

皮肉検出のための皮肉の対象の推定

肥合 智史[†] 嶋田 和孝^{††}

[†] 九州工業大学 大学院情報工学府

福岡県飯塚市川津 680-4

^{††} 九州工業大学 大学院情報工学研究院

福岡県飯塚市川津 680-4

E-mail: [†]{s_hiai, shimada}@pluto.ai.kyutech.ac.jp

あらまし 近年、皮肉を検出するタスクが多くの研究で取り組まれている。皮肉では、肯定表現を用いて、否定的意図を伝えることがあり、極性語を重要な特徴としている感情分析の分析誤りの原因となることが考えられる。よって、皮肉の検出は感情分析の精度向上に貢献する。皮肉の重要な特徴の一つに、皮肉には、皮肉によって批判される対象が存在するというものがある。本研究では、皮肉の検出の精度向上に貢献することを目的として、皮肉の対象の検出に取り組む。まず、皮肉の対象検出のためのデータセットを構築する。次に、機械学習を用いた皮肉の対象の検出手法を適用する。そして、ベースライン手法との検出精度の比較や、検出できる対象の違いを観察し、検出手法の有用性を評価する。

キーワード 皮肉, 感情分析, 意見マイニング, 皮肉の対象, マイクロブログ

Target Detection for Sarcasm Identification

Satoshi HIAI[†] and Kazutaka SHIMADA^{††}

[†] Graduate School of Computer Science and Systems Engineering, Kyushu Institute of Technology
680-4 Kawazu, Iizuka, Fukuoka 820-8502, JAPAN

^{††} Department of Artificial Intelligence, Kyushu Institute of Technology
680-4 Kawazu, Iizuka, Fukuoka 820-8502, JAPAN

E-mail: [†]{s_hiai, shimada}@pluto.ai.kyutech.ac.jp

Abstract Sarcasm identification has been treated as a task that classifies text as sarcastic or non-sarcastic. Sarcasm identification is a significant challenge for sentiment analysis because sarcasm involves a positive expression with negative meaning. Targets of sarcasm in sarcastic sentences are an important feature for sarcasm identification. In this paper, we study a detection task of targets of sarcasm in sarcastic sentences. First, we construct a dataset for the target detection task. Then we apply two detection methods using machine learning models to the dataset. We compare the methods with a baseline method. The results demonstrate the effectiveness of our method.

Key words Sarcasm, Sentiment analysis, Opinion mining, Target of sarcasm, Microblogging

1. はじめに

皮肉は、相手の欠点や弱点を意地悪く遠まわしに非難する表現である^(注1)。皮肉では、肯定表現を用いて、逆に否定的な意味を伝えることがある。

[例 1] 彼は部下に仕事を押し付ける素晴らしい上司だ。

例 1 においては、「素晴らしい」という肯定表現が用いられているが、上司への批判が示されている。文章中の極性表現は、

主に文章の内容が肯定的か否定的か分類するタスクとして研究されている感情分析における重要な特徴である。しかし、皮肉では、表層と意図が一致しないため、感情分析の誤分類の原因となる。よって、近年、より高度な文章理解のための挑戦的課題として、皮肉を検出する研究が進められている [1]。

皮肉には、皮肉によって批判される対象が存在するという特徴がある [2]。例 1 においては、「上司」が皮肉の対象である。我々は、以前、皮肉の対象に着目して皮肉の検出を行った [3]。皮肉の対象となりやすい語のリストを作成し、文章中にそのリスト中の語が存在するかを手掛かりとする検出手法を提案した。

(注1) : <https://www.weblio.jp/content/皮肉>

皮肉の対象となりやすい語のリストの作成のために、立場表現を利用した。立場表現は、例 1 中の「上司」のような立場を表す表現である。この立場表現の中から、特に、「上司」、「先生」など、皮肉に出現しやすいものを絞り込んでリスト化した。そして、検出実験により、そのリストに基づく検出手法の有効性を確認した。さらに、その後の研究 [4] で、皮肉の対象と対象に不満を持つ立場という二つの立場の間の関係に着目した検出手法を提案した。例 1 には、皮肉の対象である「上司」に加えて、その対象に不満を持つ「部下」という立場表現が含まれている。このように、二つの立場表現を含むデータについて、関係分類のための手法 [5] を利用し、その立場同士の関係を表すベクトルを構築し、それを検出に取り入れることで、検出精度が向上することを確認した。

これまでの我々の研究 [3], [4] の結果より、文章中に、皮肉の対象となる語が存在するかということは、皮肉検出において重要な手掛かりであるといえる。対象を認識するために、これまでの研究では、立場表現のリストを利用した。立場表現は、「上司」や「部下」といった一般名詞であり、人の役割といった情報を含む。役割の中には、皮肉を言われやすものがあり、皮肉に偏って出現する立場表現があることが考えられる。しかし、立場表現以外の語も、皮肉の対象となることがある。例 2 では、「君」という代名詞が皮肉の対象である。

【例 2】 君はさぼってばかりで忙しそうだね。

このような代名詞は、皮肉の対象ではない場合でも、汎用的に使われるため、リストに含めることができず、皮肉の対象である場合でも認識できない。この他にも、固有名詞が対象を表す語になることが考えられる。固有名詞は、リスト化しても各表現の出現頻度は少なく、網羅もできないことが考えられる。このような表現を対象として認識するには、その表現だけではなく、ある表現が対象となる場合の前後の表現のパターンなどの文脈を考慮する必要がある。

本研究では、皮肉検出の精度向上に貢献することを目的として、皮肉の対象を検出するタスクに取り組む。まず、皮肉の対象の検出のためのデータセットを構築する。次に、そのデータを用いて、機械学習による対象の検出モデルを構築し、文章中の各語が皮肉の対象であるかを判定する。機械学習のアルゴリズムには、Conditional Random Fields (CRFs) [6] と、リカレントニューラルネットワーク (RNN) [7] を用いて、文章全体を考慮した皮肉の対象の検出を行う。

2. 関連研究

近年、皮肉の検出が盛んに研究されている。皮肉の検出は、主に、文章が皮肉を含むかどうか分類するタスクとして取り組まれている。多くの研究で、機械学習を用いた分類手法が提案されている。機械学習のための素性には、文章中の単語 N グラム [8] や、極性表現 [9], [10] などが用いられている。

Tay ら [11] は、単語間の関係性を利用する検出手法を提案し、皮肉検出の最高精度を達成した。ニューラルネットワークによって、文章中の単語のペアごとの検出における重みを学習し、重要な単語ペアを考慮した検出を行った。我々は以前の研究 [4]

で、「上司と部下」のような二つの立場の間の関係に着目した検出手法を提案している。これは、Tay らの手法に、特に対象に注目するという視点を加えたものであると位置づけられる。

これまでの研究 [3], [4] で、皮肉の対象の存在という特徴が、皮肉検出に有用であるとわかった。これを踏まえ、本研究では、文章中の皮肉の対象を表す語の検出に取り組む。これに関連したタスクとして、文章中の人物名や組織名などの抽出を目的とした固有表現抽出がある。固有表現抽出は、文章を単語の系列と見なし、系列中のどの部分が固有表現であるかを判定する枠組みである。系列ラベリング問題として取り組まれることが多い。判定のためには、機械学習を利用した手法が用いられ、特に、系列を学習することのできる、CRFs モデルや RNN モデルが利用されている [12], [13]。本研究では、皮肉の対象の検出のために、CRFs モデルや RNN モデルを適用し、リストに基づく手法と比較して、その検出精度を確認する。

3. データセットの構築

本節では、皮肉の対象の検出に利用するデータセットの構築について説明する。皮肉の対象の検出は、皮肉を含む文章が与えられたとき、皮肉を言われている対象を検出するタスクである。皮肉を含む文章に、人手で皮肉の対象をアノテーションすることで、対象検出のためのデータセットを構築する。

以降の節で、まず、アノテーション対象のデータについて説明し、次に、データ中の皮肉の対象の種類について説明する。最後に、皮肉の対象のアノテーションの手順と結果について説明する。

3.1 皮肉を含んだ文章の収集

皮肉の検出の研究には、Twitter^(注2)への投稿を検出の対象にしたものが多い。本研究でも、Twitter への投稿から皮肉を含んだものを収集し、アノテーション対象のデータとして利用する。皮肉を含んだ投稿の収集のために、多くの研究では、投稿者が投稿につけるハッシュタグが利用されている [8]~[10]。ハッシュタグはその投稿の内容を表すタグである。英語の投稿においては、皮肉を表す「#sarcasm」というハッシュタグが付いた投稿を収集することで、皮肉を含む投稿が収集できる。本研究では、日本語の投稿をデータとして利用する。よって、同様に、日本語の投稿で、「#皮肉」というタグが付けられている投稿を収集することを試みた。しかし、「#皮肉」というタグが付けられている投稿の数は十分ではなかった。そこで、「(皮肉)」という表現を利用した。日本語の投稿には、次の例のように、「(皮肉)」という表現を用いて皮肉であることを示すものがある。

【例 3】 うちの会社はミスが発生してもお咎め無しのとても素敵な (皮肉) 会社です。

この「(皮肉)」という表現を含む投稿は、「#皮肉」というタグが付いた投稿より頻繁に投稿されていた。よって、「(皮肉)」という表現を含む投稿を 1,000 件の投稿を収集し、皮肉を含むデータ (皮肉ツイート) とみなした。

(注2) : <https://twitter.com>

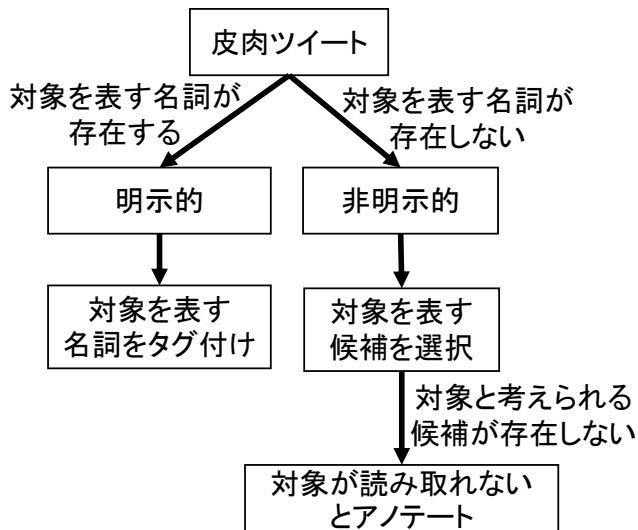


図1 アノテーション手順の概要

3.2 対象の種類：明示と非明示

皮肉ツイートには、文章中に皮肉の対象を表す語を含むものと含まないものがある。本データセットでは、これを明示・非明示とよび区別する。

次の例4, 5は、皮肉の対象を表す語を含む例である。

【例4】口先だけで何もしない政治家の鑑だ。

【例5】彼は毎回大量の宿題を出す教育熱心な先生だ。

例4では、皮肉の対象は「政治家」である。また、例5では、皮肉の対象は「彼」と「先生」である。このように、対象を表す語が複数含まれる場合もある。本研究では、一つ以上の対象を表す語を含むデータを、対象が明示的なデータと呼ぶ。アノテーションにおいては、すべての対象にタグ付けする。

次の例6, 7は、皮肉の対象を表す語を含まない例である。

【例6】何も考えずに決めるなんて、ある意味すごいよ。

【例7】開始早々メンテとか幸先がいいですね。

本研究では、このようなデータを、対象が非明示的なデータと呼ぶ。ただし、このような場合に関しても、例6は人を対象としており、例7は何らかのサービスを対象としているという区別をすることができる。皮肉の対象をリストに基づいて認識した関連研究[3]では、本文中に明示的に皮肉の対象が含まれていることが想定されていた。しかし、このように非明示的な場合にも、対象に関する情報を利用できれば、対象の存在を利用した皮肉検出手法の適用できるデータが増え、検出精度の向上に貢献できると考えられる。本アノテーションでは、非明示なデータに対しても、それに対応する対象を付与する。

3.3 皮肉の対象のアノテーション

3.1節で収集した皮肉ツイートに対して、皮肉の対象のアノテーションを行った。このアノテーションの手順を説明する。アノテーションの概要を図1に示す。

まず、アノテータには、皮肉ツイート本文と、本文中の名詞の一覧が提示される。ツイート中の名詞の抽出には、形態素解析

のMeCab^(注3)を用いた。アノテータは、本文中の名詞に、皮肉の対象を表しているものがあるかを判断する。あれば、対象が明示的なツイートであり、なければ、対象が非明示的なツイートであると判断する。次に、明示的な場合と、非明示的な場合のそれぞれのアノテーションについて説明する。

まず、対象が明示的な場合について説明する。明示的であると判断した場合、名詞中の、対象を表すものを選択し、その名詞に皮肉の対象であるとタグ付けする。このとき、皮肉の対象であると考えられる名詞が複数ある場合、そのすべてを対象とタグ付けする。例えば、例5では、「彼、毎回、大量、宿題、教育、熱心、先生」という名詞リストが得られるが、この中から、対象を表す「彼」、「先生」に、対象であるとタグ付けする。

次に、対象が非明示的な場合について説明する。例6, 7のように、皮肉の対象が非明示的であるが、読み取ることができる場合を想定し、皮肉の対象の6種類の候補のリストを準備した。今回は、「人、団体(国や組織等)、動物、もの、場所、サービス」という候補リストを準備した。皮肉の対象が暗示的である場合、アノテータは、このリストから皮肉の対象として読み取れるものを選択する。適切な対象がリストに含まれていない場合、アノテータは、対象が読み取れないツイートとみなして、そのことを示すタグを付与する。

以上の手順によって、各皮肉ツイートに対して、皮肉の対象をアノテーションした。収集した1,000ツイートを、500ツイートずつの2セットに分け、各セットについて、異なるアノテータ2名ずつにアノテートを依頼した。アノテーションの結果、2名のいずれかが、対象が読み取れないとアノテートしたツイートは1,000件中、合計134件であった。以降、このデータを除外してアノテーションの結果を分析する。

まず、明示と非明示の判断の結果について説明する。2名のアノテータの両方が、対象が明示的であると判断したツイート数は、合計547件であり、対象が非明示的であると判断したツイート数は、合計146件であった。明示か非明示かの判断の一致率を確認した結果、2つのデータセットにおける κ 値の平均は、0.50であった。Landisら[14]らの指標によると、0.50は中等度の一致(Moderate agreement)である。

明示で一致した547件のツイートについて、各アノテータは、どの名詞が対象を表しているかをタグ付けした。この547件のツイートには、11,457単語が含まれており、そのうち3,803語が名詞であった。その名詞のうち、2名のアノテータの両方が、皮肉の対象であるとタグ付けした名詞は、592語であった。各アノテータがタグ付けした名詞の一致率を確認した結果、2つのデータセットにおける κ 値の平均は、0.65であった。Landisら[14]らの指標によると、0.65は十分な一致(Substantial agreement)である。2名のアノテータの両方が、タグ付けした592語は、466件ツイートに含まれていた。本研究では、この466件のツイートを明示的に皮肉の対象を含むツイートとして扱う。

次に、非明示で一致した146件のツイートについてそれぞれ、各アノテータは、候補リストから対象となるものを選択した。

(注3) : <http://taku910.github.io/mecab/>

表 1 各候補に対するツイート数

	人	団体	動物	もの	場所	サービス
ツイート数	97	10	0	2	0	3

どの候補が選ばれたかの一致率を確認した結果、2つのデータセットにおける κ 値の平均は、0.51であった。Landisら[14]らの指標によると、0.51は中等度の一致(Moderate agreement)である。2名が皮肉の対象として選択した候補が一致したツイート数は、合計112件であった。本研究では、この112件のツイートを対象が非明示的なツイートとして扱う。また、リスト中のどの候補で一致したかを確認し、各候補に対するツイート数を表1に示した。112件のうち、97件が「人」で一致しており、偏りの大きい結果となった。

4. 皮肉の対象の検出

本節では、皮肉の対象の検出実験について説明する。我々の以前の研究[3]では、本文中に皮肉の対象が明示的に含まれていることを想定し、立場表現リストを用いた語のマッチングにより、対象を認識している。一方、本研究においては、前節で作成したデータ用いた機械学習による対象の検出を行う。本稿では、対象が明示的な場合について、リストに基づく手法と、機械学習による手法によって検出できる対象の違いに分析の焦点をおき、機械学習による検出の有用性を評価する^(注4)。本節では、まず、機械学習による皮肉の対象の検出手法と、リストに基づくベースライン手法を説明する。次に、検出手法の評価のための実験とその結果について説明する。

4.1 皮肉の対象の検出手法

皮肉の対象の検出手法を説明する。ツイート中のどの単語が皮肉の対象であるかを検出する。我々は、このタスクを系列ラベリングのタスクであるとみなし、CRFsモデルとRNNモデルをそれぞれ利用した検出を行う。

CRFsモデルによる手法は系列ラベリングにおいて、一般的な手法の一つである。素性としては、前後二単語までの単語の表層と品詞を利用した。また、実装には、CRF++^(注5)を利用した。

RNNモデルについては、特に、Gated Recurrent Unit (GRU)[15]を用いる。GRUも多くの研究で用いられているモデルである。このモデルは、文章の各単語の情報を順に入力として、その入力それぞれに皮肉の対象であるかどうかを出力する。各単語の情報は、分散表現と品詞情報を連結したベクトルである。分散表現は、word2vec[16]のskip-gramモデルによって作成した、200次元のベクトルである。分散表現の作成のためのデータには、ランダムに収集した約2,300万件のTwitterへの投稿を利用した。また、品詞情報は、どの品詞であるかをone-of-k表現で表したベクトルである。

(注4)：対象が非明示な場合の検出も皮肉の検出精度向上に貢献することが考えられるが、本稿では扱わない。

(注5)：<https://taku910.github.io/crfpp/>

表 2 データセットの詳細

	ツイート数	対象を表す単語数	総単語数
訓練	419	523	10,555
評価	47	69	902
合計	466	592	11,457

表 3 実験結果

	P	R	F
ベースライン	0.50	0.22	0.30
CRFs	0.71	0.25	0.36
RNN	0.04	0.01	0.02

4.2 ベースライン

我々の以前の研究[3]では、文章中の皮肉の対象を、立場表現のリストに基づいて認識し、皮肉検出に利用した。本研究では、リストで網羅できない皮肉の対象を検出することを目指している。よって、以前の研究[3]のリストに基づく簡単なベースライン手法を設定した。ベースライン手法は、文章中の各単語について、リストに含まれていれば、その単語を皮肉の対象であると判定し、リストに含まれていなければ、皮肉の対象でないと判定する。リストには、以前の研究[3]で利用したものをそのまま利用する。このリストには、「にとっては」などの手掛かり表現を用いて収集された立場表現のうち、皮肉データにおいて出現頻度の高いものを選択して得られた472語の立場表現が含まれている。

4.3 データセット

次に、検出手法の評価実験に利用するデータについて説明する。3.3節で構築したデータセットうち、本文中に皮肉の対象を表す語を明示的に含む466件のツイートを利用する。この466件のツイートを、419件の訓練用データと47件の評価用データに分けて利用する。各データセットについての詳細を表2に示した。

4.4 実験結果・考察

各単語が皮肉の対象であるかどうかを正しく判定できたかを評価した。その結果を表3に示した。表中のP, R, Fはそれぞれ、適合率、再現率、F値を示している。

ベースライン、CRFs、RNNのそれぞれのF値が、0.30、0.36、0.02となっており、RNNモデルを利用した検出の精度が、ベースライン手法やCRFsを利用した検出精度を大きく下回った。また、RNNモデルでの検出の適合率である0.04は、すべての単語を対象であると判定した場合の適合率である69/902 = 0.08を下回っている。これは、モデルが複雑なものであるのに対し、利用した訓練データが419件と小規模であったためであると考えられる。また、CRFsモデルを利用した検出の精度が、リストに基づくベースラインの精度をすべての指標で上回った。よって、CRFsを利用した検出では、リストでは網羅できていなかった皮肉の対象を正しく検出できていることがわかる。

次に、検出精度の高かったCRFsを利用した手法で検出される皮肉の対象について、リストに基づくベースライン手法と比較して、考察を行う。両方の手法で、検出できた対象について

説明する。次のような例において、「客」という皮肉の対象を表す語は、両方の手法で、正しく検出された。

【例 8】 予約している商品を取りに来ない熱心な客がいるらしい。

ベースライン手法においては、「客」が立場表現の一つであり、リストに含まれているため、検出された。

一方、ベースライン手法では、皮肉の対象を検出できず、CRFs を利用した手法では検出できた例に次のようなものがあった。

【例 9】 Twitter のおかげで映画のネタバレが聞けた。

例 9 の皮肉の対象は、「Twitter」である。ベースライン手法では検出できず、CRFs を利用した手法では検出できた対象のほとんどはこのような固有名詞であった。ベースライン手法の用いたリストは、立場表現を集めたものであり、リスト中の各立場表現は、一般名詞である。よって、「Twitter」のような固有名詞を検出することはできない。

しかし、固有名詞であるからといって、皮肉の対象であるわけではない。CRFs を利用した手法で対象を検出できなかった例に次のようなものがあった。

【例 10】 進撃の巨人 2 期か〜こっちの巨人は最下位に向けて進撃中ですね・・・ #giants

このツイートには、「進撃の巨人」という固有名詞が含まれているが、太字の「巨人」という語のみが皮肉の対象である。今回、CRFs モデルの学習には、単語の表層と品詞という手掛かりのみを利用した。しかし、このような例を正しく検出するためには、「進撃中」のような皮肉として使われている肯定的な表現のかかり元を考慮することや、「#giants」というハッシュタグが示すような、そのツイートの中心となっているトピックを考慮することが必要であると考えられる。

以上の結果より、一般名詞である立場表現のみに絞ったリストに基づく手法に対して、CRFs モデルを用いた手法では、立場表現だけでなく、固有名詞などもベースラインを上回る精度で検出できたため、CRFs モデルを用いた手法は、皮肉の対象の検出に有用であるといえる。そして、今後は、CRFs モデルの学習のための素性の改良が必要であると考えられる。

5. おわりに

本研究では、皮肉検出の精度向上に貢献することを目的として、皮肉の対象の検出に取り組んだ。まず、日本語の皮肉を含んだ Twitter への投稿を収集し、人手でのアノテーションを行い、皮肉の対象の検出のためのデータセットを構築した。次に、構築したデータセットに、皮肉の対象検出のための手法を適用し、評価実験を行った。その結果、関連研究 [3] で利用された立場表現リストに基づくベースラインの検出精度を、CRFs モデルを利用した検出手法の検出精度が上回った。さらに、CRFs を利用した手法とベースライン手法で検出される皮肉の対象の差について考察をおこない、提案手法の有用性を確認した。

今後の課題としては、文章中の皮肉の対象の検出について、データセットの拡張や、そのデータを用いたより複雑なモデルの適用が挙げられる。また、対象が明示的でないデータにおける対象の推定にも今後取り組んでいく。非明示なデータでは、

表 1 で示した通り、データ数が、「人」という候補を対象とするデータに偏っている。よって、データ数が比較的少ないクラスについても、正しく推定できる手法を検討する必要がある。そのための手掛かりとして、皮肉として使われている肯定表現が、どのような対象に使うものかという知識が考えられる。例えば、「名作」という表現が使われていれば、対象は「もの」であることが考えられる。よって、知識に基づくルールベースの判定を取り入れた手法などを検討する。

文 献

- [1] A. Ghosh, G. Li, T. Veale, P. Rosso, E. Shutova, J. Barnden, and A. Reyes, “Semeval-2015 task 11: Sentiment analysis of figurative language in twitter,” Proceedings 9th Int. Workshop on Semantic Evaluation (SemEval2015), Co-located with NAACL, pp.470–478, 2015.
- [2] J.D. Campbell and A.N. Katz, “Are there necessary conditions for inducing a sense of sarcastic irony?,” Language Resources and Evaluation, vol.49, no.6, pp.459–480, 2012.
- [3] S. Hiai and K. Shimada, “Sarcasm detection using features based on indicator and roles,” Proceedings of SCDM 2018, pp.418–428, 2018.
- [4] 肥合智史, 嶋田和孝, “関係ベクトルを利用した皮肉の検出,” 言語処理学会 第 24 回年次大会 発表論文集, pp.829–832, 2018.
- [5] K. Hashimoto, P. Stenetorp, M. Miwa, and Y. Tsuruoka, “Task-oriented learning of word embeddings for semantic relation classification,” Proceedings of CoNLL 2015, pp.268–278, July 2015.
- [6] J.D. Lafferty, A. McCallum, and F.C.N. Pereira, “Conditional random fields: Probabilistic models for segmenting and labeling sequence data.,” Proceedings of ICML 2001, pp.282–289, 2001.
- [7] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Comput., vol.9, no.8, pp.1735–1780, Nov. 1997.
- [8] A. Reyes, P. Rosso, and T. Veale, “A multidimensional approach for detecting irony in twitter.,” Language Resources and Evaluation, vol.47, no.1, pp.239–268, 2013.
- [9] E. Riloff, A. Qadir, P. Surve, L. De Silva, N. Gilbert, and R. Huang, “Sarcasm as contrast between a positive sentiment and negative situation.,” Proceedings of EMNLP 2013, pp.704–714, 2013.
- [10] A. Joshi, V. Sharma, and P. Bhattacharyya, “Harnessing context incongruity for sarcasm detection.,” Proceedings of ACL-IJCNLP 2015, pp.757–762, 2015.
- [11] Y. Tay, A.T. Luu, S.C. Hui, and J. Su, “Reasoning with sarcasm by reading in-between,” Proceedings of ACL 2017, pp.1010–1020, 2018.
- [12] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, and C. Dyer, “Neural architectures for named entity recognition,” Proceedings of NAACL-HLT 2016, pp.260–270, 2016.
- [13] X. Ma and E. Hovy, “End-to-end sequence labeling via bi-directional lstm-cnns-crf,” Proceedings of ACL 2106, pp.1064–1074, 2016.
- [14] J.R. Landis and G.G. Koch, “The measurement of observer agreement for categorical data,” Biometrics, vol.33, no.1, pp.159–174, 1977.
- [15] K. Cho, B. vanMerriënboer, Ç. Gülçehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning phrase representations using rnn encoder–decoder for statistical machine translation,” Proceedings of EMNLP 2014, pp.1724–1734, Oct. 2014.
- [16] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality,” Advances in Neural Information Processing Systems 26, pp.3111–3119, 2013.