Exploiting Syntactic Structures 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 recognition is an interesting and challenging task in natural language processing. This paper proposes to exploit syntactic structure features to enhance humor recognition. Our method achieves significant improvements compared with humor theory driven baselines. We found that some syntactic structure features consistently correlate with humor, which indicate interesting linguistic phenomena. Both the experimental results and the analysis demonstrate that humor can be viewed as a kind of style and content independent syntactic structures can help identify humor and have good interpretability.

1 Introduction

Humor, as a human-specific attribute, plays an important role in human communication. In addition to the tremendous help for human social success, humor also has positive and far-reaching effects on human psychology and physical health (Martineau, 1972; Anderson and Arnoult, 1989; Lefcourt and Martin, 1986). In recent years, the development of artificial intelligence has reinforced the human requirement for the intelligence of machines. As one of the qualities that embodies human wisdom, humor has attracted wide interests and attention (Mihalcea and Strapparava, 2005; Friedland and Allan, 2008; Zhang and Liu, 2014; Yang et al., 2015). The establishment of humor understanding mechanism promotes the development of language intelligence.

With the encouragement of exploring the essence of humor, great progress has been made in the research of humor theories in recent decades. Well recognized theories including superiority theory (Gruner, 1997), relief theory (Rutter, 1997) and incongruity theory (Suls, 1972) have been successively put forward, which explain the origin and essence of humorous feelings. Inspired by these theories, many computational methods are designed to model humor and make some achievements (Mihalcea and Strapparava, 2005; Friedland and Allan, 2008; Zhang and Liu, 2014; Yang et al., 2015).

Although many linguistic cue features have been studied, one aspect is often ignored that humor is a kind of style as well. An interesting question is whether we can explain humor from the perspective of styles. In this paper, we attempt to answer this question by exploring the syntactic structures, which are content independent and style related, for recognizing humor in text. The main contributions can be summarized as below:

(1) We extract production rules, dependency relations and statistics of structural elements as features to model humor. The syntactic structures significantly improve the performance of humor recognition.

(2) We demonstrate that some syntactic structures, which are differently distributed in humorous and non-humorous texts, indicate some interesting linguistic phenomena about humor.

2 Humor Recognition

The main goal of humor recognition is to judge whether the given text expresses humor (Mihalcea and Strapparava, 2005). It can be viewed as a typical classification task.

*corresponding author
This work is licenced under a Creative Commons Attribution 4.0 International Licence. Licence details: http://creativecommons.org/licenses/by/4.0/

Proceedings of the 27th International Conference on Computational Linguistics, pages 1875–1883
Santa Fe, New Mexico, USA, August 20-26, 2018.

Inspired by highly-recognized theories, many studies put forward interpretable features to model humor. We build the baseline features by following the work of (Yang et al., 2015). Next, we explain the features in detail.

Incongruity Structure. Incongruity theory (Suls, 1972) is a widely accepted theory. It believes that the core of humor is inconsistency or conflict. Following the work of (Yang et al., 2015), we describe incongruity through the following two features: (1) the largest semantic distance between word pairs in a sentence. (2) the smallest semantic distance between word pairs in a sentence.

Ambiguity. Another important way to create humor is semantic ambiguity (Miller and Gurevych, 2015), in which misunderstandings of author's intentions often produce unexpected feelings (Bekinschtein et al., 2011). In order to model the semantic ambiguity, we use WordNet (Fellbaum and Miller, 1998) to obtain all senses of word w in current text t. Then the probability of producing semantic ambiguity will be calculated through $\log \prod_{w \in t} sense(w)$, where $sense(w)$ indicates the number of the meanings that word w has. In addition, we also compute the sense farmost and sense closest features which are described in (Yang et al., 2015) to measure ambiguity.

Interpersonal Effect. Interpersonal effect is considered to be closely related to humor (Zhang and Liu, 2014). Some studies suggest that the occurrence of emotions and subjectivity increases the possibility of humor, so we use the resources in (Wilson et al., 2005) to build the following features: (1) the number of words with positive and negative polarity. (2) the number of subjective words.

Phonetic Style. According to the description in (Mihalcea and Strapparava, 2005), phonetic is also an important factor to create comedy effects. In this paper, we follow the work of (Mihalcea and Strapparava, 2005) and (Yang et al., 2015) to extract alliteration chains and rhyme chains in text by using CMU speech dictionary[1]. The alliteration chain is a set of words which have the same first phonemes and the words in rhyme chain have the same last syllable. The corresponding features are as follows: (1) the number of alliteration chains and rhyme chains. (2) the length of the longest alliteration chain and rhyme chain.

KNN features. In addition to humor-related features, we also use the semantic similarity as an indicator to build a KNN feature set that contains the labels of the top 5 sentences in the training data, which are closest to the target instance.

As described, most commonly used features are inspired by humor theories and linguistic knowledge. Few work researches syntactic styles of humorous expressions.

3 Exploiting Syntactic Structures for Humor Recognition

Syntactic information has been successfully used for analyzing text styles. We are going to exploit syntactic structures to reveal stylistic characteristics of humor. We use both constituent parsing and dependency parsing and derive features from parsed trees.

Consider the following sentence:

Never go to a doctor whose office plants have died.

The constituent tree is shown in Figure 1(a), while the dependency tree is shown in Figure 1(b). Both are provided by the Stanford parser (Klein and Manning, 2003) .

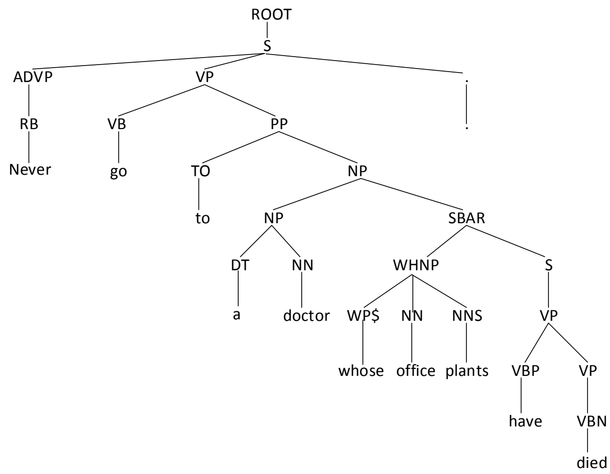
We mainly construct 3 types of features. Such features indicate statistics of basic structural elements and their relations.

3.1 Statistics of Structural Elements

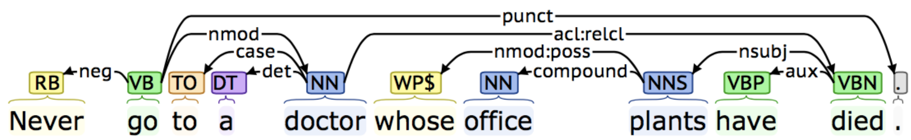
Statistics of structural elements on constituent have been proven to be effective in evaluating the linguistic quality of text (Nenkova et al., 2010). We borrow some features for human recognition.

- Complexity Metrics: Humorous texts and non-humorous texts may differ in the way they express intentions. We expect to measure the differences from the perspective of sentence complexity. Therefore, the number of noun phrases($NPcount$), verb phrases($VPcount$), prepositional

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



(a) a constituent tree.



(b) a dependency tree.

Figure 1: Examples of syntactic parsing.

phrases($PPcount$) and subordinating conjunctions($SBARcount$) are counted respectively as features.

- Phrase Length Ratio: The length ratio(LR) of PP, NP, VP is computed respectively. For each phrase type, the value of its length ratio equals the number of words in the phrase divided by the length of the sentence.

- Average Phrase Length: The average length of a phrase is calculated by dividing the number of words per phrase by the number of corresponding phrase types. It is worth noting that there are two different ways to calculate, one (APL_1) is to consider nested phrases. Consider a VP phrase (VP_1...(...VP_2...)), the length of VP is equal to *length*(VP_1)+ *length*(VP_2). Another way (APL_2) is to only consider the maximum length of the phrase, so the phrase length is *length*(VP_1).

- Normalized Phrase Length: This feature (NPL) just is considered for PP, NP and VP. The value is equal to the average length of each phrase type divided by the length of the sentence. The nested phrases are not considered in this situation.

- Phrase Ratio: Phrase ratio (PR), for each phrase type, is calculated by dividing the number of phrases that appear in the sentence by the length of the sentence.

- Ratio of PP or NP within a VP ($RPNV$): If a VP contains NP or PP, then this value is equal to the average length of the NP or PP divided by the length of the VP.

- Modifiers Change: There are several ways to modify the head noun. This feature is designed to model the modification of the head noun. Finally, two types of feature values are considered. One (MN) is the length of all the modified structures corresponding to the head noun. The other (NMN) is the normalization of the length of the modified structures.

3.2 Production Rules

Production rules describe the composition of phrases at all levels to form a complete sentence. We extract all production rules except the specific words at leaves of each instance as features.

The resulting features, corresponding to the example, include: *ROOT*→*S*, *S*→*ADVP VP .*, *ADVP*→*RB*, *VP*→*VB PP*, *PP*→*TO NP*, *NP*→*NP SBAR*, *NP*→*DT NN*, *SBAR*→*WHNP S*, *WHNP*→WP$ NN NNS, *S*→VP, *VP*→*VBP VP*, *VP*→*VBN*.

For a given production rule a, we set its feature value as $v(a) := count(a)/count(*)$, where $count(a)$ denotes the number of production rule a, and $count(*)$ represents the number of all production rules extracted from the current instance.

3.3 Dependency Relations

Compared with constituent parsing, dependency parsing indicates relation types between words. Thus, we directly design features for dependency relations. We build a feature for a dependency relation and normalize the count of the dependency relation by the number of all dependency relations in the sentence as the feature value.

We also attempt to combine part of speech (POS) tags and dependency relations. For the example shown in Figure 1(b), there is a dependency from *go* to *never* with type *neg*, therefore we construct a feature $VB \circ neg \circ RB$, where VB and RB are the POS tags of *go* and *never* respectively. But such features don't perform as well as using dependency relations only.

4 Experiment and Analysis

4.1 Data

The data includes humorous sentences as positive instances and non-humorous sentences as negative instances. The humorous sentences are from the work by (Mihalcea and Strapparava, 2005).

One-Liner includes10200 short humorous texts, which are crawled from the Web sites that containing the tags of $humor$, $humour$, $oneliner$, one-$liner$, $funny$ and $joke$.

They also provide a set of non-humorous short texts.

RPBN contains about 10000 sentences from Reuters titles, Proverbs and British National Corpus.

We found that the average number of tokens in a sentence is 9 in PRBN and 12 in One-Liner. Since we used statistics of structural elements and relations as features, such bias may affect the evaluation, although we have normalized most features. To reduce artifacts and increase the data diversity, we also extracted sentences from the newswire dataset provided by CoNLL-2012 for the task of coreference resolution (Pradhan et al., 2012). These sentences are samples from the OntoNotes corpus (Hovy et al., 2006). We filtered out the texts shorter than 5 words to reduce data noise. We finally sampled 10000 sentences to form a set of negative instances. Their average sentence length is 16.

We conducted experiments on three datasets. All three datasets use One-Liner as the positive instances. The first data set uses RPBN as negative instances, noted as RPBN as well; the second dataset uses sentences extracted from OntoNotes as negative instances, noted as CoNLL and the third dataset was built by sampling negative instances from both RPBN and CoNLL, half from RPBN and half from CoNLL, to keep a balanced positive and negative instance distribution, noted as Mixed.

Random Forest in Scikit-learn (Pedregosa et al., 2011) is used as the classifier. We ran 10-fold cross-validation on the datasets and the average performance would be reported.

4.2 Baselines

- Humor Theory driven Features (HTF). This method uses all features described in Section 2 except the KNN features. Thus all the used features are motivated by the humor theories. These features don't depend on specific content. We expect to use these features to capture the nature of humor.

- Human Centric Features (HCF). The method is a complete re-implementation of the proposed method in (Yang et al., 2015), which uses all features described in Section 2, i.e., HTF plus the KNN features.

Feature	RPBN				CoNLL				Mixed				
	Acc.	P	R	F_1	Acc.	P	R	F_1	Acc.	P	R	F_1	Average ΔF_1
HTF	0.709	0.702	0.749	0.725	0.801	0.771	0.862	0.814	0.707	0.683	0.793	0.734	—
HCF	0.786	0.777	0.816	0.796	0.908	0.889	0.935	0.911	0.821	0.797	0.87	0.832	—
Word2Vec	0.772	0.777	0.778	0.777	0.894	0.881	0.915	0.897	0.799	0.79	0.825	0.807	—
HCF+Word2Vec	**0.81**	0.81	0.821	**0.815**	**0.915**	0.90	0.937	**0.918**	**0.825**	0.801	0.872	**0.835**	—
HTF+Syntactic	0.787	0.768	0.834	0.800	0.871	0.846	0.912	0.878	0.798	0.777	0.847	0.810	+7.2%
HCF+Syntactic	0.814	0.794	0.858	**0.825**	**0.922**	0.91	0.939	**0.925**	**0.850**	0.827	0.891	**0.858**	+2.3%
Word2Vec+Syntactic	0.797	0.798	0.807	0.803	0.902	0.886	0.926	0.906	0.806	0.801	0.822	0.811	+ 0.4%
HCF+Word2Vec+Syntactic	**0.817**	0.813	0.832	0.823	0.915	0.90	0.936	0.918	0.837	0.822	0.866	0.844	+0.6%

Table 1: Humor recognition performance of four baselines and their combinations with syntactic structure features. HTF: Humor Theory driven Features, HCF: HTF plus KNN features.

- Word2Vec. The method makes use of the semantic representation of the sentences by using the averaged word embeddings as features. Similar to KNN features, it is also sensitive to content. We used the pre-trained word embeddings that were learned using the Word2Vec toolkit (Mikolov et al., 2013) on Google News dataset.[2]

- HCF+Word2Vec. The method combines HCF and Word2Vec. Since the combination of two strong methods, it achieved best evaluation scores in (Yang et al., 2015). It is more dependent on content, since both KNN features and Word2Vec features are used.

4.3 Results

The results are shown in Table 1 with accuracy ($Acc.$), precision (P), recall (R) and F_1 score. To verify the effectiveness of proposed features, we add syntactic structure features (Syntactic) to the four baselines respectively. The performance can be improved to varying degrees.

We can see that after adding syntactic features, HTF achieves significant improvements on all datasets, with an average improvement of 7.9% in accuracy and 7.2% in F_1 score. This indicates that syntactic structure features can complement humor theory driven features and benefit humor recognition. Syntactic structure features don't depend on specific content of texts and their performance is consistent on different datasets, whose sentences have different distribution of the number of tokens. It means that syntactic features can capture some key properties of humor.

Other baselines all consider content features and their performance is greatly superior to HTF. However, such great improvements may come from the artifacts in the datasets rather than capture the nature of humor. The non-humorous samples in our experiments contain news titles, which may involve different vocabularies compared with humorous samples. It is highly probable that these models match specific content and topics better. Even so, after adding syntactic features, all these baselines still achieve improvements, although the margins become small.

Feature	RPBN				CoNLL				Mixed			
	Acc.	P	R	F_1	Acc.	P	R	F_1	Acc.	P	R	F_1
HTF	0.709	0.702	0.749	0.725	0.801	0.771	0.862	0.814	0.707	0.683	0.793	0.734
HTF+DR	0.774	0.756	0.823	0.788	0.868	0.837	0.917	0.875	0.783	0.755	0.848	0.799
HTF+PR	0.783	0.775	0.81	0.793	**0.877**	0.856	0.91	**0.882**	0.797	0.784	0.83	0.806
HTF+SE	0.772	0.75	0.832	0.789	0.869	0.850	0.901	0.875	0.79	0.761	0.855	0.805
HTF+Syntactic	**0.787**	0.768	0.834	**0.800**	0.871	0.846	0.912	0.878	**0.798**	0.777	0.847	**0.810**

Table 2: Contributions of individual syntactic structure feature types. HTF: Humor theory driven features, DR:dependency relations, PR: production rules , SE: statistics of syntactic elements, Syntactic: DR+PR+SE.

In addition to reflecting the effects of the syntactic structure features as a whole, Table 2 shows the results of adding individual syntactic structure features on the basis of HTF. We can see that all three

[2] `https://code.google.com/archive/p/word2vec/`

Production Rules	Pearson	One-Liner	RPBN
S→NP VP	0.236	51%	22%
ROOT→S	-0.229	75%	74%
NP→PRP	0.209	59%	29%
ROOT→NP	-0.188	10%	18%
SBAR→IN S	0.173	20%	6%
WHADVP→WRB	0.150	13%	3%
SBAR→S	0.138	16%	6%
ROOT→SBARQ	0.136	7%	0.8%
NP→NP SBAR	0.120	17%	7%
SBAR→WHADVP S	0.112	10%	3%

(a) Discriminative production rules in One-Liner and RPBN non-humorous dataset

Production Rules	Pearson	One-Liner	CoNLL
NP→PRP	0.352	59%	25%
S→NP VP	0.207	51%	37%
WHADVP→WRB	0.193	13%	3%
NP→NNP NNP	-0.179	4%	8%
ROOT→SBARQ	0.166	7%	0.4%
NP→NP SBAR	0.163	17%	8%
SBAR→WHADVP S	0.150	10%	3%
WHNP→WP	0.148	11%	3%
PP→IN NP	-0.136	55%	81%
SBAR→IN S	0.097	20%	15%

(b) Discriminative production rules in One-Liner and CoNLL non-humorous dataset

Table 3: The top discriminative production rules and their distributions in corresponding humorous and non-humorous datasets, sorted by Pearson correlation coefficient. Rules in bold have the same correlation trend in Table 3(a) and Table 3(b).

types of syntactic structure features improve the performance on three datasets. Generally, their contributions are close. Production rule features perform slightly better. This may mean that different kinds of syntactic representations have similar effect. Next, we analyze the three types of syntactic structure features respectively.

Statistics of Structural Elements. The performance of features related to the statistics of structural elements is fairly good on three datasets. We calculate the Pearson correlation coefficient between the specific features and humor on RPBN and CoNLL datasets. We found that not all structural elements have the same correlation on two datasets but some are consistently discriminative. The features that have best positive correlations are the phrase ratio of verb phrases (PR_VP) and the number of subordinating conjunctions (SBAR_Count). In contrast, the features that have consistent negative correlations are the length ratio (LR_NP) and the average length of noun phrases (APL_2_NP).

Production Rules. As shown in Table 2, production rule features can lead to great improvements. We also compute the Pearson correlation coefficient between production rules and humor and the ratio of each production rule in humorous instances or in non-humorous instances of two datasets. Table 3 shows the results. The production rules are ranked according to the correlation coefficient. We can see that some rules are discriminative on both datasets such as *S→NP VP*, *SBAR→IN S* and *WHADVP → WRB*. Their distributions in humorous and non-humorous texts are quite different. This analysis demonstrates that syntactic structures can provide useful information for distinguishing humorous texts from non-humorous texts.

Dependency Relations. Table 4 shows the discriminative power of dependency relations with Pearson correlation coefficient. We can see that the most discriminative features are related to the dependency relation *nsubj*, *compound*, *case*, *aux*. The relation *case* and *compound* are less in humorous instances. The relation *nsubj* together with pronouns (*PRP*) occurs more in humorous instances, while the relation *nsubj* together with noun phrases are less in humorous instances.

Generally, three types of features describe syntactic features from different angles so that their contributions have overlap but also supplement each other in certain degree. One interesting question is that do these features indicate some linguistic interpretation? We attempt to discuss it in next part.

4.4 Linguistic Interpretation

The experimental results show that our proposed syntactic structures do help to identify humor. In this section, we explain how these syntactic structures relate to linguistic phenomena. We found that humorous texts have the following characteristics that can be explained by the syntactic features.

1) Humorous texts use simpler words but more complex syntactic structures.

We can see that compound phrases appear much less in humorous texts. This can be revealed by the average length of noun phrases and the dependency relation *compound*, which have negative correlations

Dependency Relations	Pearson	One-Liner	RPBN
NNP○ compound○ NNP	-0.167	8%	17%
VBP○ nsubj○ PRP	0.163	18%	5%
VBZ○ nsubj○ NNP	-0.138	2%	7%
NNS○ compound○ NNP	-0.127	2%	6%
VB○ nsubj○ PRP	0.121	16%	6%
VB○ aux○ VBP	0.114	7%	2%
VBZ○ dobj○ NN	-0.112	5%	8%
VBP○ mark○ IN	0.105	6%	2%
NNP○ case○ IN	-0.104	5%	9%
VBP○ advmod○ WRB	-0.104	4%	0.5%

(a) Discriminative dependency relations in One-Liner and RPBN

Dependency Relations	Pearson	One-Liner	CoNLL
VBP○ nsubj○ PRP	0.229	18%	3%
VB○ nsubj○ PRP	0.220	16%	3%
NNP○ case○ IN	-0.203	5%	25%
NNP○ compound○ NNP	-0.190	8%	34%
VB○ neg○ RB	0.174	12%	4%
NNS○ amod○ JJ	-0.165	10%	32%
CD○ compound○ CD	-0.161	0.1%	7%
VB○ aux○ VBP	0.160	7%	0.8%
NNS ○ case○ IN	-0.151	15%	36%
NN○ nmod:poss○ PRP$	0.142	15%	8%

(b) Discriminative dependency relations in One-Liner and CoNLL

Table 4: The top discriminative dependency relations and their distributions in corresponding humorous and non-humorous datasets, sort by Pearson correlation coefficient. Dependency relations in bold have the same correlation trend in both datasets.

with humor. The reason may be that many jokes are about events in life so that common words are more often used, while complex and professional words are less used. In contrast, humorous texts often have subordinate conjunctions, which means that subordinate clauses are often used. This can be seen in features involving $SBAR$, such as *NP→NP SBAR*, *SBAR→WHADVP S* and *SBAR→IN S*, all have positive correlations with humorous instances. Considering the following sentence,

I've learned that we are responsible for what we do, $unless$ we are celebrities.

Here, subordinated conjunction $unless$ is used to bring an $unexpected$ feeling to the readers, which results in comedy effect. To break the expectation, complex syntactic structures are often utilized.

2) Humorous texts are more vivid and specific.

We can see that aux relation appears much more in humorous instances as shown in Table 4. This is because humorous texts usually describe some details to let reader imagine the situation, so that auxiliary words are used more to enhance such descriptions.

In addition, aux often related to negation. For example, the feature $VB \circ aux \circ VBP$ mostly describes "do/don't + verb" pattern. The use of negation is usually related to an attitude change or contrast effect, which may lead to humor.

We also see that *WHADVP→WRB* and *SBAR→WHADVP S* have a positive correlation. This is because interrogative adverb why are often used to produce a satirical effect, such as

If ignorance is bliss, why aren't more people happy?

Such rhetoric questions are useful to enhance the expressiveness of a sentence.

3) Humorous texts are more like conversations.

We can see that personal pronouns often appear in humorous texts, including first-person, second-person and third-person pronouns. The corresponding features include the production rule *NP→PRP* and the dependency relation $VBP \circ nsubj \circ PRP$. These words appear more often in conversations. Besides, the rhetoric questions as we discussed earlier also build an effect like conversation.

This phenomenon can be explained by superiority theory (Keith-Spiegel, 1972; Gruner, 1997) that humor is the result of comparing with others. As a result, it is unavoidable to mention people with pronouns. Also, many jokes involve dialogues between people so that they are naturally a kind of conversation.

5 Related Work

Much existing work models humor according to psychological theories. For example, in early studies, Taylor and Mazlack (2004) proposed a computational approach to identify humor based on the humor theory of Raskin (1984). This method took into account the constraint set of all possible jokes, which took wordplay as a component. The algorithm used in their method learned the statistical patterns of text in N-grams, which provided a heuristic focus for the location of wordplay. Purandare and Litman (2006) identified humor by modeling the distinction of prosodic characteristics between humorous and non-humorous speech by utilizing acoustic-prosodic and linguistic features. Mihalcea and Pulman (2009) presented and analyzed the most frequent features including human-centeredness and negative polarity, which solved two questions related to humor recognition: One is whether humorous and non-humorous texts are separable. Another is whether the characteristics of humor are special.

Recently, Zhang and Liu (2014) designed features according to influential humor theories, linguistic rules and affective dimensions. These features can be applied to distinguish humorous tweets from non-humorous tweets. Motivated by humor theories, Yang et al. (2015) modeled humor by using the semantic structures including incongruity structure, ambiguity, interpersonal effect and phonetic styles. We used their method as one of our baselines. A recent work (Liu et al., 2018) studied sentiment association in discourse for humor recognition.

Some work considers content features. Mihalcea and Strapparava (2005) reported that using content-based features could achieve big improvements. However, content features may overfit the dataset, because in most existing work, non-humorous instances are sampled according to some assumptions which may bring in artifacts. Our work describes humorous expressions by exploring syntactic structures and treats humor as a kind of style. The syntactic structure features don't depend on specific content or topics.

6 Conclusion

In this paper, we have presented a method for humor recognition by exploiting syntactic structure features. We derive features from parsing trees and demonstrate that such features can improve the performance of humor recognition. Moreover, they are not dependent on specific content, which reduces the risk of satisfying artifacts during dataset construction and helps capture the essence of humor.

We analyzed the linguistic insights behind the features and found that humorous texts tend to 1) use simple words but with complex sentence structures; 2) be more vivid with auxiliary adverbs, negations and rhetoric questions; 3) be like conversations, involving more personal pronouns and questions. These observations indicate stylistic characteristics of humor and provide an opportunity to humor computation in a new perspective.

Acknowledgements

The research work is funded by the National Natural Science Foundation of China (No.61402304), Beijing Municipal Education Commission (KM201610028015, Connotation Development) and Beijing Advanced Innovation Center for Imaging Technology.

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.

C Fellbaum and G Miller. 1998. *WordNet:An Electronic Lexical Database*. MIT Press.

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.

Eduard Hovy, Mitchell Marcus, Martha Palmer, Lance Ramshaw, and Ralph Weischedel. 2006. Ontonotes: the 90% solution. In *Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, pages 57–60.

Patricia Keith-Spiegel. 1972. Early conceptions of humor: Varieties and Issues. *Psychology of Humor*, pages 3–39.

Dan Klein and Christopher D Manning. 2003. Accurate unlexicalized parsing. In *Meeting of the Association for Computational Linguistics*, pages 423–430.

Herbert M. Lefcourt and Rod A. Martin. 1986. *Humor and Life Stress*. Springer New York.

Lizhen Liu, Donghai Zhang, and Wei Song. 2018. Modeling sentiment association in discourse for humor recognition. *Association of Computational Linguistics*.

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

Rada Mihalcea and Stephen Pulman. 2009. Characterizing humour: An exploration of features in humorous texts. In *International Conference on Computational Linguistics and Intelligent Text Processing*, pages 337–347.

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.

Ani Nenkova, Jieun Chae, Annie Louis, and Emily Pitler. 2010. Structural features for predicting the linguistic quality of text - applications to machine translation, automatic summarization and human-authored text. In *Empirical methods in natural language generation*, pages 222–241.

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.

Sameer Pradhan, Alessandro Moschitti, Nianwen Xue, Olga Uryupina, and Yuchen Zhang. 2012. Conll-2012 shared task: Modeling multilingual unrestricted coreference in ontonotes. In *Joint Conference on EMNLP and CoNLL - Shared Task*, pages 1–40.

Amruta Purandare and Diane Litman. 2006. Humor: prosody analysis and automatic recognition for f*r*i*e*n*d*s*. In *Conference on Empirical Methods in Natural Language Processing*, pages 208–215.

Victor Raskin. 1984. *Semantic Mechanisms of Humor*. Springer Science & Business Media.

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

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

Julia M. Taylor and Lawrence J. Mazlack. 2004. Computationally recognizing wordplay in jokes. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 26.

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*, pages 347–354. Association for Computational Linguistics.

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.