Automatic Humor Classification on Twitter

Yishay Raz
Shanghai Jiao Tong University
Department of Computer Science & Engineering
800 Dongchuan Road
Shanghai, 200024, China
yishayraz@yahoo.com

Abstract

Much has been written about humor and even sarcasm automatic recognition on Twitter. The task of classifying humorous tweets according to the type of humor has not been confronted so far, as far as we know. This research is aimed at applying classification and other NLP algorithms to the challenging task of automatically identifying the type and topic of humorous messages on Twitter. To achieve this goal, we will extend the related work surveyed hereinafter, adding different types of humor and characteristics to distinguish between them, including stylistic, syntactic, semantic and pragmatic ones. We will keep in mind the complex nature of the task at hand, which emanates from the informal language applied in tweets and variety of humor types and styles. These tend to be remarkably different from the type specific ones recognized in related works. We will use semi-supervised classifiers on a dataset of humorous tweets driven from different Twitter humor groups or funny tweet sites. Using a Mechanical Turk we will create a gold standard in which each tweet will be tagged by several annotators, in order to achieve an agreement between them, although the nature of the humor might allow one tweet to be classified under more than one class and topic of humor.

1 Introduction

The interaction between humans and machines has long extended out of the usability aspect. Nowadays, computers are not merely a tool to extend our lacking memory and manpower, but also serve a larger role in communications, entertainment and motivation. These may be found in such systems as Chatterbots, gaming and decision making. Humor is extremely important in any communicative form. It affects not only feelings but also influences human beliefs. It has even shown to encourage creativity. Enabling a machine to classify humor types (and topics) can have many practical applications, such as automatic humor subscriptions that send us only those messages that will make us laugh. It can serve as a basis for further research on humor generation of witty and adequate replies by conversational agent applications. We tend to expose more about ourselves in humor than in regular prose. In the next section we will highlight several research results from the fields of psychology and sociology that show this, and explore the differences in humor produced by different groups. This knowledge can be used to identify the latent attributes of the tweeters, e.g. gender, geographical location or origin and personality features based on their tweets. Aggressiveness in humor can be viewed as a potential warning sign and teach us about the authors mental well-being.

We will now look at some examples of funny tweets from one of the sites, and then review the different types, topics and the way in which the human brain operates to get the joke. We will also see how computers can imitate this:

1. ”And he said unto his brethren, A man shall not poketh another man on facebook for thine is gayeth” #lostbibleverses

2. if life gives you lemons, make someone’s paper

Proceedings of the NAACL HLT 2012 Student Research Workshop, pages 66–70,
Montréal, Canada, June 3-8, 2012.

cut really sting

3. Sitting at a coffee bean and watching someone get arrested at the starbucks across the street. True story.

4. One of Tigers mistresses got 10 million dollars to keep quiet. I gotta admit I'm really proud of that whore.

5. There is a new IPod app that translates Jay Leno into funny.

6. May the 4th be with you...

Example (1) has a hashtag that could help us understand the special stylistic suffixes some words in the sentence bear. Googling the first part yields more than 2 million hits, since this is a common biblical verse. This makes it a wordplay joke that paraphrases a known phrase. But the main reason this is funny is the observation that a very common Facebook action is gay. Therefore, the type of this humor would be classified as observational and the topic Facebook. The latter could be observed by a computer if we allow it to recognize the named entity facebook, which in many cases would serve as the topic. The type, which we recognize as gay, will appear in our lexicon. Since it appears after a copula, we can infer that this is not a regular gay joke. If it was an outing tweet it would not be funny. For both processes, we require a part of speech tagger and a NE recognizer. We can find these two tools at tt http://www.cs.washington.edu/homes/aritter/, developed especially for Twitter by Alan Ritter. Example (2) has no NE or any special lexicon word associated with it. A Google search of the first part of the sentence, within the quotes, will yield 639,000 results. So we can infer it is of wordplay type. But why is it funny? The topic is human weakness, as described by Mihalcea (2006). We laugh at the manifestation of human misanthropy and the satisfaction in gloating. This relates to the relief theory of humor, as the joke is allowing us to speak about our tabooed and unsocial feelings. How can the computer understand this? It is a tricky and complex task. We could parse the sentence to find out that the reader is advised to make someones cut sting, and we could use a semantic ontology or a lexicon to teach the computer that sting is a negative experience, which will lead to drawing the correct conclusion. We believe a comprehensive understanding of the sentence is not mandatory, but if necessary, we can use the work of Taylor (2010) as reference. Example (3) ends with the short sentence true story, which tells us that this is an anecdote. The present progressive tense of the verbs implies the same. To understand this short sentence we need a semantic effort, or a lexicon of such terms that confirm the anecdotal nature of the tweet. The NE Starbucks could be set as the viable topic. Example (4) has a proper noun as NE, Tigers, recognized by its capital first letter. This is also the topic, and the type is probably vulgarity, that can be recognized by the last word in it. Example (5) is an insult, and the topic is the proper name Jay Leno. This research will likely conclude that we prefer the human NE over the non-human one, when instructing the computer how to choose our topic. To recognize that this is an insult to Leno, we need to know he is a comedian, and that the tweet suggests that he is not funny. An internet search will discover the former. For the latter, we must understand what translate something into funny means. The semantics of the verb and its indirect object that follows the preposition into should clarify this. This can be achieved by parsing the tweet, looking up the semantics of translate and comedian in a semantic ontology, and concluding that Leno is not funny. This is contradictory to his profession and can be viewed as an insult. Example (6) is a pun, or a wordplay, in taxonomy of Hay (1995). No topic. The pun is based on the phonologic resemblance of forth and force and the immortal quote from Star Wars. According to Wikipedia, May 4th is actually an official Star Wars day because of this pun, and an internet search can teach our computer what type of tweet this is. Alternatively, with more original phonological puns, phonologic ontologies (which have not been researched thoroughly) can be a proper reference source.

The remainder of the paper is organized as follows: related work is reviewed in section 2 . Section 3 briefly describes the data used in the experiments and evaluates the results. Section 4 describes the task and algorithm of humor classification and section5 gives ideas for further research.

2 Related Work

We will survey the research work related to our thesis in 4 different points of reference.

2.1 Humor Recognition

While the classification of different data, identifying whether tweets are humorous, sarcastic, or neither, has been examined closely in recent years, I am unaware of any research that has been done on automatic humor classification by type or topic. One of the first studies on computational humor was done by Binsted and Ritchie (1997), in which the authors modeled puns based on semantics and syntax. This work paved the way for humor generation research works, such as LIBJOG (Raskin and Attardo 1994), JAPE (Binsted and Ritchie 1994, 1997) and HAHAcronym (Stock and Strapparava, 2003). The two former systems were criticized as pseudo-generative because of the template nature of their synthesized jokes. The latter is also very limited in its syntax. Only in later studies was the recognition of humor examined. Mihalcea and Strapparava (2005) used content and stylistic features to automatically recognize humor. This was done, however, on a more homogenous set of data, one-liners, that, unlike tweets, are formal, grammatically correct and often exhibit stylistic features, such as alliteration and antonyms, which seldom appear in tweets. Davidov et al. (2010) recognized sarcastic sentences in Twitter. They used a semi-supervised algorithm to acquire features that could then be used by the classifier to decide which data item was sarcastic. In addition to these lexical patterns, the classifier also used punctuation-based features (i.e. number of !). This procedure achieved an F-score of 0.83 on the Twitter dataset and the algorithm will be carefully examined in my research.

2.2 Humor Theories

There are three theories of humor mentioned in related works: the incongruity theory, the superiority theory and the relief theory. The incongruity theory suggests that the existence of two contradictory interpretations to the same statement is a necessary condition for humor. It was used as a basis for the Semantic Script-based Theory of Humour (SSTH) (Raskin 1985), and later on the General Theory of Verbal Humour (GTVH) (Attardo and Raskin 1991). Taylor (2010) found that the semantic recognition of humor is based on this theory and on humor data that support it. We can see that examples (1)-(5) in section 1 do not comply with this theory. It appears that some humorous statements can lack any incongruity.

The superiority theory claims that humor is triggered by feelings of superiority with respect to ourselves or others from a prior event (Hobbes 1840).

The relief theory views humor as a way out of taboo and a license for banned thoughts. Through humor the energy inhibited by social etiquette can be released and bring relief to both the author and audience. Freud, as early as 1905, supported this theory and connected humor to the unconscious (Freud, 1960). Minsky (1980) embraces the theory and observes the faulty logic in humor as another steam-releasing trait. Mihalcea (2006) enumerated the most discriminative content-based features learned by her humor classifier. The more substantial features were found to be human-centric vocabulary, professional communities and human weaknesses that often appear in humorous data. We think these features of humor, more than the three theories mentioned above, will be of greatest value to our task.

2.3 Humor Types

We will then explore what research has been performed on the actual content and types of humor, aside from the computer recognition point of view. There are many taxonomies of humor (Hay, 1995), and the one that best suits our data contains the following categories:

1. Anecdotes
2. Fantasy
3. Insult
4. Irony
5. Jokes
6. Observational
7. Quote
8. Role play
9. Self deprecation

10. Vulgarity

11. Wordplay

12. Other

We believe that most of our humorous tweets will fall into one of the first 11 categories.

3 Data

Our task is to categorize the different humorous tweets. A little about Twitter: Twitter is a popular microblogging service with more than 200 million messages (tweets) sent daily. The tweet length is restricted to 140 characters. Users can subscribe to get all the tweets of a certain user, and are hence called followers of this user, but the tweets are publically available, and can be read by anyone. They may be read on the Twitter website, on many other sites, and through Twitter API, an interface that allows access to a great amount of tweets and user attributes. Aside from text, tweets often include url addresses, references to other Twitter users (appear as ¡user¿) or content tags (called hashtags and appear #¡tag¿). These tags are not taken from a set list but can be invented by the tweeter. They tend to be more generic since they are used in Twitters search engine to find tweets containing the tag. Our humorous tweet dataset is derived from websites such as `http://www.funny-tweets.com` that publish funny tweets, and can be further expanded by subscribing to all tweets by comedians who appear on these sites. Another option is a thorough check of tweets of Twitter Lists like ComedyWorld/ and features comedians who send messages to all of their followers.

3.1 Evaluation

To evaluate our results we must find out which type and topic of humor every classified tweet belongs to. We are spared from the challenging task of deciding whether a tweet is funny or not, since all of our data was already deemed funny by the publishing sites. Categorizing humor is of course very complex, due to the fuzzy nature of the taxonomy and the subjectivity of this task. One tweet can be related to more than one topic, and belong to more than one humor type. Nevertheless, the only way to achieve a gold standard for such classification is through human annotation, which can be accomplished through the use of a mechanical Turk.

4 Humor Classification

We will use a semi-supervised algorithm with a seed of labeled tweets as input. This will produce a set of distinguishing features for the multi-class classifier. A few feature types will be examined: syntactical, pattern-based, lexical, morphological, phonological and pragmatic. Here are some examples which refer to the task of classifying the examples given in section 1:

Syntactic Features

- transitiveness of the verb
- syntactic ambiguity

Pattern-based Features

- Patterns including high-frequency and content words as described in the algorithm in Davidov and Rappoport (2006)

Lexical Features

- Lexicon words like Gay
- Existence of NEs (like Facebook and Starbucks)
- Meaning of the verb and its objects(make someones cut sting)
- Lexical ambiguity

Morphological Features

- The tense of the verbs in the tweet
- Special word morphology (like the biblical eth suffix in our example (1))

Phonological Features

- existence of a word that appears on a homophones list (which could help with pun recognition)

Pragmatic Features

- Thee amount of results obtained from a search engine query of the tweet of the verbs in the tweet

Stylistic Features

- Existence of smiley characters
- Punctuation, like !

The topic of a tweet will also be retrieved from automatically retrieved features when it does not appear as a NE in the tweet.

5 Future Work

Further research could be done to classify the tweeters of the humorous tweets based on attributes of gender, age, location, etc. This could be achieved using the type and the topic of the tweets as additional features to semi-supervised classifiers. This idea was inspired by related work that found a correlation between humor and gender. In the Gender and Humor chapter of her thesis, Hay (1995) surveyed old research that claimed women are less inclined towards humor than men. Freud (1905) claimed women do not need a sense of humor because they have fewer strong taboo feelings to repress. This perception is slowly changing, with more contemporaneous work claiming that humor is different between genders. Hay concludes that:

- men are more likely to use vulgarity and quotes than women
- women are more likely to use observational humor

To a lesser degree:

- men tend to use more role play and wordplay
- women are more likely to use jocular insults

We did not find any relevant correlation studies between age, origin, and other attributes with humor, but such research has likely been explored.

References

Attardo, S., Raskin, V. 1991. *Script theory revisited: Joke similarity and joke repre- sentation model. Humor: International Journal of Humor Research 4*, 3-4.

Cheng, Z., Caverlee, J., Lee, K. 2010. *You Are Where You Tweet: A Content-Based Approach to Geolocating Twitter Usersl*. Proceeding of the ACL conference 2010

Davidov, D., and Tsur, O. 2010. *Semi-Supervised Recognition of Sarcastic Sentences in Twitter and Amazon* Computational Linguistics, July, 107-116.

Freud, S. 1905. *Der Witz und seine Beziehung zum Unbewussten*

Freud, S. 1960. *Jokes and their relation to the unconscious* International Journal of Psychoanalysis 9

Hay, J. 1995. *Gender and Humour: Beyond a Joke*. Master thesis.

Hobbes, T. 1840. *Human Nature in English Works*. Molesworth.

Mihalcea, R. 2006. *Learning to laugh automatically: Computational models for humor recognition*. Computational Intelligence, 222.

Mihalcea, R. and Strapparava, C. 2005. *Making Computers Laugh: Investigations in Automatic Humor Recognition*. roceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing

Minsky, M. 1980. *Jokes and the logic of the cognitive unconscious*. Tech. rep., MIT Artificial Intelligence Laboratory.

Pennacchiotti, M. and Popescu, A. 2011. *Democrats , Republicans and Starbucks Afficionados: User Classification in Twitter*. Statistics, 430-438.

Rao, D., Yarowsky, D., Shreevats, A. and Gupta, M. 2010. *Classifying Latent User Attributes in Twitter*. Science, 37-44.

Raskin, V. 1985. *Semantic Mechanisms of Humor*. Kluwer Academic Publications

Solomon, R. 2002. *Ethics and Values in the Information Age*. Wadsworth.

Taylor, J. M. 2010. *Ontology-based view of natural language meaning: the case of humor detection*. Journal of Ambient Intelligence and Humanized Computing, 13, 221-234.

Ziv A. 1988. *National Styles of Humor*. Greenwood Press, Inc.