweets. They have pros and cons over each other. We used all methods for generating tweets. However, the classifier mentioned in Chapter 4 is used as a grader through them. The tweets with the highest probabilities on the correct side are used.

The straight-forward method which is choosing the next word with the highest probability gives good results. It is able to form smooth sentences. The tweets of this approach have a reasonable length. Also, they have a good representation of the standpoint of the bot. However, the resulting tweet might be predictable according to

the given beginning words. We feel that the model is restricted to the given primes. The beginning of the tweet specifies the main theme of the tweet.

Adding randomness to word selection procedure gave us what we want. This approach ends up with more creative tweets. It does not limit the model to narrow the space and give a chance to explore new paths. Generated tweets are more open to produce statements and be wordy. Randomness prevents model to finish the generation early and promotes it to continue by explaining its idea. This manner frequently comes up with really high-quality tweets. However, it sometimes builds meaningless or contradictory tweets.

Using a heuristic function for choosing the best next-word is another approach that we applied. Beam search chooses the most promising branches and explores the space accordingly. The structure of generated tweets was good enough comparing to real tweets. On the other hand, this approach made our bot more discreet. The outcome was brief and to the point. The semantic was hidden inside the words and needs to be clarified for further meanings.

All of the aforementioned methods are used for our bot. However, we needed an internal scoring mechanism to select the relatively best sample among them. Therefore, we generate multiple tweets with multiple approaches and test it with the classifier we trained in Chapter 4. It, in addition to classification, also gives the score of the semantic. More clearly, if we want to generate right-winger tweets, desired tweets should pass the classifier with the right label probability above 85%.

We observed that LSTM has a huge representation power on languages. In contrast to traditional rule-based n-gram models, LSTM extracts the properties of words. In other words, it can arrange phrases that do not exist in the training set. For example, even a 2-gram *"strengthen military"* does not exist in the corpus, the model can produce the tweet *"we are going to strengthen military"*. This is only possible with being aware of that *"military"* word is a noun and *"strengthen"* verb is usually followed with nouns.

Evaluation results have been shown that our model can produce human-like tweets. Human evaluators are mostly unable to distinguish the generated tweets from the real

ones. *Meaning* and *Quality* scores of produced and real tweets are very close. The generated tweets have a common theme and do not contradict in itself. It is highly skilled to express itself clearly and to attack the dissenting opinion acutely. If the chaos in Twitter is considered, we have a strong belief that our bot is not prone to be disclosed.

Syntactic measurement of our model also reveals that it has extracted the fundamental properties of the language. The perplexity score of the randomly generated tweets is close to the one with the real users' posts. It reveals that the arrangement of the words to form up a sentence is very similar to the real sentences. This statistical measure gives an idea about the performance of our model.

5.5 Chapter Conclusion

In this chapter, we proposed a model that produces tweets by next-word prediction. Sampling is done with 3 different approaches and the resulting tweet is picked among them according to its score. The score is determined by the classifier which specifies the weight of the tweet between sides of the debate. The results of our experiments are acceptable for a Twitter bot that can support an argument. Two main issues are tested with experiments. Namely, whether generated tweets look like real tweets and whether they include an opinion that corresponds with the perspective of the desired side. Three different evaluations are performed to quantify the results. All tests are passed when comparing the results of generated tweets with the results of the real ones.

The text generation part we propose might be enough for the users that browsing the homepage of our bot account. It states regular sentences, expresses opinions, supports a view, etc. However, a Twitter account that only tweets but not replies the mentions may seem odd. In the next chapter, Chapter 6, we introduce a different approach for generating texts to address this issue. The next model produces tweets that fit with both context and the semantic of the tweet desired to reply.

CHAPTER 6

REPLY GENERATION TO AN OPPOSITE VIEW

To mimic real Twitter accounts, posting meaningful tweets is very important but not enough. It should also interact with other users to support its idea. Interaction in Twitter is done with two mechanisms; reply and mention. Mention is used if a user wants a specific user to get notification indicates that she/he is talking about something related. Tagging real user accounts just before or after the generated tweet is adequate. Reply task is unlike the tweet generation which is covered in the previous chapter. The given answer should be related to the origin post. Specifically, it should not include statements about a totally different argument.

In this chapter, we apply a quite different model than the one used in Chapter 5. The main component is still LSTM but the arrangement is different. Seq-seq LSTM fits our issue perfectly. However, starting text generation before seeing the whole text comes up with unrelated output tweets. The model needs to be fed by the origin tweet until the end to catch the correct meaning. Therefore, we pick the encoder-decoder model. The encoder part is in charge of extracting the meaning from the source tweet. The relation between the arrangement of words and the semantic is kept inside the states. This encoding is important to bring a relevant reply out. The decoder part is subject to select the next-words for forming a sentence from the encoding vector.

This model is very popular for addressing several NLP problems including text generation such as question-answering, neural-machine-translation, etc. [47][21]. It can comprehend the given sequence and produce a new one from it. Human-evaluators and kn-BLEU are used for measuring the performance of our model. BLEU becomes a common metric of machine translation studies [43]. Due to the nature of BLEU, it restricts creativity. Therefore, we propose a new metric kn-BLEU which is based on

BLEU.

6.1 Extracting the Meaning from the Tweets

In this section, we point out how we feed the LSTM to get the meaning of the source tweet. Similar to Chapter 4, we intend to project tweets into the encoding space. In this manner, we need to encode the meaning into a fixed-length vectors. For this purpose, we used LSTM with encoder-decoder architecture. The data set consists of source tweets as inputs and replies as labels; $D = (S, R)$, where $S = \{s_1, s_2, s_3, \ldots, s_N\}$ and $R = \{r_1, r_2, r_3, \ldots, r_N\}$. Words in the sentences are represented with GloVe embeddings just we did in other chapters. We use the same notation for repersentations of the tweets here as $\{v_{s1}, v_{s2}, v_{s3}, \ldots, v_{sm}\}$ for source and $\{v_{r1}, v_{r2}, v_{r3}, \ldots, v_{rm}\}$ for reply.

The encoder gets source tweet as one word at a time. At each time, the model is fed with v_{si} and the state is updated accordingly. While getting new embeddings as input, the meaning starts to be encoded into the state vector, $h_i = f(h_{i-1}, v_{si})$ where f is the abstract function representation of the network. Unlike the model we applied in Chapter 5, we do not work on intermediate states because the meaning is not concluded yet. Since one word can directly turn the meaning to the opposite in natural languages, the model should wait until the sentence is over. The resulting state represents the meaning of the tweet and can be used for various purposes. In Chapter 4, we used this encoding intending to classify the sentiment. In this chapter, we use it for generating replies to it.

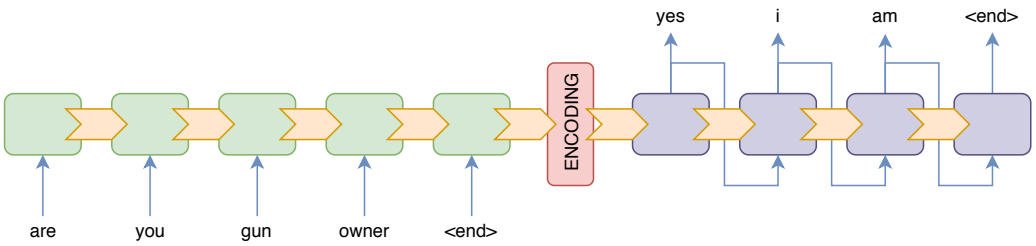

Figure 6.1: Encoder-decoder architecture for reply generation.

We use another sequence of cells, called decoder, for text generation. It gets the last

state of the LSTM cells as input and forms a sentence that is a reply to it. In addition to the final state of the encoder, we applied the attention mechanism to our model for getting a more related reply. There is a layer responsible for extracting the relation between words of source sentences and target sentences. Namely, it encodes the context into a vector. This information helps the model to generate the replies within the context. Recent researches have shown that using this mechanism improves the model performance [48].

6.2 Generating Reply According to the Encoding

Extracted features are fed into the second part of our model, decoder. It is responsible for forming a new sentence conforming with the source tweet. The procedure is done basically with predicting one word at a time. The following word depends on the encoder state, attention vector and the current state of the decoder. Therefore, the conditional probability of the sentence can be formulated as the summation;

$$\log p(y|s_i) = \sum_{j=1}^{m} \log p(y_j|y_{<j}, v_j) \tag{6.1}$$

Similar to Chapter 5, the prediction of the following word is done among the vocabulary. The model assigns a probability of occurrence for each word in the vocabulary and selects the most probable one. Hence, we apply the softmax activation function to get probability values from logits. Sequence loss is used for optimizing our model. It sums up the cross-entropy loss across the predicted sequence. The weights of the loss along the sequence are set equal since the meaning depends on all words equally.

Cells in the model should not work depending on each other. Every cell should learn its job perfectly and if it can emit important information by itself, then it should. Therefore, we applied dropout to our model to make it more robust to over-fitting. It shutdowns random nodes and behave like they do not exist. Selected nodes differ at each iteration of training. In this way, nodes undertake more responsibility and improve their skills. Furthermore, we used it to add randomness to our model while sampling replies. Since LSTM networks are very powerful, it is prone to generate

memorized sentences. It makes our bot weak on hiding itself. We observed that using dropout on the generation phase is also useful. It changes the followed path from an ordinary one to the original but still related one.

6.3 kn-BLEU - New Metric on Measuring the Relatedness

BLEU is a metric for measuring the quality of the translation by comparing a candidate translation to the reference translations. It becomes very popular because of being quick, inexpensive and language-independent [43]. Since there may be alternatives to correct translation, it can get multiple reference translations to compare. The score is based on searching the n-grams of the candidate in the reference translations and the length of them. The calculation ignores the word order. Unigrams can be used in this metric.

For training a Twitter bot that can participate in discussions, the BLEU score does not fit perfectly to measure the quality of the reply. This is mainly because of that we can not talk about a reference replies for the tweets. Besides, we want our bot to produce original tweets that spoil the metric. Therefore, we propose a new metric kn-BLEU where kn stands for k-nearest. It is based on BLEU. Instead of comparing the generated translations with targeted ones, we widen the space of references with similar contents. Since we want to measure the relatedness of the generated tweet, it is valid for our purpose. Chosen words can give a thought about the context. Hence, we worked on unigrams while testing our metric. Figure 6.2 visualizes the flow of proposed metric.

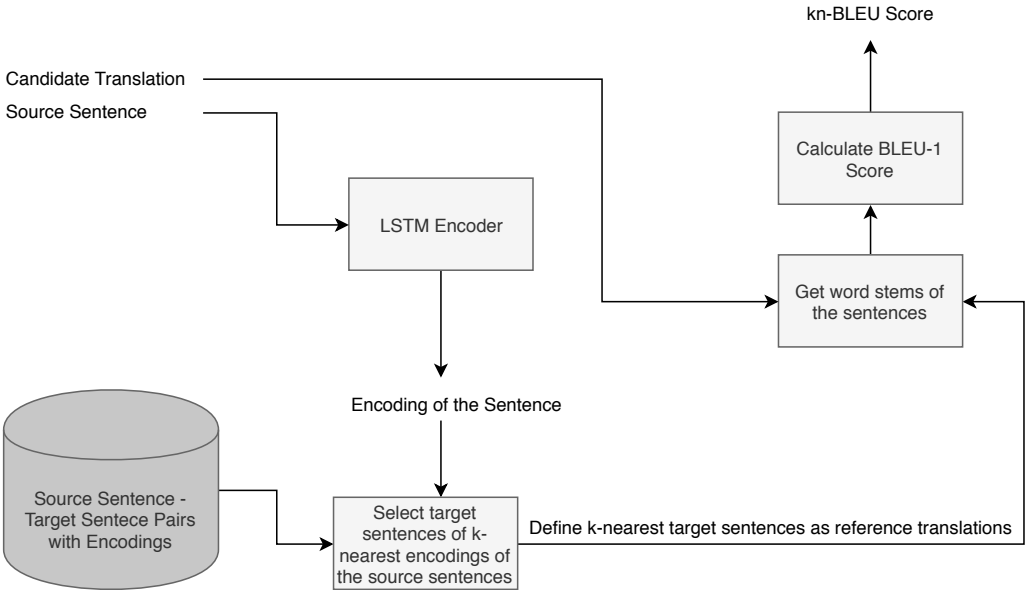

Figure 6.2: The flow of kn-BLEU metric that measures the relatedness of the translation.

The main procedure is encoding the sentences into fixed-length sentences. We used the encoder-decoder LSTM model for this aim. The model is trained for regenerating the text itself. The output of the encoder contains very informative features about the sentence. It moves the sentences from natural languages to the hypothetical context space. We used this space to find the sentences with the closest semantic to the source sentence. We observed that the closest sentences might have similar meanings or words. Some of the samples can be found in Table 6.1. To establish the functionality of the approach, the model is trained with the source sentences of the WMT news-test-2014 dataset in the German language. The table contains the matching reference translations of them.

Table 6.1: k-nearest context to the given sentence with the WMT news-test-2014 test dataset where k is 5.

Source Sentence	Closest Sentences
Maybe we ’re more " Mexican " than I thought .	(1) We ’re not FC Barcelona ! (2) This year , we ’re really committed to winning the championship . (3) In Mexico , we ’re seen as Americans . (4) " We get treated like criminals , " Mr. Orozco said . (5) However , Mexican law recognises the possibility .
" Weapons were stolen by Russian officers and by the Caucasians , " says Baranets .	(1) There were twenty of us , including Sharon Stone and John Kennedy Jr . (2) Several bills were blocked by vetoes of Democratic governors . (3) Next are " black weapons , " stolen by criminals from representatives of defence agencies . (4) Both projects were destined for military use of rocket engines . (5) Manufacturers of weapons make their contribution , according to Baranets .

(continued)

Table 6.1 *(continued)*

Source Sentence	Closest Sentences
The Ministry of the Interior does not put arms from the illegal market back into circulation	(1) Nonetheless , the Ministry of the Interior asserts that the situation of the spread of illegal arms is under control . (2) The eight planets of our solar system , plus the dwarf planet Ceres . (3) This image was put together from the X-ray images captured by various telescopes . (4) The Russian Ministry of the Interior is proposing to toughen up the law for owners of civil weapons . (5) The latter caused a sensation by explaining the mechanics of the public contracts " system " to the commission .
- What do you mean ?	(1) Do you want him ? ’ (2) How do you select them ? (3) What now ? (4) - What do you think about voice control ? (5) How do you explain this progression ?
It ’s not an easy task .	(1) And that ’s not all . (2) That ’s not going to happen . (3) There ’s an awful lot of questions . (4) That ’s fair . (5) That ’s a difficult one .

We used the closest sentences as alternative translations in the BLEU. Since we want to measure the word selection quality of our model, unigrams are compared. Moreover, the same word can exist in several different forms. We are not interested in the form of the word. Thus, we applied stemming to the words before applying BLEU

on the sentences. The following examples are retrieved from Neural Machine Translation (seq2seq) Tutorial to visualize the need for stemming [49]. For one of the sentences, reference translation is *"Republican leaders justified their policy by the need to combat electoral fraud ."* and the neural machine translation is *"Republicans are justifying their policy with the need to combat electoral fraud ."*. Machine translated version can be accepted as perfect in the context manner. However, the BLEU score is low because of the difference in the forms of the words. BLEU-1 score with the original translation is 69.23 but the stemmed BLEU-1 score is 84.62.

We applied both BLEU-1 and kn-BLEU with $k = 5$ to the machine translation outputs of the model trained by Luong et al. To compare them logically, we applied stemming to the sentences before getting BLEU-1 score. Our proposed method gives the average score of 64.72 where the BLEU-1 is 53.65. Increased 3 and stayed the same 3 samples are given in Table 6.2. The results are promising in measuring the context relevancy. The new metric is not a suitable tool for detecting the quality of the grammar. However, it gives consequential information about the relatedness to the context over selected words.

Table 6.2: BLEU-1 and kn-BLEU scores on increased 3 and stayed the same 3 samples on news-test-2014 test dataset where k is 5.

Reference Translation	Machine Translation	BLEU-1	kn-BLEU
Manning testified Thursday about his arrest in Iraq and his transfer to Kuwait , where he was held for nearly two months before being transferred to the brig at Marine Base Quantico in Virginia in July 2010 .	On Thursday , his arrest and postponement to Kuwait , where he held almost two months before he was imprisoned in July 2010 , was held in the prison of the navy of Virginia .	53.46	78.13

(continued)

Table 6.2 *(continued)*

Reference Translation	Machine Translation	BLEU-1	kn-BLEU
It was through him that the scandal broke in 2011 , in an in-depth investigation into corruption related to road construction contracts in Quebec , to which the liberal Prime Minister at the time , Jean Charest , had consented only reluctantly .	The scandal was revealed by an in-depth investigation into the untrustworthiness of the contracts on road construction in Quebec , which the then prime minister had reluctantly agreed to .	51.29	75.86
A referendum is due to be held within the next two weeks .	A referendum should be held in 14 days .	35.62	66.67
In accordance with the wishes of the military , the Defence budget review will be not submitted to Parliament , but to a National Defence Council .	In line with the military ' s wish , the review of the defence budget will not be presented to Parliament , but to a national defence council .	85.19	85.19
The shelves are full to capacity .	The shelves are filled up to the top .	55.56	55.56
Increasingly , an unrelieved patient will have the option of having such palliative sedation .	Patients without pain reduction will have an increasingly common ability to back this date .	0.4	0.4

6.4 Flow of Replying Opposite View Tweets

There is an automated flow for searching dissident tweets on Twitter and post a reply to them. This flow focuses on generating highly related replies to the target tweet.

Also, produced tweets should be noncontradictory with its point of view. Namely, if the bot is supposed to be leftwinger, then it should not generate pro-gun content. Our bot collects tweets across Twitter by given tokens. The mentions directed to our bots are also included. It does not interact with all of these tweets.

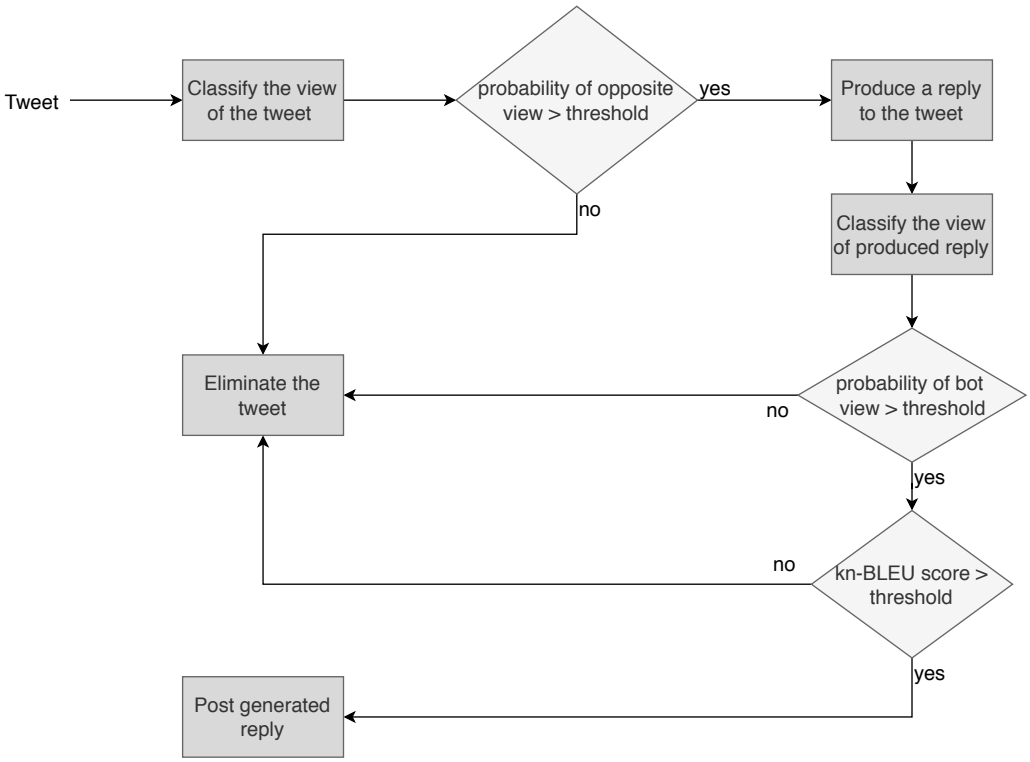

Figure 6.3: The flow of producing replies to the target tweets.

First, it filters them according to the pre-trained classifier that is explained in Chapter 4. If the classifier predicts the label of the given tweet with the probability higher than the given threshold, then it is taken into account. Other tweets are accepted as the bot is not sure and simply ignored. Selection procedure is described in Algorithm

6.1 with a pseudocode.

Algorithm 6.1: Algorithm to choose tweets that should be replied.

```
Process ChooseTweetsToReply(collected_tweets, label_of_opposite,
  threshold):
    selected_tweets ← []
    classifier ← new TweetClassifier()
    for tweet in collected_tweets do
        probabilities ← classifier.classify(tweet)
        if probabilities[label_of_opposite] > threshold then
            selected_tweets.insert(tweet)
        end
    end
    return selected_tweets ;
```

After choosing which tweets should be replied, we feed our model with these tweets to generate replies. The model emits new sequences of words to form tweets and produced tweets need to be checked before getting posted. Our bot checks the generated tweets if it is under the desired quality in both relatedness and the meaning manners. In other words, if the reply is not related to the target tweet, it does not post it. The relatedness check is done with kn-BLEU metric which is described in Section 6.3. Algorithm 6.2 demonstrate this automated process with pseudocode. It gets the tweets to reply and produce a reply to each of them. Generated replies are checked and the ones which passed the relatedness check are returned. Overall flow is visualized in Figure 6.3.

Algorithm 6.2: Algorithm to choose reply among machine produced replies.

```
Process ChooseReplyToPost(selected_tweets, label_of_intended_side,
 label_threshold, kn-BLEU_threshold):
    classifier ← new TweetClassifier()
    replier ← new TweetReplier(label_of_intended_side)
    scorer ← new kn-BLEU_model(label_of_intended_side)
    generated_replies ← []
    for tweet in selected_tweets do
        generated_reply ← replier.generate(tweet)
        score ← scorer.score(tweet, generated_reply)
        probabilities ← classifier.classify(generated_reply)
        if probabilities[label_of_intended_side] > label_threshold then
            if score > kn-BLEU_threshold then
                generated_replies.insert((tweet, generated_reply))
            end
        end
    end
    return generated_replies ;
```

6.5 Experiments

Similar to other models, we used GloVe embeddings which are 200-dimensional vectors and trained with the Twitter data set to symbolize words. Missing words in the GloVe are tagged with special keyword <*unk*>. The architecture we applied is 1 layer LSTM with the size of 256. It consists of two parts, namely encoder, and decoder. Dropout is applied to the cells for preventing memorize the data and adding randomness while sampling. ADAM optimizer with 10^{-2} learning rate is used for training. Two models are trained separately for two views of the debate similar to Chapter 5. Configurations for these models are identical. Learning lasted for 150 epochs. For filtering the produced tweets meaning threshold is fixed to 80% and the kn-BLEU threshold is set to 45. In addition to training, we used dropout for sampling. The dropout rate for sampling is 50%.

The training set and test set contain 250,000 and 65,000 tweets and reply tuple respectively for each model. Produced results showed that the LSTM model can overfit on training data and produce the same replies again. For measuring kn-BLEU of generated replies, we train another model to produce the tweet itself with these 250,000 training set tweets. k of the kn-BLEU is set to 5 for the experiments. 3,000 different tweets are produced for both sides. Before filtering unrelated replies out, we get the average kn-BLEU scores for both models. Table 6.3 shows these results. If the score of a reply is 100, then it is the same with one of k nearest tweets in the training set. It is 0 if any of the words are not included in any of them. The kn-BLEU of generated replies, 44.98 and 41.64 show that the model is able to produce highly related replies to the target tweets.

Table 6.3: Average kn-BLEU scores of generated replies for both sides of the debate with k=5.

	kn-BLEU
Real Right-Winger Replies	43.72
Generated Right-Winger Replies	44.98
Real Left-Winger Replies	44.51
Generated Left-Winger Replies	41.64

Human evaluation is done on generated and real replies like we did in Chapter 5. There are 60 tweets that are randomly chosen from the tweets passed the threshold. Real and fake light-winger and right-winger replies are with equal proportion. Tweets used in the evaluation are listed in Table A.5. *Relatedness* and *Quality* are two scores that we want participants to mark. Scores are in scale between 1-5 where 5 stands for the best. Table 6.4 includes the results of human evaluation. *Relatedness* refers to how the given reply is related to the target tweet. *Quality* is the measure if the generated reply looks like a human-generated one.

Table 6.4: Human evaluation results for generated and real replies.

	Relatedness	Quality
Real Right-Winger Replies	4.24	3.76
Generated Right-Winger Replies	4.07	3.62
Real Left-Winger Replies	4.14	3.56
Generated Left-Winger Replies	3.92	3.43

6.6 Chapter Discussion

In this chapter, we tried to generate replies to the target tweets. To generate original tweets, we used dropout also in the sampling process. It made the model to not follow the accustomed path while predicting the next word. Alternative paths also involve related output to the given input. Produced replies sometimes have opposite meanings. This weakness is overcome with the classifier explained in Chapter 4. The classifier is fed with the generated tweet and if the probability of the desired label passes the threshold, then it is accepted.

Similar to meaning, the relevancy of the reply is crucial. We proposed a new metric for measuring the relatedness of the produced reply. It is based on an existing measure named BLEU which is used for mainly measuring the quality of the neural machine translations. Some deficiencies of BLEU lead us to come up with kn-BLEU. We took the advantage of this metric for filtering the machine-generated replies. If kn-BLEU score of a reply is above the predefined threshold, we acknowledge that it is related to the target tweet. The average kn-BLEU score of the generated replies with a separate test set is relatively high. As displayed in Table 6.3, our model is capable of capturing the context of target tweet and select related words.

We have applied kn-BLEU metric on neural machine translation to be more consistent. The results bring out that it does not quantify the quality of the translation. However, our proposed method measures the relatedness of the chosen words to the context of the source sentences. Table 6.1 shows the success of our approach. In-

creasing the size of the data set retrieves more related content. As future work, new mechanisms that weights the importance of words in the sentence can be included while creating the context space.

Keeping our bot silent is better than posting unrelated or meaningless tweets. Accommodating with real users' Twitter usage, it should not take every tweet or mention into consideration. This ignorance behavior is more human-like. Therefore, two-layer of filtration made the bot able to post content only that is both relevant and supportive to the desired side of the debate.

Human evaluation results reveal that it is successful in the mean of forming sentences. Even the quality of bot tweets are less than the real ones, it has still close scores. These results might be interpreted as the bot replies like real users and it is very difficult to be get disclosed.

6.7 Chapter Conclusion

In this chapter, we developed the most important component of our Twitter bot. Considering the bot as a combination of different components, our Twitter bot is ready for the action after the reply module is ready. We have proposed a new metric for measuring the relatedness of the produced content to the target tweet. Results reveal that the generated replies are mostly relevant and reflect the main idea of the desired side of the debate. We have shown that it is possible to measure the relatedness of two content with the use of LSTM networks. Human evaluators also indicated that the bot is able to produce human-like replies from both meaning and quality perspectives.

CHAPTER 7

CONCLUSION AND FUTURE WORK

In this thesis, we propose multiple models for developing a bot that can join a discussion on Twitter. The bot should mimic real users in several ways. It is capable of capturing the meaning of a tweet in a debate, generating new tweets that reflect the opinion of the side of the debate, reply with the content that is related to the target tweet, and retweet the tweets that support the same opinion.

The first model that is required for our bot is classifier. It gets a tweet and returns the probability of which side it belongs to. The results show that it is capable of predicting the label correctly 74.07% of the time. Narrowing the tweets to take into account the probability threshold increases the accuracy. If the model is sure for the label above 85%, the accuracy leaps up to 89.99%. Retweet operation is done with the help of the classifier model. If the selected tweet reflects the idea above the threshold, the bot decides to retweet it.

The tweet generation model uses LSTM as the main component. Generated tweets mostly reflect the political view of the bot. However, sometimes it gets confused. In order to minimize this deficiency, the classifier model is used as a scoring mechanism for generated tweets. If the tweet does not reflect the idea correctly, it is eliminated and not posted.

The third model that is used for replying includes seq-seq LSTM architecture. It is fed by a tweet that is intended to reply and produces a sequence. Similar to the next-word prediction model, sometimes it generates tweets that do not reflect the idea outrightly or not related to the target tweet. The classifier model is applied to the generated tweet to measure the sentiment. Relatedness issue is handled with our

proposed metric, namely kn-BLEU. If any of the measurements do not exceed the threshold, the generated tweet is eliminated.

The bot performs its tasks with the operation of these three models coherently. It is able to carry out all possible activities that can be done on Twitter, such as posting a tweet, retweeting, replying, or mentioning.

7.1 Future Work

In this thesis, we used GloVe embeddings for word representations. Other methods and approaches can be tested to check whether it is the best way to represent vectors or not. Deep contextualized word representations might be selected for encoding the tweets [50]. Furthermore, generated tweets contain some special tags, such as <*number*>, <*hashtag*>, <*user*>, etc. Filling these blanks with logical matching is crucial. Bidirectional-LSTMs might be applied to the tweets for this purpose.

The accounts in our dataset generally tweets about the gun debate. However, tweeting about different topics may lower the probability of being disclosed. Converting the bot from closed-domain to open-domain might be a challenging but impressive study. Even though we use human evaluation to determine whether our bot can be easily disclosed or not, the best way is to see what happens in the action. Therefore, the bot can be released to Twitter and the interaction it gets can be collected.

REFERENCES

[1] T. Rosenstiel, J. Sonderman, K. Loker, M. Ivancin, and N. Kjarval, "Twitter and the news: How people use the social network to learn about the world.," *American Press Institute*, 2015.

[2] O. Varol, E. Ferrara, C. A. Davis, F. Menczer, and A. Flammini, "Online human-bot interactions: Detection, estimation, and characterization," in *Eleventh international AAAI conference on web and social media*, 2017.

[3] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543, 2014.

[4] I. L. Liu, C. M. Cheung, and M. K. Lee, "Understanding twitter usage: What drive people continue to tweet.," *Pacis*, vol. 92, pp. 928–939, 2010.

[5] L. Rook, "An economic psychological approach to herd behavior," *Journal of Economic Issues*, vol. 40, no. 1, pp. 75–95, 2006.

[6] M. Forelle, P. Howard, A. Monroy-Hernández, and S. Savage, "Political bots and the manipulation of public opinion in venezuela," *arXiv preprint arXiv:1507.07109*, 2015.

[7] P. Howard, B. Kollanyi, and S. C. Woolley, "Bots and automation over twitter during the second us presidential debate," 2016.

[8] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach," *IEEE transactions on neural networks*, vol. 8, no. 1, pp. 98–113, 1997.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[10] J. Park and I. W. Sandberg, "Universal approximation using radial-basis-function networks," *Neural computation*, vol. 3, no. 2, pp. 246–257, 1991.

[11] S. Lai, L. Xu, K. Liu, and J. Zhao, "Recurrent convolutional neural networks for text classification," in *Twenty-ninth AAAI conference on artificial intelligence*, 2015.

[12] J. Weizenbaum *et al.*, "Eliza—a computer program for the study of natural language communication between man and machine," *Communications of the ACM*, vol. 9, no. 1, pp. 36–45, 1966.

[13] M. Kochen, "On the generality of parry, colby's paranoia model," *Behavioral and Brain Sciences*, vol. 4, no. 4, pp. 540–541, 1981.

[14] H. Al-Zubaide and A. A. Issa, "Ontbot: Ontology based chatbot," in *International Symposium on Innovations in Information and Communications Technology*, pp. 7–12, IEEE, 2011.

[15] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[16] F. Karim, S. Majumdar, H. Darabi, and S. Chen, "Lstm fully convolutional networks for time series classification," *IEEE Access*, vol. 6, pp. 1662–1669, 2017.

[17] T.-H. Wen, M. Gašić, N. Mrkšić, P.-H. Su, D. Vandyke, and S. Young, "Semantically conditioned LSTM-based natural language generation for spoken dialogue systems," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, (Lisbon, Portugal), pp. 1711–1721, Association for Computational Linguistics, Sept. 2015.

[18] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances in neural information processing systems*, pp. 3104–3112, 2014.

[19] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in neural information processing systems*, pp. 3111–3119, 2013.

[20] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135–146, 2017.

[21] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," *arXiv preprint arXiv:1409.1259*, 2014.

[22] A. Turing, "Mind," *Mind*, vol. 59, no. 236, pp. 433–460, 1950.

[23] R. Epstein, "The quest for the thinking computer," *AI magazine*, vol. 13, no. 2, pp. 81–81, 1992.

[24] L. Bradeško and D. Mladenić, "A survey of chatbot systems through a loebner prize competition," in *Proceedings of Slovenian Language Technologies Society Eighth Conference of Language Technologies*, pp. 34–37, 2012.

[25] Y. Wilks, *Close engagements with artificial companions: key social, psychological, ethical and design issues*, vol. 8. John Benjamins Publishing, 2010.

[26] R. Wallace, "The elements of aiml style," *Alice AI Foundation*, vol. 139, 2003.

[27] B. Wilcox, S. Wilcox, B. Psych, and D. F. Arts, "Suzette, the most human computer," *Agent's Processing, Cognition: https://www. chatbots. org/images/uploads/research_papers/9491. p df*, 2010.

[28] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, *et al.*, "Google's neural machine translation system: Bridging the gap between human and machine translation," *arXiv preprint arXiv:1609.08144*, 2016.

[29] H.-Y. Shum, X.-d. He, and D. Li, "From eliza to xiaoice: challenges and opportunities with social chatbots," *Frontiers of Information Technology & Electronic Engineering*, vol. 19, no. 1, pp. 10–26, 2018.

[30] R. Sarikaya, "The technology behind personal digital assistants: An overview of the system architecture and key components," *IEEE Signal Processing Magazine*, vol. 34, no. 1, pp. 67–81, 2017.

[31] G. Neff and P. Nagy, "Automation, algorithms, and politics| talking to bots: Symbiotic agency and the case of tay," *International Journal of Communication*, vol. 10, p. 17, 2016.

[32] P. Indyk and R. Motwani, "Approximate nearest neighbors: towards removing the curse of dimensionality," in *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, pp. 604–613, ACM, 1998.

[33] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain.," *Psychological review*, vol. 65, no. 6, p. 386, 1958.

[34] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *nature*, vol. 323, no. 6088, p. 533, 1986.

[35] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *Journal of machine learning research*, vol. 11, no. Dec, pp. 3371–3408, 2010.

[36] T. Mikolov, M. Karafiát, L. Burget, J. Černockỳ, and S. Khudanpur, "Recurrent neural network based language model," in *Eleventh annual conference of the international speech communication association*, 2010.

[37] S. Hochreiter and J. Schmidhuber, "Lstm can solve hard long time lag problems," in *Advances in neural information processing systems*, pp. 473–479, 1997.

[38] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with lstm," 1999.

[39] V. Peirson, L. Abel, and E. M. Tolunay, "Dank learning: Generating memes using deep neural networks," *arXiv preprint arXiv:1806.04510*, 2018.

[40] Y. Wang, M. Huang, L. Zhao, *et al.*, "Attention-based lstm for aspect-level sentiment classification," in *Proceedings of the 2016 conference on empirical methods in natural language processing*, pp. 606–615, 2016.

[41] M. Sundermeyer, R. Schlüter, and H. Ney, "Lstm neural networks for language modeling," in *Thirteenth annual conference of the international speech communication association*, 2012.

[42] B. T. Lowerre, "The harpy speech recognition system," tech. rep., CARNEGIE-MELLON UNIV PITTSBURGH PA DEPT OF COMPUTER SCIENCE, 1976.

[43] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "Bleu: a method for automatic evaluation of machine translation," in *Proceedings of the 40th annual meeting on association for computational linguistics*, pp. 311–318, Association for Computational Linguistics, 2002.

[44] C.-Y. Lin, "Rouge: A package for automatic evaluation of summaries," in *Text summarization branches out*, pp. 74–81, 2004.

[45] R. Rosenfeld, "Two decades of statistical language modeling: Where do we go from here?," *Proceedings of the IEEE*, vol. 88, no. 8, pp. 1270–1278, 2000.

[46] C. D. Manning, "Part-of-speech tagging from 97% to 100%: is it time for some linguistics?," in *International conference on intelligent text processing and computational linguistics*, pp. 171–189, Springer, 2011.

[47] D. Wang and E. Nyberg, "A long short-term memory model for answer sentence selection in question answering," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pp. 707–712, 2015.

[48] M.-T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," *arXiv preprint arXiv:1508.04025*, 2015.

[49] M. Luong, E. Brevdo, and R. Zhao, "Neural machine translation (seq2seq) tutorial," *https://github.com/tensorflow/nmt*, 2017.

[50] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," *arXiv preprint arXiv:1802.05365*, 2018.

APPENDIX A

GENERATED TWEETS

Table A.1: Tweets used in Meaning and Quality test explained in Chapter 5.

Tweet	Tweet Owner
Q1) these are the good mentally stable people that have ar <number> s <url> <end>	Leftwinger / Democrat / Anti-Gun
Q2) we have to know the details to prevent this from happening again . and hows that working for ya ? did knowing he was an orphan suddenly clear things up for you ? you dont care about the signs you want excuses . <hashtag> <end>	Leftwinger / Democrat / Anti-Gun
Q3) these kids are being used for political gain and are being told exactly what to say ! ! said by the same people who retweet russian troll propaganda . <hashtag> <hashtag> <end>	Leftwinger / Democrat / Anti-Gun
Q4) these kids are destroying their life . may not new and having a beat . just think of the core savvy women can make every usual boil on <hashtag> <url>	Leftwinger / Democrat / Anti-Gun
Q5) its time to change lobbies bent = so important rather than saving our children are <hashtag> <url>	Leftwinger / Democrat / Anti-Gun
Q6) its time to arm our teachers is literally the dumbest shit ive heard . you dont fight to end gun violence with more guns . <hashtag> <end>	Leftwinger / Democrat / Anti-Gun
Q7) it is time for the victims to be the change we need to see emma gonzalez <hashtag> these teens are inspirationalin the wake of tragedy they have been focused eloquent and determined rattling the nation out of its complacency and apathy on the issue of gun control . <url> <end>	Leftwinger / Democrat / Anti-Gun

(continued)

Table A.1 *(continued)*

Tweet	Tweet Owner
Q8) we have to be strong on another extremist in the world by bribing senators for $. <hashtag> <url>	Leftwinger / Democrat / Anti-Gun
Q9) these are the people working for gun manufacturer supporters not republican willing and allows nra and politicians to sell their toys so their lives are yours ! <hashtag> <hashtag> <hashtag>	Leftwinger / Democrat / Anti-Gun
Q10) we need to stop these gun lobby i pledge . have kids the rest of the amazing friends voices that are tone please watch this <url>	Leftwinger / Democrat / Anti-Gun
Q11) we need to give teachers guns to stop ppl with guns ! is akin to we need to teach girls not to be raped so ppl cant rape ! both statements entirely overlook the root of the problems & the victims of violence should not be responsible for preventing that violence . <hashtag> <end>	Leftwinger / Democrat / Anti-Gun
Q12) gun control is too hard <hashtag> rains hell on for blaming school shootings on <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <url> <url> <end>	Leftwinger / Democrat / Anti-Gun
Q13) i do not need this rifle this man gave up his ar <number> style gun because of the florida school shooting buzzfeed news <url> <end>	Leftwinger / Democrat / Anti-Gun
Q14) gun control is accountable . email <user> that nra bribe tap dance about kids violence is hush money . <url>	Leftwinger / Democrat / Anti-Gun
Q15) it is time to revisit my grandstanding to hide about peoples families over school assault . weed in school get robbed public speak of this hat too . <url>	Leftwinger / Democrat / Anti-Gun
Q16) i do not . from the damage the nra lines your pockets they do not even have empty words . <hashtag> <hashtag> <hashtag>	Leftwinger / Democrat / Anti-Gun
Q17) i think it is actually a new liberal defend trump except keep fighting patriots <hashtag>	Rightwinger / Republican / Pro-Gun

(continued)

Table A.1 *(continued)*

Tweet	Tweet Owner
Q18) i love how these celebrities come to twitter make a shit storm with their activist tweet and then never get involved in the thread ? what the hell ? <hashtag> <hashtag> <hashtag> <end>	Rightwinger / Republican / Pro-Gun
Q19) i think it s a great time for the <user> to think about expanding into the car rental business . <hashtag> <end>	Rightwinger / Republican / Pro-Gun
Q20) you are a liberal useful idiot off the wrong side of the left is false . <hashtag>	Rightwinger / Republican / Pro-Gun
Q21) this is a bizarre interview she claims she was talking to the suspect while gun shots were still happening else where in the school and at times is smiling . <url> <end>	Rightwinger / Republican / Pro-Gun
Q22) you re a leftist haters while plausible at all day so quickly again ! <hashtag> <user> <hashtag>	Rightwinger / Republican / Pro-Gun
Q23) i love how fascists & the second amendment does i thought it was <hashtag> but upped their communist $ for <hashtag> to <hashtag>	Rightwinger / Republican / Pro-Gun
Q24) you are a truly american young man we admired you and we wish we have more young ppl like you ! and not opportunists all over the fake news taking advantage of the horrible situation ! <url> <end>	Rightwinger / Republican / Pro-Gun
Q25) broward county sheriff boasts about preserving our constitution ! god bless our <hashtag> <hashtag> <hashtag>	Rightwinger / Republican / Pro-Gun
Q26) you re a puppet of the sick left . you can not disarm the american people . you have no right . guns are not the problem . people are the problem . you do not stop these people without the proper tool . that tool is a gun . <hashtag> <hashtag> <hashtag> is dishonest <url> <end>	Rightwinger / Republican / Pro-Gun
Q27) i wonder if <user> places within their contracts that you must be a liar to obtain the position at the network fake news cnn s chris cuomo pushes false story about buying ar <number> without id then lies about it <url> <end>	Rightwinger / Republican / Pro-Gun
Q28) broward county sheriff scott israel represents what american police culture looks like by <user> <hashtag> <url> <end>	Rightwinger / Republican / Pro-Gun

(continued)

Table A.1 *(continued)*

Tweet	Tweet Owner
Q29) i m not to own ar <number> shall not the militia shall not be infringed . <hashtag> can qual can be educated on the <hashtag> push <hashtag>	Rightwinger / Republican / Pro-Gun
Q30) this is a real airline place scary she most would love to get used in concerning place <user> inciting violence upon leftist liberals <hashtag>	Rightwinger / Republican / Pro-Gun
Q31) i m not sure this person fully grasps what guns are supposed to do . <hashtag> <url> <end>	Rightwinger / Republican / Pro-Gun
Q32) i wonder if the sheriff nation is divided and imprisoned in a government <number> sole outcome for their recent issues only make work for liberals . <hashtag>	Rightwinger / Republican / Pro-Gun
Q33) assault weapons are saint of the second . oregon overrated georgia gov . smith & wesson . im complicit . <url>	Leftwinger / Democrat / Anti-Gun
Q34) i think the gop came through . also gun lovers remind them of young leaders .	Leftwinger / Democrat / Anti-Gun
Q35) guns do not belong in the hands of civilians they re power combined with their kk views . maybe we have to deal with the text of those of us <number> m+ lifes and i ll happily scream about guns .	Leftwinger / Democrat / Anti-Gun
Q36) assault weapons are so misunderstood is the not all men of the gun debate . <url> <end>	Leftwinger / Democrat / Anti-Gun
Q37) the right to own an assault rifle is not more important than one life . says a gun owner <url> <end>	Leftwinger / Democrat / Anti-Gun
Q38) good guy with a gun to the rescue once again ! <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <number> . <url> <number> . <url> <url> <end>	Leftwinger / Democrat / Anti-Gun
Q39) this is a community & amp a state that is in deep pain & amp they want action to make sure this doesnt happen again . <url> <end>	Leftwinger / Democrat / Anti-Gun

(continued)

Table A.1 *(continued)*

Tweet	Tweet Owner
Q40) good guy with a simple gun contest . and their empty fake tears get thought anytime . lands offshore and backing the top are stone go ! ! maga ! ! !	Leftwinger / Democrat / Anti-Gun
Q41) guns do not kill people do guns do not kill people with guns do <hashtag> <end>	Leftwinger / Democrat / Anti-Gun
Q42) we the people have to do it if <hashtag> and <user> the disgusting human offal wont <hashtag> <url> <end>	Leftwinger / Democrat / Anti-Gun
Q43) a bunch of nope . <hashtag> <url> <end>	Leftwinger / Democrat / Anti-Gun
Q44) a bunch of make money . fuck you . money . kids over guns . <hashtag> <hashtag>	Leftwinger / Democrat / Anti-Gun
Q45) this is a shameful idea . lobbyist aligning with and making sure that children r shame ! <url>	Leftwinger / Democrat / Anti-Gun
Q46) the right to confront the old white male terrorist organization . ar <number> differ from a bear . the only guarantee to kids and others	Leftwinger / Democrat / Anti-Gun
Q47) i think the best way to deal with this is to ignore him . the things that pres trump tweets is nothing that will have a lasting impact unless its a negative lasting impact on the people around us . <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <end>	Leftwinger / Democrat / Anti-Gun
Q48) we the people we re gon na represent ! ! you vote poc & people get blown away ! yay this wild west that d die .	Leftwinger / Democrat / Anti-Gun
Q49) these kids are engaging in <hashtag> circus liberals have different views can damage we learn about guns/ stupid or dangerous . <url>	Rightwinger / Republican / Pro-Gun
Q50) i agree . shall not be infringed ! there s no necessary clause here s the suggestion for supporting ones stay . <hashtag>	Rightwinger / Republican / Pro-Gun

(continued)

Table A.1 *(continued)*

Tweet	Tweet Owner
Q51) i dont think that shall not be infringed the second amendment does not exist in the world where the constitution is important . <url>	Rightwinger / Republican / Pro-Gun
Q52) why is this purpose but thanks you and the globalists armed men of children to school killer or not <hashtag> instead of <hashtag> protect em ! <hashtag>	Rightwinger / Republican / Pro-Gun
Q53) i will not relinquish my <hashtag> right . fascist <hashtag> people need to stop attempting to diminish my birthright . an armed citizenry doesnt board the trains to concentration camps . <url> <end>	Rightwinger / Republican / Pro-Gun
Q54) it s time for your meds ray . nurse ratched is waiting <hashtag> <hashtag> <hashtag> <url> <end>	Rightwinger / Republican / Pro-Gun
Q55) i want to remove self freedom . <hashtag> deletes <hashtag> question . lordy ! <hashtag> what ? do u s think ? <hashtag>	Rightwinger / Republican / Pro-Gun
Q56) why is this an either/or choice ? <url> <end>	Rightwinger / Republican / Pro-Gun
Q57) it s time ! <hashtag> <url> jerome corsi aux le agencies are corrupt ! <hashtag>	Rightwinger / Republican / Pro-Gun
Q58) these kids are just kids and the shallowness & deceptive nature of their advocacy is illustrative of that . <url> <end>	Rightwinger / Republican / Pro-Gun
Q59) i dont think he even knows what an assault rifle is <end>	Rightwinger / Republican / Pro-Gun
Q60) do you know no gun shall infringe <number>. protects gun rights private ownership of individuals violating individual freedoms . <hashtag>	Rightwinger / Republican / Pro-Gun
Q61) i will not that <number> rep government carries <hashtag> gun guide any man child should remember so other leftist group hoodlums would have permits like them ! <hashtag> <url>	Rightwinger / Republican / Pro-Gun

(continued)

Table A.1 *(continued)*

Tweet	Tweet Owner
Q62) do you know about the act of <number> teach this in your school <url> via <user> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> or you can go watch <hashtag> that should save your country <hashtag> or <hashtag> is anybody still freaking doubting they lie ? <end>	Rightwinger / Republican / Pro-Gun
Q63) i want to know if any useless russian s trollers knew <hashtag> or <user> <user> <user> <hashtag> <user> <hashtag> <user> <user> <user> <hashtag> <end>	Rightwinger / Republican / Pro-Gun
Q64) i agree . i support the nra too . <url> <end>	Rightwinger / Republican / Pro-Gun
Q65) we do not need comfort . we need change <hashtag> <hashtag> <hashtag> <user> <user> <end>	Leftwinger / Democrat / Anti-Gun
Q66) we do not after paul brightest hope for a classroom in school . you re common factor . who is necessary to achieve the issue of kids s social welfare basis ? <number> nd amendment would be diminished enough ? its entertainment !	Leftwinger / Democrat / Anti-Gun
Q67) this is not a republican or a democrat thing you are either with us or against us we are giving every politician a clean slate if you are accepting money from the nra there is a badge of shame on you . this was fox news sunday ! <hashtag> <hashtag> <hashtag> <url> <end>	Leftwinger / Democrat / Anti-Gun
Q68) the fact that suffers gun violence from conception to mass slaughter . <hashtag> <hashtag> <hashtag>	Leftwinger / Democrat / Anti-Gun
Q69) why do you say youre from brooklyn and not pennsylvania ? <url> <end>	Leftwinger / Democrat / Anti-Gun
Q70) the nra has spent republicans in the pockets of <hashtag> to start caring about you !	Leftwinger / Democrat / Anti-Gun
Q71) the only thing necessary for the triumph of evil is for good people to do nothing . edmund burke essentially fuck your thoughts and prayers . <hashtag> <end>	Leftwinger / Democrat / Anti-Gun

(continued)

Table A.1 *(continued)*

Tweet	Tweet Owner
Q72) this is not too just sort of validated human lives including children in your classroom . <hashtag> the <hashtag> <hashtag> is listening .	Leftwinger / Democrat / Anti-Gun
Q73) i am so proud of all the students at my school sarah lerner teacher at marjory stoneman douglas high school . <hashtag> <end>	Leftwinger / Democrat / Anti-Gun
Q74) why do you think <number> yearold white male you can survive an increase in the effectiveness of these gun murders ? <url>	Leftwinger / Democrat / Anti-Gun
Q75) i want to worship them but an assault weapon is tone deaf and disrespectful . id say you must come out here	Leftwinger / Democrat / Anti-Gun
Q76) the nra has a history of supporting the civil rights movement . nra armed and trained blacks to defend against democrat kk ! <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <url> <url> <end>	Leftwinger / Democrat / Anti-Gun
Q77) the fact that some of the students are showing more maturity and political action than many of our elected officials is a testament to how disgusting and broken our political system is right now in america . but were trying to fix that . amen <url> <end>	Leftwinger / Democrat / Anti-Gun
Q78) i am so ashamed of sane people about us enough . president forced a visit to <hashtag> <url>	Leftwinger / Democrat / Anti-Gun
Q79) the only thing <user> backwards republican listening session is . <hashtag> what would have happened if he was murdered <url>	Leftwinger / Democrat / Anti-Gun
Q80) i want to fight for this country . im willing to die for it . do the right thing . its unjust & irresponsible to let <hashtag> who are willing to risk their lives for this country remain in limbo without <hashtag> we need to act now to help these <hashtag> <url> <end>	Leftwinger / Democrat / Anti-Gun
Q81) if you are an nra member who believes in the constitution i will follow all ! we must join together to fight the left ! semper fi ! <url> <end>	Rightwinger / Republican / Pro-Gun

(continued)

Table A.1 *(continued)*

Tweet	Tweet Owner
Q82) you need to be taken to us we must senate / the previous leftist and <hashtag> office is very govt without power bc of my opinion answers ! <user>	Rightwinger / Republican / Pro-Gun
Q83) take the guns first and go through due process second ? no no no no no . that s not how it works . <end>	Rightwinger / Republican / Pro-Gun
Q84) if you are not disarmed they would never use your box police to be protected . <hashtag> <url>	Rightwinger / Republican / Pro-Gun
Q85) let s get to the truth.the ar <number> is not an assault rifle . it s the name the left has given it because of its profile . it is a semiautomatic rifle like any other clip or magazine fed rifle is . u must pull the trigger each time to fire a round you can not hold the trigger dn . <end>	Rightwinger / Republican / Pro-Gun
Q86) you need to listen to all the students stop the far left liberal attacks this pushing a democratic narrative wo not change laws <hashtag> <url> <end>	Rightwinger / Republican / Pro-Gun
Q87) what about the men who landed on normandy beach ? did they say no we re not going to get off the boat because they re shooting big guns at us ? <url> <end>	Rightwinger / Republican / Pro-Gun
Q88) i ve been thinking the exact same thing <hashtag> <hashtag> <hashtag> or just some very cold <hashtag> <hashtag> <hashtag> <hashtag> <url> <end>	Rightwinger / Republican / Pro-Gun
Q89) i ve been left to understand why we need <hashtag> and not be infringed upon . take guns away from them to arm themselves !	Rightwinger / Republican / Pro-Gun
Q90) what is the right matters with dismay america supports their feelings on gun free zones they say and do not exercise their ability to protect themselves . <hashtag> <user> <hashtag> <url>	Rightwinger / Republican / Pro-Gun
Q91) let s get out off the msm & standing up for our rights sit <hashtag>	Rightwinger / Republican / Pro-Gun
Q92) take the guns of the leftist attempts to own a gun show will use more <hashtag> facilities responsibly . <hashtag>	Rightwinger / Republican / Pro-Gun

(continued)

Table A.1 *(continued)*

Tweet	Tweet Owner
Q93) what about the same premise is the broward law & failures in the <user> with to this issue alex harris this is how evil lies ignoring ! ! <url>	Rightwinger / Republican / Pro-Gun
Q94) what is the average police response time in you area ? <end>	Rightwinger / Republican / Pro-Gun
Q95) make no mistake . the <hashtag> want to disarm us . that is was and always will be their endgame and they dont care how they get there . <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <url> <end>	Rightwinger / Republican / Pro-Gun
Q96) make no mistake folks will fight all i will . <user> hold back up program education = suspension ! that s important ! expose them <hashtag>	Rightwinger / Republican / Pro-Gun
Q97) i think it is time for us to have a conversation about what the right to bear arms means in the modern world rice said . i dont understand why civilians need to have access to military weapons . <hashtag> <hashtag> <url> <end>	Leftwinger / Democrat / Anti-Gun
Q98) it s the adults and the <hashtag> they are humankind im sorry for love for those active thoughts and prayers not well give this s so much . they will do ! ! the kids or ireland has become much more powerful you have to represent americas future . we must see this heroic movement . might not .	Leftwinger / Democrat / Anti-Gun
Q99) it s not be like the one day at when in november to send to chad low and faithful servant . its like the rest of us have a voice and shake the stand . that we should do nothing . it matters . maybe the <user> starting to do it ? they do and qualify . unwise <user> did give everything you have a responsibility to rally for keeping the <number> million . <hashtag>	Leftwinger / Democrat / Anti-Gun
Q100) this is what democracy looks like students lobby and organize in tallahassee urging lawmakers to act on gun control . <url> <hashtag> <hashtag> <end>	Leftwinger / Democrat / Anti-Gun
Q101) do not be smearing of future children from a computer . afraid to faux news it represents a racist not playing at dinner	Leftwinger / Democrat / Anti-Gun
Q102) this is what s the nras affiliated terrorist plot voice <hashtag> <hashtag> <hashtag>	Leftwinger / Democrat / Anti-Gun

(continued)

Table A.1 *(continued)*

Tweet	Tweet Owner
Q103) there is a tremendous difference in the amount of energy associated with a bullet from an ar <number> rifle & amp that of a handgun with a rifle round you have massive tissue disruption <url> <end>	Leftwinger / Democrat / Anti-Gun
Q104) it s the children the world almost breaks who grow up to save it . frank warren <hashtag> <end>	Leftwinger / Democrat / Anti-Gun
Q105) it s not so much staying alive it s staying human that s important . what counts is that we do not betray each other . orwell <number> <hashtag> <url> <end>	Leftwinger / Democrat / Anti-Gun
Q106) do not be a dope vote like knope . fuck the nra shirt available at <url> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> art by <user> <url> <end>	Leftwinger / Democrat / Anti-Gun
Q107) the problem is not the guns its the people so lets help the people . adorable healthcare ? nah we dont give handouts they need a job so lets help them find jobs nah they can do that on their own im tired . <hashtag> <hashtag> <end>	Leftwinger / Democrat / Anti-Gun
Q108) i think it takes good delivery no bills i thought <user> which runs ring <url> via <user> <url>	Leftwinger / Democrat / Anti-Gun
Q109) i can not believe it should be blind & diseased . by sensible action . bend one pair & help wont be able to even call it by kids that still stfu & human talk !	Leftwinger / Democrat / Anti-Gun
Q110) the problem is the primary channel but when paranoia fears insults through the world have fun playing field . <url>	Leftwinger / Democrat / Anti-Gun
Q111) i can not accept that their right to enjoy their hobby supersedes my right to send my own children to school to a movie theater or to a concert and to know that they are safe . <url> <end>	Leftwinger / Democrat / Anti-Gun
Q112) there is a gun . backwards it has been working in way of a gun fantasy . <hashtag>	Leftwinger / Democrat / Anti-Gun
Q113) amen ! ! <hashtag> is lobbying with globalists groups punishing <hashtag> members too . you need a little out <number> <hashtag>	Rightwinger / Republican / Pro-Gun

(continued)

Table A.1 *(continued)*

Tweet	Tweet Owner
Q114) i have never happens indifferent to me . in the long time liberals destroy people from choosing to defend themselves against children from evil . <url>	Rightwinger / Republican / Pro-Gun
Q115) amen ! ! ! ! no due process with the demorats ! ! ! <url> <end>	Rightwinger / Republican / Pro-Gun
Q116) i can not like to come nor will reasonably infringed anyone educated with any other gun . <url>	Rightwinger / Republican / Pro-Gun
Q117) i can not even believe what i m reading on the boards right now . this is seriously fucked . freedom is ours and that includes body spirit and mind ! from darkness into light where we go one we go all . <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <end>	Rightwinger / Republican / Pro-Gun
Q118) is not it a symbolic leftist antifa etc ? <hashtag> <url>	Rightwinger / Republican / Pro-Gun
Q119) fake news cnn pushes boycotts first <url> <hashtag> <user> <url>	Rightwinger / Republican / Pro-Gun
Q120) fake news cnn dope chris cuomo savaged after promoting fake ar <number> story <url> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <end>	Rightwinger / Republican / Pro-Gun
Q121) i have never murdered anyone but i own guns i do want to be able to protect myself ! ! i value all my rights ! <url> <end>	Rightwinger / Republican / Pro-Gun
Q122) how can you advocate for the <hashtag> when you are trying to boycott the leading advocate of the <number> nd amendment the. <user> <hashtag> <url> <end>	Rightwinger / Republican / Pro-Gun
Q123) of course they theoretical only well citizens owners would resist their tyrannical deeds . nukes call them constitution for freedom ! <url>	Rightwinger / Republican / Pro-Gun
Q124) we ca not even trust to import government & communists or fascists how many americans are doing mao pol for <hashtag> and ones .	Rightwinger / Republican / Pro-Gun

(continued)

Table A.1 *(continued)*

Tweet	Tweet Owner
Q125) we ca not have a debate about guns if liberals keep lying about them <url> <hashtag> <hashtag> <end>	Rightwinger / Republican / Pro-Gun
Q126) how can you tell that commie gunrunner ? a full boycott stuff ! youve made my big mistake ! you are being used . thanks .	Rightwinger / Republican / Pro-Gun
Q127) of course they did thats what they do best <url> <end>	Rightwinger / Republican / Pro-Gun
Q128) is not it odd that liberals only seem to care about the loss of life when it s committed by a gun ? and my thoughts and prayers are with the victims and their families and friends ! <hashtag> <end>	Rightwinger / Republican / Pro-Gun

Table A.2: Belongings of tweets used in Meaning and Quality test explained in Chapter 5.

Q1)	Real User	Q2)	Real User	Q3)	Real User	Q4)	Bot
Q5)	Bot	Q6)	Real User	Q7)	Real User	Q8)	Bot
Q9)	Bot	Q10)	Bot	Q11)	Real User	Q12)	Real User
Q13)	Real User	Q14)	Bot	Q15)	Bot	Q16)	Bot
Q17)	Bot	Q18)	Real User	Q19)	Real User	Q20)	Bot
Q21)	Real User	Q22)	Bot	Q23)	Bot	Q24)	Real User
Q25)	Bot	Q26)	Real User	Q27)	Real User	Q28)	Real User
Q29)	Bot	Q30)	Bot	Q31)	Real User	Q32)	Bot
Q33)	Bot	Q34)	Bot	Q35)	Bot	Q36)	Real User
Q37)	Real User	Q38)	Real User	Q39)	Real User	Q40)	Bot
Q41)	Real User	Q42)	Real User	Q43)	Real User	Q44)	Bot
Q45)	Bot	Q46)	Bot	Q47)	Real User	Q48)	Bot
Q49)	Bot	Q50)	Bot	Q51)	Bot	Q52)	Bot

(continued)

Table A.2 *(continued)*

Q53)	Real User	Q54)	Real User	Q55)	Bot	Q56)	Real User
Q57)	Bot	Q58)	Real User	Q59)	Real User	Q60)	Bot
Q61)	Bot	Q62)	Real User	Q63)	Real User	Q64)	Real User
Q65)	Real User	Q66)	Bot	Q67)	Real User	Q68)	Bot
Q69)	Real User	Q70)	Bot	Q71)	Real User	Q72)	Bot
Q73)	Real User	Q74)	Bot	Q75)	Bot	Q76)	Real User
Q77)	Real User	Q78)	Bot	Q79)	Bot	Q80)	Real User
Q81)	Real User	Q82)	Bot	Q83)	Real User	Q84)	Bot
Q85)	Real User	Q86)	Real User	Q87)	Real User	Q88)	Real User
Q89)	Bot	Q90)	Bot	Q91)	Bot	Q92)	Bot
Q93)	Bot	Q94)	Real User	Q95)	Real User	Q96)	Bot
Q97)	Real User	Q98)	Bot	Q99)	Bot	Q100)	Real User
Q101)	Bot	Q102)	Bot	Q103)	Real User	Q104)	Real User
Q105)	Real User	Q106)	Real User	Q107)	Real User	Q108)	Bot
Q109)	Bot	Q110)	Bot	Q111)	Real User	Q112)	Bot
Q113)	Bot	Q114)	Bot	Q115)	Real User	Q116)	Bot
Q117)	Real User	Q118)	Bot	Q119)	Bot	Q120)	Real User
Q121)	Real User	Q122)	Real User	Q123)	Bot	Q124)	Bot
Q125)	Real User	Q126)	Bot	Q127)	Real User	Q128)	Real User

Table A.3: Disclosure tests of generated Tweets explained in Chapter 5.

	Choices	Tweet Owner
Q1)	(a) we need to stop these gun lobby i pledge . have kids & the rest of the amazing friends voices that are tone please watch this <url>	Leftwinger / Democrat / Anti-Gun
	(b) we need to give teachers guns to stop ppl with guns ! is akin to we need to teach girls not to be raped so ppl cant rape ! both statements entirely overlook the root of the problems & amp the victims of violence should not be responsible for preventing that violence . <hashtag> <end>	
	(c) we need to arm our teachers with more money in their pockets . robert runcie <hashtag> <hashtag> <hashtag> <hashtag> <end>	
	(d) we need to tackle the difficult issue of mental health . no . weve all been saying this for years . what about guns ? did he even use the words guns in his speech ? what is your plan ? <hashtag> <hashtag> <end>	
Q2)	(a) it s time for politicians to do something about guns . <url> <hashtag> <hashtag> <hashtag> <hashtag> <end>	Leftwinger / Democrat / Anti-Gun
	(b) it s time corporations publicly make a choice between what they value more the <hashtag> or the lives of our children . <hashtag> <hashtag> <hashtag> <hashtag> <url> <end>	
	(c) it s time to enact <hashtag> . mourning and <number> dead floridians want congress to act to enact <hashtag> bipartisan legislation . now show <hashtag> <hashtag>	
	(d) it s time to do something american ceos are getting bolder on guns <url> <hashtag> <end>	
Q3)	(a) these are the darkest days in at least half a year [aides] say and they worry just how much farther pres . trump and his administration may plunge into unrest and malaise before they start to recover . as one official put it we havent bottomed out . <url> <end>	Leftwinger / Democrat / Anti-Gun
	(b) these are the victims of the florida high school shooting <url> <url> <end>	
	(c) these are the people working for gun manufacturer supporters not republican willing and allows nra and politicians to sell their toys so their lives are yours ! <hashtag> <hashtag> <hashtag>	
	(d) these are the good mentally stable people that have ar <number> s <url> <end>	

(continued)

Table A.3 *(continued)*

	Choices	Tweet Owner
Q4)	(a) it is not too soon to talk about gun safety . its too late . its too late for the seventeen who have lost that were lost . <user> <hashtag> <end> (b) it is not too soon to talk about gun control it is too late because of the <number> people that lost their lives . <hashtag> <hashtag> <user> <end> (c) it is not too soon it is too late <hashtag> <end> (d) it is not great . the nra finances nra is supposed to speak out & devin nunes and last year n says republicans have wasted $ <number> <hashtag> <number> <hashtag> <hashtag> <hashtag>	Leftwinger / Democrat / Anti-Gun
Q5)	(a) these kids are being used for political gain and are being told exactly what to say ! ! said by the same people who retweet russian troll propaganda . <hashtag> <hashtag> <end> (b) these kids are gon na to save us all . amen to that ! ! <url> <end> (c) these kids are destroying their life . may not new and having a beat . just think of the core savvy women can make every usual boil on <hashtag> <url> (d) these kids are such great speakers too ! they re amazing ! follow <user> join us nationwide april <number> th may june however long it takes because we re not settling for a token change in gun controls ! ! ! <hashtag> <hashtag> <url> <end>	Leftwinger / Democrat / Anti-Gun
Q6)	(a) this is the order to ban assault rifles & tillis and they will never have access to such a weapon ! <hashtag> (b) this is the price of freedom . it seems more like bondage . <hashtag> shooting <hashtag> <end> (c) this is the way i have to grieve . i have to make sure that everybody knows that this isnt something that is allowed to happen . <hashtag> <hashtag> <hashtag> emma gonzlez leads a student outcry on guns this is the way i have to grieve <url> <end> (d) this is the first time across all the dustcovered years of our history centuries filled with courage and honor that the elected commanderinchief chose to tweet instead of plan to defend the country . from <user> <url> <end>	Leftwinger / Democrat / Anti-Gun

(continued)

Table A.3 *(continued)*

	Choices	Tweet Owner
Q7)	(a) you have to be pretty careful if youre saying that youre not going to fly on this airline because of that or were not going to use this railroad because of that buffet said . <user> <user> <hashtag> <hashtag> <hashtag> <hashtag> <url> <end>	Leftwinger / Democrat / Anti-Gun
	(b) you have to stand up for some things in this world . marjory stoneman douglas from voice of the river <hashtag> <hashtag> <hashtag> <hashtag> <user> <user> <user> <hashtag> <end>	
	(c) you have to be <number> in this country to buy a beer but you only have to be <number> to buy an ar <number> rifle <end>	
	(d) you have to keep guns away . save the bottom to your wet dreams . demand the <user> is not a bigger contributing cause . <hashtag>	
Q8)	(a) orphaned troubled past let the excuses begin hes a murderer and a terrorist stop with the bs <url> <end>	Leftwinger / Democrat / Anti-Gun
	(b) orphaned troubled past comments is akin to deflection . next step the medal of honor . <url>	
	(c) orphaned troubled past quit tryna make excuses for this kid he killed <number> innocent people enough said <url> <end>	
	(d) orphaned troubled past . cry me a fucking river . no fucking excuse <url> <end>	
Q9)	(a) as long as the democrats agree to build my stupid wall and deport all the daca recipients <url> <end>	Leftwinger / Democrat / Anti-Gun
	(b) as long as a gunowner s male is responsible . other industries are really around money to <user> & guns in innocents is willfully threatened by gun lobby .	
	(c) as long as the <hashtag> and other corporations are allowed to buy our politicians it wont matter how many children are killed in america because those victims are not the ones writing them their checks the nra is <url> <user> <end>	
	(d) as long as children continue to be injured and killed by guns in this country pediatricians will not rest in our pursuit to keep them safe . <user> in the wake of <hashtag> i urge you to advance meaningful gun safety legislation . <url> <end>	

(continued)

Table A.3 *(continued)*

	Choices	Tweet Owner
Q10)	(a) this is how you do it . <url> <end>	Leftwinger /
	(b) this is how you attract the leaders of tomorrow and today ! well done <user> thank you we the people <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <url> <end>	Democrat / Anti-Gun
	(c) this is how we need politicians . from a fellow gop cowards so exactly how much you profit from you pool . <hashtag>	
	(d) this is how it works <hashtag> <url> <end>	
Q11)	(a) we have to be strong on another extremist in the world by bribing senators for $. <hashtag> <url>	Leftwinger / Democrat / Anti-Gun
	(b) we have to keep them safe from the dangers outside ! do you realize that out there is our country ? dont just protect us from our country make our country safer . were going to spend the rest of our lives out there . <hashtag> <url> <end>	
	(c) we have to harden our schools not soften them up . a gunfree zone to a killer or someone that wants to be a killer that s like going in for an ice cream president trump says during school safety meeting in wake of florida school shooting <url> <url> <end>	
	(d) we have to know the details to prevent this from happening again . and hows that working for ya ? did knowing he was an orphan suddenly clear things up for you ? you dont care about the signs you want excuses . <hashtag> <end>	
Q12)	(a) its time to arm our teachers is literally the dumbest shit ive heard . you dont fight to end gun violence with more guns . <hashtag> <end>	Leftwinger / Democrat / Anti-Gun
	(b) its time to eliminate the misconception that ar stands for assault rifle and tell the world what it really stands for americas rifle . nra website . sandy hook aurora orlando san bernardino las vegas sutherland springs parkland america . <url> <end>	
	(c) its time to change lobbies bent = so important rather than saving our children are <hashtag> <url>	
	(d) its time to start ignoring trump . <hashtag> <url> <end>	

(continued)

Table A.3 *(continued)*

	Choices	Tweet Owner
Q13)	(a) i do not need this rifle . a man gave up his assault rifle . <url> <end>	Leftwinger /
	(b) i do not . from the damage the nra lines your pockets they do not even have empty words . <hashtag> <hashtag> <hashtag>	Democrat / Anti-Gun
	(c) i do not need this rifle this man gave up his ar <number> style gun because of the florida school shooting buzzfeed news <url> <end>	
	(d) i do not believe in compromise you want to compromise in weakness if you are winning you do not have to compromise marjory stoneman douglas at age <number> . <end>	
Q14)	(a) it is time for the victims to be the change we need to see emma gonzalez <hashtag> these teens are inspirationalin the wake of tragedy they have been focused eloquent and determined rattling the nation out of its complacency and apathy on the issue of gun control . <url> <end>	Leftwinger / Democrat / Anti-Gun
	(b) it is time to face up to the fact that the problem is deepseated complex and requires a societal examination of conscience . <hashtag> <hashtag> <hashtag> <user> <user> <user> <user> <user> <user> <user> <url> <end>	
	(c) it is time to see how we deal with the foolishness of the second amendment to end . karen keysgamarra fcps board member at large <hashtag> <hashtag> <end>	
	(d) it is time to revisit my grandstanding to hide about peoples families over school assault . weed in school get robbed public speak of this hat too . <url>	
Q15)	(a) gun control is accountable . email <user> that nra bribe tap dance about kids violence is hush money . <url>	Leftwinger / Democrat / Anti-Gun
	(b) gun control is not the same as making guns illegal . he used an ar <number> i dont understand how you can even compare a knife to an assault rifle . it is so uncommon to go into schools and commit mass slaughter with knives this doesnt make any sense . <url> <end>	
	(c) gun control is about saving the kids the proabortion left that thinks an abortion is appropriate at any stage of pregnancy <url> <end>	
	(d) gun control is too hard <hashtag> rains hell on for blaming school shootings on <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <url> <url> <end>	

(continued)

Table A.3 *(continued)*

	Choices	Tweet Owner
Q16)	(a) david hogg is fair game for critics <url> exactly right so how about the left cutting the bulls t ? ? ? <hashtag> <hashtag> <end> (b) david hogg is currently trending due to allegations that hes a crisis actor who is <number> years old and working for cnn . <hashtag> <hashtag> <url> <end> (c) david hogg is being feckless pawns and shamelessly coached antitrump antigun push for gun control . <url> (d) david hogg is simply too young and misinformed to be taken seriously . he has blamed the <hashtag> on every republican and expressed a lack of understanding about how government works . <url> <end>	Rightwinger / Republican / Pro-Gun
Q17)	(a) this is a real airline place scary she most would love to get used in concerning place <user> inciting violence upon leftist liberals <hashtag> (b) this is a bizarre interview she claims she was talking to the suspect while gun shots were still happening else where in the school and at times is smiling . <url> <end> (c) this is a true hero while <number> armed deputies stood by and did nothing <hashtag> <hashtag> <hashtag> <hashtag> <user> <url> <end> (d) this is a staged set up to make gun owners look bad wow just wow little desperate lefties <url> <end>	Rightwinger / Republican / Pro-Gun
Q18)	(a) i wonder if the <hashtag> took the life of one <user> kids or <user> grandkids would they be so quick to defend gun rights ? do not make us wait for one of your loved ones to die . enact reasonable <hashtag> <hashtag> <end> (b) i wonder if the sheriff nation is divided and imprisoned in a government <number> sole outcome for their recent issues only make work for liberals . <hashtag> (c) i wonder if <user> places within their contracts that you must be a liar to obtain the position at the network fake news cnn s chris cuomo pushes false story about buying ar <number> without id then lies about it <url> <end> (d) i wonder if high school students are taught the importance of our constitutional rights . <url> <end>	Rightwinger / Republican / Pro-Gun

(continued)

Table A.3 *(continued)*

	Choices	Tweet Owner
Q19)	(a) sheriff scott israel needs to resign blames his own failings on our constitution and rights ! shame on him ! <hashtag> <hashtag> <hashtag> <end> (b) sheriff scott israel needs to resign <hashtag> <number> of his deputies stood down while kids were getting shot and killed in <hashtag> <user> <user> <hashtag> are fighting back ! <url> <end> (c) sheriff scott israel accused of <hashtag> of a <number> year old girl in <number> . <hashtag> <hashtag> <hashtag> <hashtag> voat <url> <url> <end> (d) sheriff scott israel to keep hands out chuck schumer s crazy socialists <url> drain the swamp already . <hashtag>	Rightwinger / Republican / Pro-Gun
Q20)	(a) i m not going to sleep victim s father calls for school safety laws at trump session <hashtag> <hashtag> <url> <end> (b) i m not to own ar <number> shall not the militia shall not be infringed . <hashtag> can qual can be educated on the <hashtag> push <hashtag> (c) i m not sure this person fully grasps what guns are supposed to do . <hashtag> <url> <end> (d) i m not too surprised this video s been scrubbed from the internet . but can back the claim up via <user> radio program yesterday . he interviewed parkland high school student who said they had a drill for exact same scenario <number> wks ago . <hashtag> <hashtag> <url> <end>	Rightwinger / Republican / Pro-Gun
Q21)	(a) to all the anti <number> nd amendment anti gun douchebags who are in a constant state of hysteria and bloviate lies about ar <number> s and demand more gun laws we are prepared to die for our freedom and beliefs are you ? foreign & domestic <end> (b) to all the liberals and leftists that want to put out that way as lawabiding citizens understood this discussion . <url> (c) to all the liberals tweeting <hashtag> the parkland florida shooting happened in a gun free zone . it is illegal to bring a gun to a gun free zone . how will more laws stop shootings from happening ? <end> (d) to all the people who support gun control dont blame a gun for a persons action and if you retards due some research ar dose not stand for assault rifle it actually stands for armalite rifle named after the company who made it and the firearm was not built just for war and <end>	Rightwinger / Republican / Pro-Gun

(continued)

Table A.3 *(continued)*

	Choices	Tweet Owner
Q22)	(a) sounds like a hot commie cuck judge almost suggesting them lying democrats & leftism let this tragedy to get these personal things listen and choose again . <hashtag> (b) sounds like a wonderful start agreeable to everyone ! ! <url> <end> (c) sounds like a make america great again idea <hashtag> <hashtag> <user> <user> <url> <end> (d) sounds like a weak liberal judge if he s been to court that many times and still roaming the town . <url> <end>	Rightwinger / Republican / Pro-Gun
Q23)	(a) i m a mom & i would do it in a heartbeat to protect my kids . i ve had a home invasion while unarmed . i have no doubt how it will go down if they return . <hashtag> <url> <end> (b) i m a mom . that is why i own guns . <user> <hashtag> <url> <end> (c) i m a <hashtag> right to bear arms shall not be infringed in freedom of rights not free nor shall abridge our rights shall not be infringed (d) i m a prolife woman that believes in the <hashtag> a member of the <user> and women wearing pussy customs are just plain sick liberals . do not vote for democrats . <url> <end>	Rightwinger / Republican / Pro-Gun
Q24)	(a) that s a two way street dirtbags and u need usmuch more than we need u think about thatwhen you file your next sec earnings s form ! <url> <end> (b) that s a no brainer . easier said than done . still we need to try . everyone can spot mental crys for help . when you spot the gun it may be to late . <hashtag> <url> <end> (c) that s a $ <number> question ! <url> <end> (d) that s a disgrace ! if it s coincidental that leftists can negotiate with antifa ignorant libtards <url>	Rightwinger / Republican / Pro-Gun
Q25)	(a) broward county sheriff boasts about preserving our constitution ! god bless our <hashtag> <hashtag> <hashtag> (b) broward county sheriff no stranger to criticism for past handling of a mass shooting <url> <url> <end> (c) broward county sheriff scott israel represents what american police culture looks like by <user> <hashtag> <url> <end> (d) broward county sheriff dept connected to cair & hamas terrorists ! <url> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <user> <url> <end>	Rightwinger / Republican / Pro-Gun

(continued)

Table A.3 *(continued)*

	Choices	Tweet Owner
Q26)	(a) you are a liberal useful idiot off the wrong side of the left is false . <hashtag>	Rightwinger /
	(b) you are a moron <url> <end>	Republican /
	(c) you are a truly american young man we admired you and we wish we have more young ppl like you ! and not opportunists all over the fake news taking advantage of the horrible situation ! <url> <end>	Pro-Gun
	(d) you are a lemming believing only one side . not one gun has killed one person it takes a person to kill another . why not outlaw cars alcohol knives and all the other mechanics in which a person kills another . wake up ! <url> <end>	
Q27)	(a) i love how they automatically think the focus is on the white house and not the fbi that did the background check ask yourself why would the wh knowing the optics hide this if they knew and ask yourself why the fbi would hold info back or make the info convoluted <url> <end>	Rightwinger /
		Republican /
		Pro-Gun
	(b) i love how fascists & the second amendment does i thought it was <hashtag> but upped their communist $ for <hashtag> to <hashtag>	
	(c) i love how these celebrities come to twitter make a shit storm with their activist tweet and then never get involved in the thread ? what the hell ? <hashtag> <hashtag> <hashtag> <end>	
	(d) i love how liberal celebrities and politicians who live behind gated communities often with armed security are chanting for <hashtag> behind their army of moronic leftists . normal law abiding citizens should have a right to protect themselves . <url> <end>	
Q28)	(a) you re a leftist haters while plausible at all day so quickly again ! <hashtag> <user> <hashtag>	Rightwinger /
		Republican /
	(b) you re a puppet of the sick left . you can not disarm the american people . you have no right . guns are not the problem . people are the problem . you do not stop these people without the proper tool . that tool is a gun . <hashtag> <hashtag> <hashtag> is dishonest <url> <end>	Pro-Gun
	(c) you re a fucking lunatic if you think this is not about gun control <url> <end>	
	(d) you re a <hashtag> git er dun ! <smile> <url> <end>	

(continued)

Table A.3 *(continued)*

	Choices	Tweet Owner
Q29)	(a) i think it s very possible that some of these people are arranging the school shootings too . for $ $. there was another shooting in parkland fla one year before this recent one and that guy also said he heard voices in his head telling him to do it . <url> <end> (b) i think it s a great time for the <user> to think about expanding into the car rental business . <hashtag> <end> (c) i think it s more likely he s referring to some game manual <url> <end> (d) i think it is actually a new liberal defend trump except keep fighting patriots <hashtag>	Rightwinger / Republican / Pro-Gun
Q30)	(a) these people are clinically unable to accept reality . <url> <end> (b) these people are sick ! i hope america wakes up soon and sees the truth behind the globalist criminal elite . they have been hiding behind a shield of smoke and mirrors for too long and we have swallowed everything we have been fed by the msm . time for the great awakening ! <url> <end> (c) these people are evil ! <url> <end> (d) these people are still free no such guy the left will argue about <hashtag> like <hashtag> then you are clueless . stop exploiting this leftist and <hashtag> division !	Rightwinger / Republican / Pro-Gun
Q31)	(a) assault weapons are allowed through a bastardized interpretation of the second amendment . <hashtag> <end> (b) assault weapons are not the real problem ! the majority of gun murders are not mass shootings ! <hashtag> <url> <end> (c) assault weapons are saint of the second . oregon overrated georgia gov . smith & wesson . im complicit . <url> (d) assault weapons are so misunderstood is the not all men of the gun debate . <url> <end>	Leftwinger / Democrat / Anti-Gun

(continued)

Table A.3 *(continued)*

	Choices	Tweet Owner
Q32)	(a) i find it when the state houses of congress have sold congress to death to give . wake up ! ! ! <hashtag> <url> (b) i find it fascinating how the right to bear arms comes second only to free speech but is <number> amendments before outlawing slavery . i wonder when we got our priorities so out of order samuel g . <number> years old <hashtag> <hashtag> <user> <url> <url> <end> (c) i find it fascinating how the right to bear arms comes second only to free speech but is <number> amendments above outlawing slavery . i wonder when we got our priorities so out of order <hashtag> <url> <end> (d) i find it incredible that those who cant find an extra ten cents in taxes to pay for <hashtag> and <hashtag> and new <hashtag> buildings can now find for <hashtag> and bonuses for guntoting teachers . <hashtag> <url> <end>	Leftwinger / Democrat / Anti-Gun
Q33)	(a) i will be strongly pushing comprehensive background checks with an emphasis on mental health . raise age to <number> and end sale of bump stocks ! <user> says in a tweet <url> <end> (b) i will be undoing what i undid that obama initially did <url> <end> (c) i will be a professional school shooter is what <hashtag> said & amp z <hashtag> failed <number> investigate this so blood is on their hands . had this been someone of a different race / ethnicity / religion im sure it would have been handled differently . <hashtag> <hashtag> <url> <end> (d) i will be missed that kick off cockroaches . we fight for minds . yes god bless our brave young people ! <url>	Leftwinger / Democrat / Anti-Gun

(continued)

Table A.3 *(continued)*

	Choices	Tweet Owner
Q34)	(a) guns dont kill people people kill people . you know like nikolas cruz his gun didnt kill he did even if he didnt have his gun he would have still been able to kill <number> kids so <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <end> (b) guns dont kill people people kill people . if i see this as the only belief ppl have for their argument one more time im going to lose my gd mind . you can have guns if thats your thing ! ! theres just no reason to own an assault rifle ? ? ? like wtf are you going to do with it <end> (c) guns dont kill people people kill people . a person could kill a lot less people with a knife than with an assault rifle . <end> (d) guns dont kill or voters will be until getting thrown out of the white house neighbors and public bodies . their ultimate right is what it is to brutally . pow . peter mueller is watching him on the anus of the <hashtag> <hashtag> <url>	Leftwinger / Democrat / Anti-Gun
Q35)	(a) to all the vets who ask our kids it or only speak up a . <number> nest shot at say their children sacrificed for slaughtering their children . <hashtag> (b) to all the politicians taking donations from the nra shame on you . <url> <end> (c) to all the generations before us we sincerely accept your apology . we appreciate youre willing to let us rebuild the world that you fucked up . <user> on real time with bill maher love it ! these kids have guts and a sense of humor . <end> (d) to all the politicians taking contributions from the <user> shame on you . student emma gonzales . crowd echoes her chanting shame on you . she goes on to call out the president & amp <user> <hashtag> <hashtag> <hashtag> <hashtag> <url> <end>	Leftwinger / Democrat / Anti-Gun
Q36)	(a) the right to own an assault rifle is not more important than one life . says a gun owner <url> <end> (b) the right to bear arms is not the full <number> nd amendment . the <number> nd amendment reads a well regulated militia being necessary to the security of a free state the right of the people to keep and bear arms shall not be infringed . what well regulated militia is this group ? <url> <end> (c) the right to bear arms can not be greater than the right to life . spains newspaper <user> <hashtag> <hashtag> <hashtag> <end> (d) the right to confront the old white male terrorist organization . ar <number> differ from a bear . the only guarantee to kids and others	Leftwinger / Democrat / Anti-Gun

(continued)

Table A.3 *(continued)*

	Choices	Tweet Owner
Q37)	(a) i am not a crisis actor florida teens fire back at rightwing conspiracy theorists <url> i stand with the <hashtag> <hashtag> <end>	Leftwinger / Democrat / Anti-Gun
	(b) i am not a <hashtag> there is quite a backlash against the attempt of rightwingers and the <hashtag> crowd to undermine the credibility of the <hashtag> students florida teens fire back at rightwing conspiracy theorists <url> <hashtag> <hashtag> <end>	
	(c) i am not content with being in the corner anymore . this is powerful stuff . <hashtag> <hashtag> how teens want to solve americas school shooting problem pbs newshour <url> <end>	
	(d) i am not afraid of trying to shoot my country some of the current state of pa mooch . we scream mental illness ! ! <url>	
Q38)	(a) thank you for you onvergelijkbaar voudrais elise are we realize how way porn would want serious do both education to hide it . same assault weapons ban strengthened gun control laws <url> <url>	Leftwinger / Democrat / Anti-Gun
	(b) thank you for all your work <hashtag> <url> <end>	
	(c) thank you for choosing insert rental car company will you be needing a toddler seat for your semi automatic combat rifle with the bump stock and duffel bag full of <number> round magazines ? yes thanks do you know what time school lets out around here ? <hashtag> <url> <end>	
	(d) thank you for your strength dedication & amp commitment <url> <end>	
Q39)	(a) a bunch of make money . fuck you . money . kids over guns . <hashtag> <hashtag>	Leftwinger / Democrat / Anti-Gun
	(b) a bunch of nope . <hashtag> <url> <end>	
	(c) a bunch of kids lecturing me on the <hashtag> about the <hashtag> would be cute if they were not so ignorant . <end>	
	(d) a bunch of punk bishes . <url> <end>	

(continued)

Table A.3 *(continued)*

	Choices	Tweet Owner
Q40)	(a) i am a <hashtag> not an instrument of violence <url> via <user> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <end>	Leftwinger / Democrat / Anti-Gun
	(b) i am a concerned parent that fully supports the walkouts . my son survived the santana high school shooting on this day in <number> . his best friend randy gordon a senior was killed as was bryan zuckor a freshman . <number> others were injured . the pain lessens but never ends . <url> <end>	
	(c) i am a repulsive replica of a human being but i would like some attention now please . <url> <end>	
	(d) i am a tan and french army of firearms but i just want to prosecute lazy candidates . no civilian should have an assault rifle all the time .	
Q41)	(a) this is a shameful idea . lobbyist aligning with and making sure that children r shame ! <url>	Leftwinger / Democrat / Anti-Gun
	(b) this is a matter of pubic safety we urge you to listen to the american public and to the law enforcement community and support a ban on the further manufacture of assault weapons . signed presidents ford carter and reagan nearly <number> yrs ago <url> <end>	
	(c) this is a man who supports teenagers and other people [being allowed] to purchase assault weaponsand then comes down here and wants to act as though this is horrible and shouldnt happen but goes back to washington and supports it . its hypocrisy . <end>	
	(d) this is a community & amp a state that is in deep pain & amp they want action to make sure this doesnt happen again . <url> <end>	
Q42)	(a) i think the democrats are the party of lisa simpson and republicans are happily the party of homer and bart and maggie and marge <user> at the cpac gathering in national harbor maryland <hashtag> <hashtag> <hashtag> <url> <end>	Leftwinger / Democrat / Anti-Gun
	(b) i think the gop came through . also gun lovers remind them of young leaders .	
	(c) i think the best way to deal with this is to ignore him . the things that pres trump tweets is nothing that will have a lasting impact unless its a negative lasting impact on the people around us . <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <end>	
	(d) i think the democrats are the party of lisa simpson and the rrpublicans are happily the party of homer and bart and maggie and marge . ted cruz off to a great start here . <hashtag> <end>	

(continued)

Table A.3 *(continued)*

	Choices	Tweet Owner
Q43)	(a) we the people are more than <number> million <url> <end> (b) we the people we re gon na represent ! ! you vote poc & people get blown away ! yay this wild west that d die . (c) we the people have to do it if <hashtag> and <user> the disgusting human offal wont <hashtag> <url> <end> (d) we the people are trying to prevent our fellow citizens from dying and the next generation that has grown up with your bullshit is not going to take it anymore . if youre too afraid to do your job because the nra wont like you very much gtfo you snowflake . <url> <end>	Leftwinger / Democrat / Anti-Gun
Q44)	(a) guns do not kill people do guns do not kill people with guns do <hashtag> <end> (b) guns do not kill people people kill people . but people with guns kill lots of people . <hashtag> <hashtag> <hashtag> <url> <end> (c) guns do not kill people with guns do guns do not kill gun lobbies do <hashtag> <url> <end> (d) guns do not belong in the hands of civilians they re power combined with their kk views . maybe we have to deal with the text of those of us <number> m+ lifes and i ll happily scream about guns .	Leftwinger / Democrat / Anti-Gun
Q45)	(a) good guy with a gun to the rescue once again ! <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <number> . <url> <number> . <url> <url> <end> (b) good guy with a gun theory is bogus . <hashtag> <hashtag> <url> <end> (c) good guy with guns stop bad guys with guns . <hashtag> <hashtag> <url> <end> (d) good guy with a simple gun contest . and their empty fake tears get thought anytime . lands offshore and backing the top are stone go ! ! maga ! ! !	Leftwinger / Democrat / Anti-Gun
Q46)	(a) it s time for your meds ray . nurse ratched is waiting <hashtag> <hashtag> <hashtag> <url> <end> (b) it s time to end gunfree schools . <url> <end> (c) it s time ! <hashtag> <url> jerome corsi aux le agencies are corrupt ! <hashtag> (d) it s time for us & our elected officials to prioritize the protection of our children . enough with the double standards . guns defend & protect lives . <hashtag> <hashtag> <hashtag> <url> <end>	Rightwinger / Republican / Pro-Gun

(continued)

Table A.3 *(continued)*

	Choices	Tweet Owner
Q47)	(a) i will not support their philosophy . they attack the <user> but none of the shooters were nra . but all of the shootets were on facebook . <hashtag> <user> <url> <end> (b) i will not that <number> rep government carries <hashtag> gun guide any man child should remember so other leftist group hoodlums would have permits like them ! <hashtag> <url> (c) i will not relinquish my <hashtag> right . fascist <hashtag> people need to stop attempting to diminish my birthright . an armed citizenry doesnt board the trains to concentration camps . <url> <end> (d) i will not give hogg anymore attention & focus on . <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> we ca not hear you ! <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> hogg <url> <end>	Rightwinger / Republican / Pro-Gun
Q48)	(a) i dont think he understands how due process works you get that first ! and the attack on <hashtag> rights to go along with that . this was an all out dumpster fire . <url> <end> (b) i dont think that shall not be infringed the second amendment does not exist in the world where the constitution is important . <url> (c) i dont think he even knows what an assault rifle is <end> (d) i dont think this is the way to keep kids and teachers safe . im barely paid enough to teach kids much less carry a weapon and risk getting shot at by the police when they show up in response to an active shooter call . hire professional security to keep our kids safe every day . <url> <end>	Rightwinger / Republican / Pro-Gun
Q49)	(a) i agree . i just became a member of nra and am not looking for discounts . i wo not do business with those companies but that s because they became political when it was not necessary . i wanted to ensure our <hashtag> rights remain as well as our full constitution . <url> <end> (b) i agree . the left is trying to vilify the nra that makes our country safer through protecting the <hashtag> and educating people on the safe and lawful use of firearms . these companies need to understand the fascist left are running a scam . there will be a backlash if you fall for it ! <url> <end> (c) i agree . i support the nra too . <url> <end> (d) i agree . shall not be infringed ! there s no necessary clause here s the suggestion for supporting ones stay . <hashtag>	Rightwinger / Republican / Pro-Gun

(continued)

Table A.3 *(continued)*

	Choices	Tweet Owner
Q50)	(a) just joined the law abiding citizens for raising the purchase less liberal rhetoric . this only few make gun free zones . (b) just joined the <user> <number> year membership . hope you ll do the same . <hashtag> <end> (c) just joined the nra today if you re not a member please consider becoming one <url> <end> (d) just joined the nra ! stand and fight <hashtag> <end>	Rightwinger / Republican / Pro-Gun
Q51)	(a) you do not believe guns are protected so far left msm needs to receive him held in support & backtracking on his twitter handle <hashtag> (b) you do not need an ar <number> in <number> blackclad knifewielding attackers including two women killed at least <number> people and injured about <number> at a railway station in the southwest chinese city of kunming . four of the assailants were k <url> <end> (c) you do not know what your talking about . president trump said they he would like to take the guns of sick people ! ! ! not everyone . <url> <end> (d) you do not be a gun owner to join the nra . it s an investment in guarding our constitution right . no sane person would ever <hashtag> <end>	Rightwinger / Republican / Pro-Gun
Q52)	(a) these kids are really pushing the limit to anyone who can tolerate their s t. they already used up all my sympathy for them . wonder how long anyone else can endure ? <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <url> <end> (b) these kids are just kids and the shallowness & deceptive nature of their advocacy is illustrative of that . <url> <end> (c) these kids are engaging in <hashtag> circus liberals have different views can damage we learn about guns/ stupid or dangerous . <url> (d) these kids are not smart enough democrats and the media are only using these kids to push their agenda . <url> <end>	Rightwinger / Republican / Pro-Gun

(continued)

Table A.3 *(continued)*

	Choices	Tweet Owner
Q53)	(a) do you know the single biggest lie americans must believe to buy into federal gun control has nothing to do with the wording of the <hashtag> its all here in the key to understanding the constitution <url> <hashtag> <hashtag> <hashtag> <hashtag> <end> (b) do you know about the act of <number> teach this in your school <url> via <user> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> or you can go watch <hashtag> that should save your country <hashtag> or <hashtag> is anybody still freaking doubting they lie ? <end> (c) do you know what is more passionate that politics ? gun politics as it overrules the logic & drives the heat with emotion that you can no longer own a gun any gun and that confiscation is the ultimate goal people do not want to give up their guns & will fight to keep them <hashtag> <url> <end> (d) do you know no gun shall infringe <number>. protects gun rights private ownership of individuals violating individual freedoms . <hashtag>	Rightwinger / Republican / Pro-Gun
Q54)	(a) why is this such a hard concept to understand ? <url> <end> (b) why is this being covered up ? <user> <user> <user> <user> this is remarkable ! <hashtag> <url> <end> (c) why is this purpose but thanks you and the globalists armed men of children to school killer or not <hashtag> instead of <hashtag> protect em ! <hashtag> (d) why is this an either/or choice ? <url> <end>	Rightwinger / Republican / Pro-Gun
Q55)	(a) here we go . <url> <end> (b) here we go again . when will it stop ? gun control is a big problem . period . <hashtag> <end> (c) here we go again by the <hashtag> the <hashtag> <user> <user> <user> <hashtag> <url> via <user> <end> (d) here we go ! smoke crack done with the liberal socialist lying thousands of killing poor planned parenthood abortion lol i own guns are therefore will not be dictated by someone who only told us they will regret attacking <user> for . <hashtag>	Rightwinger / Republican / Pro-Gun

(continued)

Table A.3 *(continued)*

	Choices	Tweet Owner
Q56)	(a) here is the proof <url> <end> (b) here is the response to the <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <url> <end> (c) here is the federal gun purchase form . what additional questions need to be asked ? seems a working background check system and prosecution of those who lie would go a long ways to improve citizen safety . note zero gun deaths is unreasonable . <url> <url> <end> (d) here is the time <number> deputies stood outside and did not win time from <number> not words very well many threats ! ! ! <hashtag>	Rightwinger / Republican / Pro-Gun
Q57)	(a) you can not make a <hashtag> a legit argument . a weapon can not load chamber and fire itself . you talk about a mental background check . well idiots theres a thing called hippa its illegal to search health records . again <hashtag> <hashtag> <end> (b) you can not buy an automatic weapon or possess one unless you have a certain very restricted ffl license . that is the law today . get informed please . <url> <end> (c) you can not argue with ppl . asking to kill others . they are idiots . <url> <end> (d) you can not be fair trashing explaining that dems are already indoctrinated by such idiots in this part of media lynch . deal is not logical . thx <user> <hashtag>	Rightwinger / Republican / Pro-Gun
Q58)	(a) i want to see the dipshit who sold this shifty eyed <number> year old a fucking ar <number> . <end> (b) i want to know if any useless russian s trollers knew <hashtag> or <user> <user> <user> <hashtag> <user> <hashtag> <user> <user> <user> <hashtag> <end> (c) i want to remove self freedom . <hashtag> deletes <hashtag> question . lordy ! <hashtag> what ? do u s think ? <hashtag> (d) i want to start a big peaceful protest in houston to speak out for more gun reform laws if anyone is interested inbox me also retweet this so everyone will know about the protest ! <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <end>	Rightwinger / Republican / Pro-Gun

(continued)

Table A.3 *(continued)*

	Choices	Tweet Owner
Q59)	(a) what the hell . dear <user> stop kneecapping a necessary policy . <url> <end> (b) what the hell happened to nicole who worked for bush oh i forgot she was indoctrinated at the view brainwashed paid big money geezz <url> <end> (c) what the hell are he talking about it and knows what conservative opinions from the subject look sounds . leftists need to learn on their agenda . <hashtag> (d) what the hell ? ! fashion models walk runway carrying their heads in gucci show <url> via <user> <end>	Rightwinger / Republican / Pro-Gun
Q60)	(a) why is the argument surrounding gun about banning assault rifles vast majority of murders are with handguns not rifles the left is not at all philosophically consistent but then again they are all for mercilessly murdering babies . <hashtag> <hashtag> <hashtag> <hashtag> <end> (b) why is the left s throats . they are the ones who are talking about the <hashtag> and the <hashtag> <hashtag> (c) why is the media making this shooter famous ! ? <hashtag> <hashtag> stop giving him attention and push the drug companies that made him do this . <end> (d) why is the focus on guns ? the focus should be on the fact that bipolar & schizoid people can not be dealt with due to privacy laws <url> <end>	Rightwinger / Republican / Pro-Gun

Table A.4: Answer key of disclosure test in Table A.3.

Q1) a	Q2) c	Q3) c	Q4) d	Q5) c	Q6) a	Q7) d	Q8) b
Q9) b	Q10) c	Q11) a	Q12) c	Q13) b	Q14) d	Q15) a	Q16) c
Q17) a	Q18) b	Q19) d	Q20) b	Q21) b	Q22) a	Q23) c	Q24) d
Q25) a	Q26) a	Q27) b	Q28) a	Q29) d	Q30) d	Q31) c	Q32) a
Q33) d	Q34) d	Q35) a	Q36) d	Q37) d	Q38) a	Q39) a	Q40) d
Q41) a	Q42) b	Q43) b	Q44) d	Q45) d	Q46) c	Q47) b	Q48) b
Q49) d	Q50) a	Q51) a	Q52) c	Q53) d	Q54) c	Q55) d	Q56) d
Q57) d	Q58) c	Q59) c	Q60) b				

Table A.5: Tweets used in Relatedness and Quality test explained in Chapter 6.

Dialogue	Reply Owner	Belonging
+<user> <user> get rid of the assault weapons that are used in virtually every one of these mass murders . it wont fix everything . but it might make it a little harder for some pissed off loner to get an ar <number> and kill dozens of people . <end> -<user> <user> ar <number> s are not assault weapons . assault weapons are used for hunting rifles .	Right-winger	Bot
+<user> <user> none of which would be necessary if a disturbed <number> year old didnt have an assault rifle . <end> -<user> <user> if he was wanted he would use an assault rifle .	Right-winger	Bot
+<user> nobody with more than three brain cells to rub together believes you . you re an <user> lapdog the same as the rest of the <user> bloodmongers . <hashtag> <hashtag> <hashtag> <hashtag> <url> <end> -<user> <user> <user> <user> i believe you need to have due process thats not the problem .	Right-winger	Bot
+<user> <user> <user> keep your handgun . neither you nor anyone else needs an ar <number> to defend their homes . <end> -<user> <user> <user> <user> you do know that ar does not stand for assault rifle ?	Right-winger	Bot
+<user> the best way to stop the massacres is to kill the beast . <hashtag> <end> -<user> the issue is there to be blocks instead of people who care about them to get them . <hashtag>	Right-winger	Bot
+a cheater thinks everyone else is a cheater . i avoid guys who think everyone else is armed and ready to shoot . <hashtag> <end> -<user> <user> are you trying to ! they screw <hashtag>	Right-winger	Bot
+<user> <user> she s not blocking enough kids . if she hasnt fired yet an ar <number> has already shot off a dozen rounds into her and as many kids . i really dont think you grasp how dangerous guns are . sad . <end> -<user> <user> <user> maybe he had a background check he would not have been done .	Right-winger	Bot

(continued)

Table A.5 *(continued)*

Dialogue	Reply Owner	Belonging
+<user> <user> it was recently explained to me that bump stocks and <number>+ round magazines are critical to the sport of boar hunting and therefore can not be regulated . <end> -<user> <user> <user> thats what is an argument . between an ar and bump stocks .	Right-winger	Bot
+<user> <user> <user> i have no problem saying the gun killed those kids . a gun killed those kids ! <hashtag> <end> -<user> <user> <user> there are a lot short of people . it does not fire meaning they are illegal . the ar <number> is not an assault weapon . it is not a machine gun .	Right-winger	Bot
+<user> please discontinue your relationship with nra . <hashtag> <hashtag> <end> -<user> <user> well if you give up the ar <number> i will	Right-winger	Bot
+<user> <user> <user> that was a group of terrorists . so yeah a group with knives can do a lot of damage . imagine how many would have died if they all had ar <number> s ? nice try <end> -<user> <user> <user> i agree we can thank god in a progressive of florida we have an ar <number> and pass the background check system .	Right-winger	Bot
+<user> plus you dont use an ar <number> to hunt <end> -<user> <user> an ar <number> issue is not about hunting .	Right-winger	Bot
+<user> <user> bullshit there are thousands of fbi agents and not all of them are investigating the trump / putin conspiracy a simple assault weapons ban would at the very least reduce the damage done by crazed shooters . <end> -<user> <user> <user> banning assault weapons is a waste .	Right-winger	Bot
+<user> well isnt that just the height of hypocrisy . surprised ? ! no <hashtag> <hashtag> <end> -<user> <user> the <hashtag> says nothing about nothing to do with a good guy with a handgun . <hashtag>	Right-winger	Bot

(continued)

Table A.5 *(continued)*

Dialogue	Reply Owner	Belonging
+<user> <user> <user> <user> <user> <user> oh and they need to also ban the ar <number> pistol . not just the rifle . the ar <number> pistol short barrel and lack of a stock make the gun more versatile in confined spaces . <end> -<user> <user> <user> <user> <user> ar <number> style are not assault rifles sound like a ban uses . google which specific difference is using strictly semiauto as accuracy . also uses the same as a pistol grip to begin with .	Right-winger	Bot
+because cair appears to matter more than the safety of the kids in schools in broward county . <url> <end> -<user> <user> we ban sale of assault weapons ? it s the problem	Left-winger	Bot
+support our politicians and political candidates who support our constitutional rights . <url> <end> -<user> please consider it to a march . <hashtag>	Left-winger	Bot
+i m a new member of the <user> i signed up yesterday . i have for house protection but will take a cc class . nra is not the problem . training and education is the solution . <hashtag> <end> -<user> <user> here comes to be careful what you think <hashtag> should be able to buy assault weapons	Left-winger	Bot
+they do not blame the gunman . they do not blame the sheriff s department . they do not blame the mental health department . they do not blame the school district . they do not blame the fbi . but they blame <user> <end> -<user> <user> youre not a much ! banning assault weapons talks . ban the military .	Left-winger	Bot
+you ve been on twitter since <number> & have <number> followers . you can see how popular your ideas are even among your own fellow liberals . the <number> nd amendment does not give us the right to bear arms & defend ourselves . our god does . <number> nd amendment merely reiterates our godgiven rights <url> <end> -<user> outside the <number> nd amendment that do u need assault weapons ? you do not need .	Left-winger	Bot

(continued)

Table A.5 *(continued)*

Dialogue	Reply Owner	Belonging
+<user> <user> hes not really a republican at the core and it kind of showed when he said take the guns first due process second . <end> -<user> <user> he just wants it and process and the impact of . no due process for those domestic abusers adults semiautomatics .	Left-winger	Bot
+<user> <user> <user> thing is assault weapons is not an actual gun . it does not exist . that s one reason why we do not tak <end> -<user> <user> and if i fear of the <hashtag> price they did not need a background check .	Left-winger	Bot
+to all the leftists pushing <hashtag> i d sooner die than let you take away my <number> nd amendment rights . <hashtag> <end> -<user> do not see you believe in fear you guys space about your continued support . <hashtag> <hashtag>	Left-winger	Bot
+sounds accurate all the way to the pink vagina hats . <url> <end> -<user> how is the answer to this kid pass a background check .	Left-winger	Bot
+and this is why we carry <hashtag> <hashtag> <hashtag> <hashtag> <url> <end> -<user> we need to protest the public . you have a lot of blood to protect our children . <hashtag> <hashtag>	Left-winger	Bot
+criminals will always get guns one way or another ! regardless of gun laws ! <url> <end> -<user> the <user> should have to be dead and murdered with them . go shoot to reach out . <hashtag>	Left-winger	Bot
+i love guns and i love ar <number> s . change my mind . let the games begin <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <hashtag> <end> -<user> pockets . they do not anymore . <hashtag> <hashtag> <hashtag>	Left-winger	Bot
+this is why we need an <hashtag> stupid af mfs who say shit like this . like we need ar <number> s to protect our freedom of speech assembly and due process . we need a culture change . <hashtag> <hashtag> <hashtag> <hashtag> <url> <end> -<user> preach ! ! ! thoughts and prayers are on that . i ca not vote tell u too bad guy with an ar <number> .	Left-winger	Bot

(continued)

Table A.5 *(continued)*

Dialogue	Reply Owner	Belonging
+whats a student doing with this ? <hashtag> <url> <end> -<user> i have to say they do something about what it is right ! <hashtag>	Left-winger	Bot
+<user> because an assault weapons ban would accomplish nothing . it didnt in the <number> s and it wound now . <end> -<user> i use words to talk now . how about tweeting all of guncontrol laws ? unfortunately it is not me about <hashtag> <hashtag> <url>	Left-winger	Bot
+trump demands due process for aides accused of abusing women <url> -<user> intelligent / prudent people should be asking where are our due process laws in this country ? presumption of innocence amongst them allegations are not proof if you are denying them . hightech lynchings ? . due process applies to u too . just a thought	Right-winger	Real User
+two members of the central park five criticized president trump for his tweet that a mere allegation could ruin people s lives without due process <url> <url> -<user> due process is a part of our constituion	Right-winger	Real User
+the <hashtag> boycott is working here is an update dozens getting laid off due to massive lack of revenue keep it up patriots ! we are more powerfup when we unite against the <hashtag> phuckery <hashtag> <hashtag> <url> -<user> <user> <user> <user> <user> <user> <user> <user> <user> <user> <user> while i hate contemptable nasty news cnn i do feel for the employees who are out of work because of the management and antitrump onair talent . of course they may be just as hateful . cnn has no credibility on the truth meter . <hashtag>	Right-winger	Real User
+central park five defendant calls trump out now you care about due process ? <url> <url> -<user> i think <user> does believe in due process all the time .	Right-winger	Real User

(continued)

Table A.5 *(continued)*

Dialogue	Reply Owner	Belonging
+there is a common disapproval between a large number of democrats who feel that everyday citizens should not have the right to carry semi-automatic weapons . taking away these style of guns would be a direct infringement on our <hashtag> rights . -but not to worry ! there is a massive group of <hashtag> supporters who are backed by both the <user> and <user> these typically republican citizens have a powerful voice in the <hashtag> debate .	Right-winger	Real User
+more legal guns used in st. louis area crimes <url> <url> -<user> legally owned is very different than legal to own a <number> is legal to own . doesnt mean that same <number> is legally owned after being illegally taken . if you want <hashtag> <hashtag> you yourself need common sense before we even talk about it .	Right-winger	Real User
+<number> . the white house claim again and again and again was that porter s background check was ongoing and they needed to complete the process this was a lie according to wray the background check formally ended in january <url> -<user> <user> the question is what was in the background check a year ago or not in it ?	Right-winger	Real User
+<user> yeah like the due process he wanted for the central park five . -<user> <user> there was due process in the central park five trial additional evidence was discovered all accused were exonerated of that crime	Right-winger	Real User
+<user> yeah like the due process he wanted for the central park five . -<user> <user> but allegations are not a trial therefore there has been no due process for rob porter and the other wh official .	Right-winger	Real User
+central park five exoneree slams president trump for due process remark when i first saw the tweet i was not surprised because of all the stuff that he has done already leading up to this . it goes back to <number> <url> -<user> just so everyone knows the central park five received due process . they were found guilty	Right-winger	Real User

(continued)

Table A.5 *(continued)*

Dialogue	Reply Owner	Belonging
+trump has consistently responded to the allegations of assault or abuse against women by expressing sympathy for the men being accused <url> <url> -<user> due process ! have you ever seen a woman that lied ? lol ! accusations from women are the dems new idea for regaining power . how dare anyone question whether or not these allegations are true .	Right-winger	Real User
+<user> so trump said he believed in due process on porters case yet still blamed central park five even after it was proven wrong ? hmm whats the difference between of them ? color me baffled -<user> <user> the central park <number> were found guilty in a court of law . they received due process .	Right-winger	Real User
+i m tired of hearing and writing about peter strzok & lisa page . they are still getting paid . in a corporate setting they would be shit canned along with wray for not firing them . no accountability at fbi . why americans have no faith in it . -<user> thats right to the point ! not fired because they work for federal government that gives them lots of what ? due process ! how ironic ! ! !	Right-winger	Real User
+i m tired of hearing and writing about peter strzok & lisa page . they are still getting paid . in a corporate setting they would be shit canned along with wray for not firing them . no accountability at fbi . why americans have no faith in it . -<user> yeah only civil servants seem to get due process	Right-winger	Real User
+i know it isnt popular in america to say this but i hate guns they serve no good purpose they just take human life and change other lives . we need major reform on this issue <user> you paying attention ? <hashtag> <hashtag> <url> -<user> <user> just because you have this unhealthy fear of inanimate objects doesnt negate my right to defend myself . <hashtag> needs a clue .	Right-winger	Real User

(continued)

Table A.5 *(continued)*

Dialogue	Reply Owner	Belonging
+sick of this bullsh <hashtag> the left wants to disarm us i dont care what laws they introduce or pass in the future prohibiting my right to protect myself by any means necessary . seems idiotic for those who hate guns to go after those who own them <url> -<user> it s like you ve forgotten about parkland already . when will you actually do something . you are a waste of space who does nothing for americans except the rich the haters & the racists . those young people those families need you to do something for <hashtag>	Left-winger	Real User
+. <user> cited broken homes without dads as being something that is common in these shootings . how does he explain the fact that there arent mass shootings in nations with higher divorce rates than the us like belgium spain and hungary ? <url> -<user> why are not you calling out the <user> directly . they are responsible for every murder today and for every murderer that uses an ar <number> military style assault rifle .	Left-winger	Real User
+who has a list of advertisers ? please respond here with company names phone numbers and twitter handles <url> -<user> you are damn right anger is bubbling over . over the entire country . <hashtag>	Left-winger	Real User
+we will not be gaslighted into thinking that were responsible for a tragedy that we had nothing to do with nra spokeswoman dana loesch said at the annual conservative political action conference <url> <url> -<user> the nra set themselves up as the sole protectors of the <number> nd amendment they are not . we the people are the protectors of our rights . the nra is a propaganda machine with tremendous power & we have to cut off their $ $ and get them out of dc . <hashtag>	Left-winger	Real User
+good great meeting in the oval office tonight with the nra ! -<user> you know i m all about due process opportunities to be heard and independent judicial review of executive decisions . you also know that without smart independent judges this whole liberty thing collapses quickly . and the good news about process is it can be modified .	Left-winger	Real User

(continued)

Table A.5 *(continued)*

Dialogue	Reply Owner	Belonging
+this is horrific . in november a tipster called bso to say cruz could be a school shooter in the making but deputies did not write up a report on that warning . it came just weeks after a relative called urging bso to seize his weapons . <url> -<user> <user> <user> <user> we are so proud of all of you ! the world is yours now ! run free with your hopes & dreams ! its your turn ! we are right beside you . we have your backs . stay strong take breaks stay loud . <hashtag> <hashtag> <hashtag> your are our future <user> <url>	Left-winger	Real User
+been saying this forever & it remains just as true today as its ever been the political clout of the <user> comes from its ability to mobilize its members for issue advocacy & elections not from its direct campaign giving which is relatively meager . <url> -<user> treating parkland victims changes debate on guns ar <number> wounds radically different from other mass shootingsct scan of a victim s organ looked like melon smashed by a sledgehammer only shreds remained nothing could be done to fix the injury it was fatal <url>	Left-winger	Real User
+interviews and stuff and shes on the phone all day and super busy but working hard we promise . <hashtag> <hashtag> <url> -<user> you dont drive a nascar on the street no matter how fun it might be just like you dont need an ar <number> to protect yourself when walking home at night . amen !	Left-winger	Real User
+<user> <user> <user> <user> <user> what about assault baseball bats assault hammers assault knives assault vehicles ? people use those inanimate objects to kill also you clowns . using the word assault is stupid and not necessary . -<user> are these mentally ill people the ones we should be taking guns from first and due process second ?	Left-winger	Real User
+just in senators introduce bipartisan bill to block terrorists from buying guns <url> <url> -<user> leek the damn dem response . stop bringing knives to an assault rifle showdown !	Left-winger	Real User

(continued)

Table A.5 *(continued)*

Dialogue	Reply Owner	Belonging
+protect your children . make the changes . we are not acting we are dying . <url> -<user> who gives a flying fart in space about your buddy and his family unless they stop the russians and outlaw assault weapons ! <hashtag>	Left-winger	Real User
+we all know that russia funnels money through the nra and that the nra willingly happily takes money from them which means they more than likely influence the sudden coddling and love fest of putin . the idea that robert mueller could very well be investigating them is lovely . -<user> proof in film that she lied burn her <hashtag> truth shame on you <hashtag> <hashtag> <hashtag>	Left-winger	Real User
+here are people the <user> and <user> fought on behalf of to make sure they can buy guns and ammo in any quantity <number> . the mentally ill <number> . people on terrorist watch list . when they won on the mentally ill nra s newsletter went to members celebrating the good news . seriously . -<user> pure and simple no one should be able to have assault rifles as civilians . hand guns for those who pass background checks over <number> . make it right for once . use your common sense ! <hashtag> <hashtag>	Left-winger	Real User
+if it was the goal of russia to create discord disruption and chaos within the u.s. then with all of the committee hearings investigations and party hatred they have succeeded beyond their wildest dreams . they are laughing their asses off in moscow . get smart america ! -<user> ca not get much dumber than that . what you would expect . <hashtag>	Left-winger	Real User
+this is immoral . what youve been through does not give you the high ground to falsely attribute impure motives to those who disagree . <url> -<user> this generation grew up with social media . motivating & organizing is the easy part for them . you re right to be afraid . only a moron would really underestimate the power of america s youth . <hashtag>	Left-winger	Real User