Modeling Sentiment Association in Discourse for Humor Recognition

Lizhen Liu
Information Engineering
Capital Normal University
Beijing, China
liz_liu7480@cnu.edu.cn

Donghai Zhang
Information Engineering
Capital Normal University
Beijing, China
dhzhang@cnu.edu.cn

Wei Song[*]
Information Engineering
Capital Normal University
Beijing, China
wsong@cnu.edu.cn

Abstract

Humor is one of the most attractive parts in human communication. However, automatically recognizing humor in text is challenging due to the complex characteristics of humor. This paper proposes to model sentiment association between discourse units to indicate how the *punchline* breaks the expectation of the *setup*. We found that discourse relation, sentiment conflict and sentiment transition are effective indicators for humor recognition. On the perspective of using sentiment related features, sentiment association in discourse is more useful than counting the number of emotional words.

1 Introduction

Humor can be recognized as a cognitive process, which provokes laughter and provides amusement. It not only promotes the success of human interaction, but also has a positive impact on human mental and physical health (Martineau, 1972; Anderson and Arnoult, 1989; Lefcourt and Martin, 2012). To some extent, humor reflects a kind of intelligence.

However, from both theoretical and computational perspectives, it is hard for computers to build a mechanism for understanding humor like human beings. First, humor is generally loosely defined. Thus it is impossible to construct rules to identify humor. Second, humor is context and background dependent that it expects to break the reader's common sense within a specific situation. Finally, the study of humor involves multiple disciplines like psychology, linguistics and computer science. Recently, humor recognition has drawn more attention (Mihalcea and Strapparava, 2005;
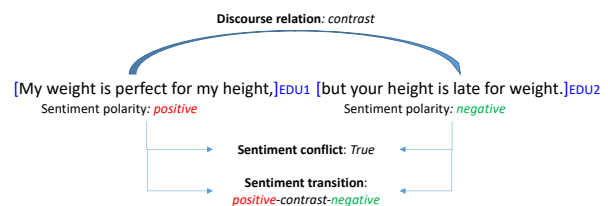

Figure 1: An example of RST style discourse parsing, sentiment polarity analysis and the features we consider in this paper.

Friedland and Allan, 2008; Zhang and Liu, 2014; Yang et al., 2015). The main trend is to design interpretable and computable features that can be well explained by humor theories and easy to be implemented in practice.

In this paper, we propose a novel idea to exploit sentiment analysis for humor recognition. Considering superiority theory (Gruner, 1997) and relief theory (Rutter, 1997), sentiment information should be common in humorous texts to express comparisons between *good* and *bad* or the emotion changes.

Existing work mainly considers statistical sentiment information such as the number of emotional words. We argue that modeling sentiment association at discourse unit level should be a better option for exploiting sentiment information. Such sentiment association in some extent can be used as sentiment patterns to describe the *expectedness* or *unexpectedness*, which is the main idea of incongruity theory (Suls, 1972).

To incorporate discourse information, we exploit RST(Rhetorical Structure Theory) style discourse parsing (Mann and Thompson, 1988) to get discourse units and relations. Combining with sentiment analysis, we derive discourse relation, sentiment conflict and sentiment transition fea-

---

[*]corresponding author

tures for humor recognition as shown in Figure 1. The experimental results show that our method can improve the performance of humor recognition on the dataset provided in (Mihalcea and Strapparava, 2005) and exploiting sentiment information at discourse unit level is a better option compared with simply using the number of emotional words as features.

2 Humor Recognition

Humor recognition is typically viewed as a classification problem (Mihalcea and Strapparava, 2005). The main goal is to identify whether a given text contains humorous expressions. Humor is a cognitive process. Thus the interpretability of models is important. Most existing work focuses on designing features motivated by humor theories from different perspectives.

2.1 Humor Theories

The highly recognized theories include superiority theory, relief theory and incongruity theory.

Superiority theory expresses that we laugh because some types of situations make us feel superior to other people (Gruner, 1997). For example, in some jokes, people appear stupid because they have misunderstood an obvious situation or made a stupid mistake.

Relief theory says that humor is the release of nervous energy. The nervous energy relieved through laughter is the energy of emotions that have been found to be inappropriate (Spencer et al., 1860; Rutter, 1997).

Incongruity theory says that humor is the perception of something incongruous, something that violates our common sense and expectations (Suls, 1972) . It is now the dominant theory of humor in philosophy and psychology.

2.2 Baseline Features

Motivated by the humor theories, many researchers design features to describe the characteristics of humor. We mainly follow the recent work of Yang et al. (2015) to build baseline features. The features are summarized as follows.

Incongruity Structure. Inconsistency is considered as an important factor in causing laughter. Following the work of Yang et al. (2015), we describe inconsistency through the following two features:

- The largest semantic distance between word pairs in a sentence

- The smallest semantic distance between word pairs in a sentence

The semantic distance is measured by computing cosine similarity between word embeddings.

Ambiguity. Ambiguity of semantic is a crucial part of humor (Miller and Gurevych, 2015), because ambiguity often causes incongruity, which comes from different understandings of the intention expressed by the author (Bekinschtein et al., 2011). The computation of ambiguity features is based on WordNet (Fellbaum, 2012). We use WordNet to obtain all senses of each word w in an instance s and measure the possibility of ambiguity by computing $\log \prod_{w \in s} num\_of\_sense(w)$, which is used as the value of an ambiguity feature. We also compute the sense farmost and sense closest features as described in (Yang et al., 2015).

Interpersonal Effect. In addition to the commonly used linguistic cues, interpersonal effect may serve an important role in humor (Zhang and Liu, 2014). It is believed that texts containing emotional words and subjective words are more likely to express humor. Therefore, we use the following features based on the resources in (Wilson et al., 2005).

- The number of words with positive polarity

- The number of words with negative polarity

- The number of subjective words

Phonetic Style. Phonetics can also create comic effects. Following (Mihalcea and Strapparava, 2005), we build a feature set which includes alliteration chain and rhyme chain by using CMU speech dictionary[1]. An alliteration chain is a set of words that have the same first phoneme. Similarly, a rhyme chain includes words with the same last syllable. The features are:

- The number of alliteration chains

- The number of rhyme chains

- The length of the longest alliteration chain

- The length of the longest rhyme chain

---

[1]http://www.speech.cs.cmu.edu/cgi-bin/cmudict

KNN. The KNN feature set contains the labels of the top 5 instances in the training data, which are closest to the target instance.

The above five feature sets are denoted as Humor Centric Features(HCF).

Word2Vec Features. Averaged word embeddings are used as sentence representations for classification.

3 Modeling Sentiment Association in Discourse

As described in humor theories and baseline features, emotional words are viewed as important indicators of humorous expressions, which trigger the subjective opinions and sentiment. Previous work only considers the number of words with different sentiment polarity, but ignores the sentiment association in discourse.

Consider the example in Figure 1 again. The first clause expresses a positive sentiment, while the second clause reveals a negative sentiment. The different sentiment polarity forms a kind of contrast or comparison.

Such sentiment association can be explained with main humor theories. For example, superiority theory says humor is the result of suddenly feeling superior when compared with others who are infirm or unfortunate. There are usually two objects, one of the objects is a laugher who feel better than the other, a weak person. The sentiment association between the *perfect weight* and the *late height* highlights such a comparison.

There are also other cases that may have sentiment association between *negatively nervous* and *positively relief* or from *expected sentiment* to *unexpected sentiment*, which can be explained with relief theory (Rutter, 1997) and incongruity theory (Suls, 1972; Ritchie, 1999).

Therefore, sentiment association should be a useful representation to reveal the nature of humor. In this paper, we utilize a discourse parser to get comparable text units and measure sentiment association among them.

3.1 Discourse Parsing

A well-written text is organized by text units which are connected to express the author's intentions through certain discourse relations.

We use the discourse parser implemented by Feng and Hirst (2012) to automatically recognize RST style discourse relations. RST struc-

ture builds a hierarchical structure over the whole text (Mann and Thompson, 1988). A coherent text is represented as a discourse tree, whose leaf nodes are individual text units called elementary discourse unites (EDUs).These independent EDUs can be connected through their relations. The parser can automatically separate a sentence into EDUs and gives discourse relations between EDUs. One of its advantage over others is that it can identify implicit relations, when no discourse marker is given.[2] There are about 77% of sentences in our dataset that don't have explicit connective.

Our goals of using discourse parsing include two aspects: First, we want to investigate whether humorous texts prefer any discourse relations to realize or enhance the effect. Second, EDUs can be used as comparable text units and enable us to measure sentiment association among them. As a result, we derive three types of features.

3.2 Discourse Relation Features

For each instance, we recognize EDUs and the relations connecting them. Then, we design boolean features to indicate the occurrence of discourse relations. The main idea is that some discourse relations such as *contrast* usually indicate a topic transition, which may be used to achieve the effect of unexpectedness.

3.3 Sentiment Conflict Feature

The sentiment conflict we proposed is a specific and descriptive feature to model a kind of incongruity. After dividing an instance into EDUs, we check the sentiment polarity of each EDU using the TextBlob toolkit[3]. The sentiment polarity is either positive, negative or neutral. The sentiment conflict feature is a boolean feature. If there are at least two EDUs and their polarity are opposite (positive vs. negative), the feature is set as *True*.

3.4 Sentiment Transition Features

Besides the heuristically designed sentiment conflict feature, we integrate sentiment polarity and discourse relations. We thought that the expected sentiment might be dependent on the discourse relation. For example, if two clauses have a *sequence* relation, their sentiment polarity may be

---

[2]PDTB-style (Prasad et al., 2008) discourse parsers can be used here as well. But we didn't find proper toolkits that can deal with implicit relations well.

[3]http://textblob.readthedocs.io/en/dev

expected to be the same, while if their relation is *contrast*, their polarity might be different.

For two EDUs with a discourse relation R, we get their sentiment polarity respectively, namely E_1 and E_2, where $E_* \in \{positive, negative, neutral\}$. We design a feature $E_1 \circ R \circ E_2$, where $\circ$ indicates a concatenation operation and E_1 and E_2 are ordered according to the order in which they appear in the instance. For sentence with more than two EDUs, we do this recursively and set a *True* value for every extracted features.

4 Experiments

4.1 Research Questions

We are interested in the following research questions:

- Whether the proposed features are useful for humor recognition?

- Whether the way we manipulate sentiment is more effective compared with previous approaches?

4.2 Settings

We conducted experiments on the dataset used by (Mihalcea and Strapparava, 2005). The dataset contains 10,200 humorous short texts and 10000 non-humorous short texts coming from Reuters titles and Proverbs and British National Corpus(RPBN).

We used the pre-trained word embeddings that are learned using the Word2Vec toolkit (Mikolov et al., 2013) on Google News dataset.[4] We used the implementation of Random Forest in Scikit-learn (Pedregosa et al., 2011) as the classifier. We ran 10-fold cross-validation on the dataset and the average performance would be reported.

4.3 Baselines

- HCF. The method includes the incongruity structure, ambiguity, interpersonal effect, phonetic style features and KNN features.

- HCF w/o KNN. Since KNN features used in HCF are content dependent. We remove KNN features from HCF to have a content free baseline.

	Acc.	P	R	F_1
Base1: HCF	0.787	0.779	0.815	0.797
KNN	0.756	0.733	0.821	0.775
Base2: HCF w/o KNN	0.71	0.706	0.745	0.725
Base3: Word2Vec	0.77	0.775	0.774	0.775
Base4: Base1+Base3	**0.808**	0.81	0.816	**0.813**
Base1+SA	0.799	0.789	0.828	0.808
Base2+SA	0.75	0.747	0.774	0.76
Base3+SA	0.783	0.788	0.787	0.788
Base4+SA	**0.814**	0.812	0.828	**0.82**

Table 1: Humor recognition results. Base1 to Base4 correspond to four baseline settings and SA represents sentiment association features.

- Word2Vec. As described in Section 2.2, this method exploits semantic representations of sentences. It is also content dependent but has better generalization capability.

- HCF+Word2Vec. This method combines HCF and Word2Vec and is the strongest setting as reported in (Yang et al., 2015).

4.4 System Comparisons

Table 1 shows the results, reported with accuracy(Acc.), precision (P), recall (R) and F_1 score. We add sentiment association features (SA) to four baseline settings. In all cases, the performance is improved.

Base2 only uses features that are motivated by humor theories without content features. After adding SA features, Base2 achieves a significant improvement of 4% in accuracy and 3.5% in F_1 score. Since SA features have good interpretability, they complement previous features very well both in theory and practice.

Base1, Base3 and Base4 all consider content features and their performance is significantly better than Base2. However, since the negative instances in the dataset include news titles, it is very likely that the model matches specific topics of the data, rather than capturing the nature of humor. We can see that the KNN method that is based on content similarity only can achieve high scores, which is unreasonable. Even so, SA features still benefit the three baseline settings, although the improvements become small. The results indicate that sentiment association features are useful for humor recognition, especially when domain specific information is not considered.

	Acc.	P	R	F_1
Base2	0.71	0.706	0.745	0.725
Base2-EWC	0.709	0.705	0.742	0.723
Base2-EWC+SA	**0.748**	0.744	0.773	**0.758**
Base4	0.808	0.81	0.816	0.813
Base4-EWC	0.808	0.808	0.818	0.813
Base4-EWC+SA	**0.812**	0.812	0.823	**0.817**

Table 2: Comparing the ways of utilizing sentiment information. Base2 doesn't consider content; Base4 utilizes full information; SA: sentiment association, EWC: emotional word count.

4.5 Comparing with Emotional Word Count

Previous work also considers sentiment information but in a different way. Among interpersonal effect features, the numbers of emotional words are used as features, noted as emotional word count, EWC for short. We want to compare the sentiment association features with EWC.

We compare them in two conditions. First, we replace EWC features with SA features in Base2, which doesn't use content information. Second, we replace EWC features with SA features in Base4, which considers all available information. As shown inTable 2, in both conditions, SA features are more effective, indicating the usefulness of analyzing sentiment polarity at EDU level.

4.6 Discussion of Sentiment Association

Table 3 shows the results of adding individual sentiment association features on the basis of Base2 and Base4. All three features are shown to be useful for humor recognition. Sentiment transition is most useful. Again, when removing content features (Base2), the improvements are large. In contrast, if considering content features (Base4), the improvements become small. This is because the content features are already very strong for distinguishing two classes.

Discourse Relation. By analyzing the data, we found that 79% of the humorous instances contain more than one EDU, while 38% of non-humorous messages contain more than one EDU. This means that humorous texts may have more complex sentence structures. The most frequent discourse relations in humorous data include *condition*, *background* and *Contrast*. In contrast, non-humorous texts contain *same-unit* and *attribution* more. The most discriminative relation is *condition*, which accounts for 4.5% in humorous instances and 2% in non-humorous instances. This may be explained with the incongruity theory, where the

	Acc.	P	R	F_1
Base2	0.71	0.706	0.745	0.725
Base2+DR	0.741	0.737	0.768	0.752
Base2+SC	0.738	0.734	0.764	0.749
Base2+ST	**0.748**	0.743	0.775	**0.759**
Base4	0.808	0.81	0.816	0.813
Base4+DR	**0.813**	0.813	824	**0.818**
Base4+SC	0.811	0.812	0.82	0.816
Base4+ST	**0.813**	0.814	0.823	**0.818**

Table 3: Comparing sentiment association features. Base2 doesn't consider content; Base4 utilizes full information; DR: discourse relation, SC: sentiment conflict, ST: sentiment transition.

setup of the text prepares an expectation for the readers, while the *punchline* breaks the expectation. *Condition* relation is often used to connect the *setup* and the *punchline*.

Sentiment polarity in Humor. According to the automatic sentiment analysis tool we use, 57% of humorous instances have non-neutral polarity, while 47% of non-humorous instances have non-neutral polarity. This means that humor truly has a positive correlation with sentiment polarities and sentiment analysis should be a useful complement to semantic analysis for measuring incongruity. In addition, as we have shown, measuring sentiment at discourse level should be more important. Combining discourse relations and sentiment polarity performs best.

5 Conclusion

In this paper, we have studied humor recognition from a novel perspective: modeling sentiment association in discourse. We integrate discourse parsing and sentiment analysis to get sentiment association patterns and measure incongruity in a new angle. The proposed idea can be explained with major humor theories. Experimental results also demonstrate the effectiveness of proposed features. This indicates that sentiment association could be a better representation compared with simply analyzing the distribution of sentiment polarity for humor recognition.

References

Craig A. Anderson and Lynn H. Arnoult. 1989. An examination of perceived control, humor, irrational beliefs, and positive stress as moderators of the relation between negative stress and health. *Basic and Applied Social Psychology* 10(2):101–117.

T. A. Bekinschtein, M. H. Davis, J. M. Rodd, and A. M. Owen. 2011. Why clowns taste funny: the relationship between humor and semantic ambiguity. *Journal of Neuroscience the Official Journal of the Society for Neuroscience* 31(26):9665.

Christiane Fellbaum. 2012. *WordNet*. Blackwell Publishing Ltd.

Vanessa Wei Feng and Graeme Hirst. 2012. Text-level discourse parsing with rich linguistic features. In *Meeting of the Association for Computational Linguistics: Long Papers*. pages 60–68.

Lisa Friedland and James Allan. 2008. Joke retrieval: recognizing the same joke told differently. In *ACM Conference on Information and Knowledge Management*. pages 883–892.

Charles R Gruner. 1997. *The game of humor: A comprehensive theory of why we laugh.*. Transaction PUblishers.

Herbert M. Lefcourt and Rod A. Martin. 2012. Humor and life stress. *Springer Berlin* .

W Mann and Sandra Thompson. 1988. Rhetorical structure theory : Toward a functional theory of text organization. *Text* 8(3):243–281.

William H. Martineau. 1972. *A Model of the Social Functions of Humor*. The Psychology of Humor.

Rada Mihalcea and Carlo Strapparava. 2005. Making computers laugh: investigations in automatic humor recognition. In *Conference on Human Language Technology and Empirical Methods in Natural Language Processing*. pages 531–538.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. pages 3111–3119.

Tristan Miller and Iryna Gurevych. 2015. Automatic disambiguation of english puns. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing* 1:719–729.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12:2825–2830.

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind K. Joshi, and Bonnie L. Webber. 2008. The penn discourse treebank 2.0. In *International Conference on Language Resources and Evaluation, Lrec 2008, 26 May - 1 June 2008, Marrakech, Morocco*. pages 2961–2968.

Graeme Ritchie. 1999. Developing the incongruity-resolution theory. *Proceedings of the Aisb Symposium on Creative Language* pages 78–85.

Jason. Rutter. 1997. Stand-up as interaction : performance and audience in comedy venues. *University of Salford* 33(4):1 – 2.

Herbert Spencer et al. 1860. The physiology of laughter. *Macmillans Magazine* pages 395–402.

Jerry M. Suls. 1972. A two-stage model for the appreciation of jokes and cartoons: An information-processing analysis. *Psychology of Humor* 331(6019):81–100.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing*. Association for Computational Linguistics, pages 347–354.

Diyi Yang, Alon Lavie, Chris Dyer, and Eduard Hovy. 2015. Humor recognition and humor anchor extraction. In *Conference on Empirical Methods in Natural Language Processing*. pages 2367–2376.

Renxian Zhang and Naishi Liu. 2014. Recognizing humor on twitter. In *ACM International Conference on Conference on Information and Knowledge Management*. pages 889–898.