Humor Identification in Microblog

Yong Ren*, Nobuhiro Kaji*, Naoki Yoshinaga*, Masaru Kitsuregawa*†
*Institute of Industrial Science, University of Tokyo
Email: {renyong, kaji, ynaga }@tkl.iis.u-tokyo.ac.jp
† National Institute of Informatics
Email: kitsure@tkl.iis.u-tokyo.ac.jp

***Abstract*—Humor is one of the most wonderful phenomenons in our life. Humor research has attracted extensive research in literature, linguistics, psychology and cognition, various theories have been proposed chronically. However, humor research has not cause much attention in NLP or text mining field. In this work, we focus on humor vs. non-humor categorization in Sina Weibo, and we explore the usage of well-known supervised and semi-supervised learning approaches in humor identification, we also exploit several conventional paradigms including hyper-parameters tuning, feature selection, machine translation, and synonym expanding in this work. We argue that empirically distinguishing non-humorous content from humorous text in Sina Weibo is still a challenging task.**

I. Introduction

Computational humor[1] appears on the scene in 90s last century. Many works [2], [3], [4], [5] hope to generate humor with the power of computer. Briefly speaking, those works focus on word level amusing such as puns or try to make humor text according to some fixed patterns. Several researches [6], [7], [8] target at discriminating humorous text from the other counterparts appear in recent years. But generally, they restrict the humor text to one liner or two-liners, and the genre or content in non-humorous text also have constrains. The potential applications of humor identification can locate in content recommendation. And we also believe that humor categorization or identification can also play positive role in humorous text generation, which is to be believed has positive role in computer human interface (CHI) venue.

Sina Wibo is the most popular micro-blog service in China, the number of registered users has surpassed 503 million by December 2012[1]. Sina Weibo looks more like the combination of Twitter and Facebook, users can post comments on others' original messages (we will use the words "message" and "text" interchangeably), meanwhile, they can also choose to transpose their comment with those original text. The authors can also reply to others and transfer their replies.

Humor expression is common in Sina Weibo, and there are various purposes: some people make use of it to show their wit; while others may choose it as a way of implicitly criticizing some people or events. The length limit in Weibo message (at most 140 Chinese characters) make the humor expression more creative and more difficult to recognize. In the following, we will show an instance and analyze the complexity of this task.

[1]http://thenextweb.com/asia/2013/02/21/chinas-sina-weibo-grew-73-in-2012-passing-500-million-registered-accounts/

"One cobras with high myopia dates with one elephant, after a brief greeting, the cobras says to the nose of that elephant 'you are so polite to bring me such a big pig'."

The message is a very representative humorous text in the following points: Firstly, it requires the readers own a certain inference capability, the shapes of the cobras and the nose of elephant are both slender, so the nearsighted cobras mistakes the nose of the elephant as his counterpart. Secondly, it involves culture background, "cobras" written in Chinese characters is "眼镜蛇", while "glasses" written in Chinese characters is "眼镜". Moreover, "glasses" has highly relatedness with "high myopia". So the appearance of composition "cobras with high myopia" brings amusing feeling. Finally, this humorous text also has connections with the readers' experience. A few Chinese children once read one cartoon tells the story that one elephant lost his nose and be mistaken as a pig.

Through the explanation above, compared with other tasks such as topic categorization and sentiment classification, we can find that humor identification is a hard task which involves inference and experience, and it is also culture dependant. There are a lot of implicit factors, unfortunately, we could not get enough context from the briefly text in Sina Weibo. In this work, we focus on conquering the problem of lacking context with the help from synonym dictionary.

The contribution of our work is that we exploit representative supervised learning algorithms (SVM) and semi-supervised learning approach (Label Propagation) in distinguishing humorous short text from non-humorous message. In order to improve the performance, we adopt several common methods including feature selection, machine translation, and synonym expanding. We conclude that humor recognition is a difficult problem.

This paper is outlined as follows: we will introduce the related work in section II, then we will explain methods we explored in this task in section III and show the experiment and results in section V. Next, we discuss the challenges based on instances and propose the potential way to resolve humor identification in VI. Finally, we conclude this work and note the future work in section VIII.

II. Related Work

During the passed years, there are a lot of research focus on generating humor automatically. Most of them [2], [3], [4], [5] only attempt to produce word level amusing expression. The representative system called JAPE [3] equips with the ability to make punning riddles. Similarity in phonology is the basis of suck kinds of research.

[8] is sort of unconventional in those humor generation studies. The goal of their work is synthesizing complete humorous short text with much flexibility. Their progress in based on Semantic-Script Theory of Humor (SSTH) which we has introduced in the discussion part VI. This work also relies on knowledge bases such as ConceptNet [9] and General Transition Network [10].

[6] make use of Naive Bayes and SVM to classify humorous text and non-humorous text. They only take one sentence joke ("one-liner") into account. The excellent result are from Reuters titles vs. one-liners and proverbs vs. one-liners. When the negative instances are British National Corpus (BNC) which are text in mixed form, the performance highly degenerates. The author also tried to detect incongruity existing in text in their following work [7].

[11] explore semi-supervised learning (SSL) approach to detect sarcastic narrative in Twitter and Amazon reviews, and they pay much attention on the feature selection. Thought they do not mention it clearly, the SSL strategy they use is quite similar to label propagation which is also exploited in our task.

III. Exploited Algorithms

A. Supervised Learning

When people formulate their task as classification, usually, SVM [12] is well accepted choice to resolve the problem.

Theoretically, SVM aims at minimizing the following objective function in Equation 1. Intuitively, it tries to classify the training instances and, at the same time, maximizing the margin (i.e., minimizing w). It is well known that the performance of SVM is dependent on the choices of kernels and corresponding hyper-parameters(i.g., C). In our work, we tune those hyper-parameters on development dataset, the process will be described in detail in the section V-B.

$$\sum_{i=1}^{l} \max(1 - y_i(w^T x_i + b), 0) + \lambda_1 \left\| w \right\|^2 \tag{1}$$

B. Semi-supervised Learning

The main disadvantage of those supervised approaches is that they demand a large amount of training data to achieve high accuracy. When we encounter the circumstance where we do not own sufficient training data due to labeling cost or difficulty in collecting samples with high quality like our task. Semi-supervised learning (SSL) [13] will rank top on our option list.

The SSL algorithms we exploit in humor identification is Label propagation (LP) [14]. Essentially, it is a graph-based SSL algorithm and owns a lot of advantages including a well-defined objective function and convergence property.

Mathematically speaking, LP aims at minimizing the following objective function 2.

Algorithm 1: Label Propagation

Input : $G = \{V, E, W\}$ is the similarity graph; w_{ij} is the elements of W

1 Initialize label matrix Y by using seed examples

2 $T = D^{-1}W$

3 **while** *Y is not convergent* **do**

4 $\quad Y = TY$

5 $\quad$ Clamp the seed examples in Y to their original values

6 **end while**

Output: Y

$$\frac{1}{2}\sum_{i,j=1}^{l+u} w_{ij}(f(x_i) - f(x_j))^2$$
$$\textit{subject to } f(x_i) = y_i, \textit{ for } i = 1, 2, \ldots, l \tag{2}$$

Equation 2, which is sometimes referred as *energy* or *smoothness*, is the common objective function in graph-based SSL method. Intuitively, this optimization can be interpreted as trying to assign the same labels to vertices that are connected by highly weighted edges, while fixing the labels of the vertices corresponding to labeled data.

It is not difficult to verify the solution to the problem in Equation 2 satisfies the following stationary conditions.

$$\begin{aligned} f(x_i) &= y_i, \textit{ for } i = 1...l \\ f(x_j) &= \frac{1}{d_j}\sum w_{ij} f(x_i), \textit{ for} j = l+1, \ldots, l+u \\ \textit{where } d_j &= \sum_j w_{ij} \end{aligned} \tag{3}$$

One step further, the Equation 3 can be transformed into matrix form $f = Pf$, where $P = D^{-1}W$ and D = diag(d_i). Then we could seek $f(x_i)$ $(i = l+1, \ldots, l+u)$ that satisfies in Equation 3 in an iterative manner.

Algorithm 1 depicts LP in detail. In the initiation section (line 1 and line 2 in Algorithm 1), it firstly initializes the label matrix Y, which has n rows and m columns, where n is the number of text including both labeled and unlabeled ones (line 1), and m is the number of labels, in our case $m = 2$. Here we define column 0 and column 1 are used to store the probability with which each text is labeled as positive and negative, respectively. The rows in Y corresponding to positive- and negative-labeled data are initialized as (1, 0) and (0, 1), respectively. After initialization, a new matrix T is built through transforming the original similarity matrix W (line 2). Then, LP enters the learning phase (from line 3 to line 6 in Algorithm 1), it propagates labels through the graph (line 4). In essence, it is an iterative matrix computation. At the end of each iteration, the seeds are re-adjusted to original value (line 5).

When label matrix Y converges, the propagation terminates. The unlabeled data are classified according to the the matrix Y: for each review, if the value of column 0 (humor value) is larger than column 1 (non-humor value), the message is classified as humorous text and vice versa.

IV. Explored Methods

We will introduce the paradigms adopted in our work for improving the performance in this section.

A. Machine Translation

The idea that categorize text with the help of Machine Translation (MT) service has been employed in work [15]. The main benefit is that we can make use of linguistic resource of target language to conquer the problem in source language. This paradigm is particularly useful when our source language is under-resourced.

In this task, we translate the Chinese text into English counterpart through Google translation service. Then we can make use of WordNet [16]. The other purpose is investigating the influence of culture background in humor understanding.

B. Synonyms Expanding

As we has noted in section IV-A, the main advantage of take the measure of MT in classification is to benefit from the available resources in the target language. In WordNet, is is very convenient for us to find the synsets (consist of words with similar meaning) for the given words, which means we can expand the words in text easily. We should note that there is the chance that one word has several different synsets, now we only consider the first synset(the primary synset).

At present, we restrict the number of synonyms of each word in our dataset no more than five. The rational for this constrain is, firstly a lot of words do not own synonyms more than five, more important, we argue that the words expanding process in our mind is not un-limited imagination, we do not make our classifier go wild. Consider the existence of polysemous words, randomly expanding may be a better way, which we will adopt in the future work.

Just we have pointed out in section I, one challenge we are facing is the limited words in Weibo messages. So enlarging the lexicon is a reasonable and straightforward choice. As a matter of fact, when we read jokes, we could realize the process of associating the words we has read with the concept in our mind. We hope we could capture such kind of association by including the synonyms of the words in the text.

Meanwhile, we also try to conduct this synonyms expanding in the original Chinese text. The synonyms thesaurus we use is called “Synonyms Cilin (HIT extended version)” [2], which totally contains 77,458 words. This synonyms thesaurus compose of lines, and each line corresponds to the synset begins with group ID, followed with the words belong to this synset.

C. Feature Selection

The last trying is feature selection. It is known that the discriminating function of features in text classification are not equal, it is a necessary procedure before the classifier are put into real usage. According to the theorems involved, there are various feature selecting approaches [17]. In this work, we employ a conventional and straightforward one: F-score. The comparison of distinguishing power among those features in our case can be seen in Figure 1, in which the horizonal axis is the index number of features and vertical axis is the negative ratio between p-value of each feature and the highest value.

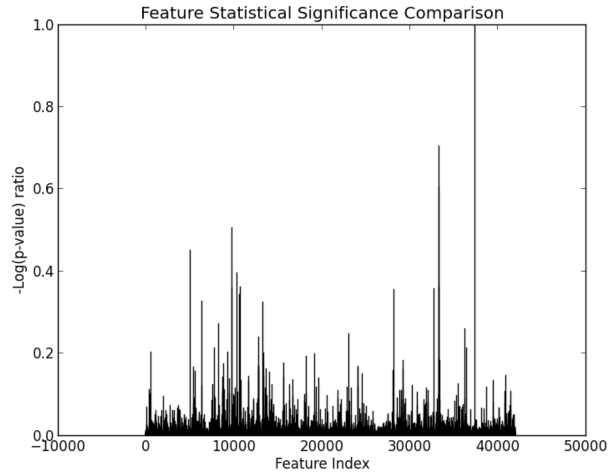

Fig. 1. Significance among features

V. Empirical Evaluation

In this section, we will present experiments we have conducted and show the finding. The purposes of our experiments are, firstly, investigating whether there is manual agreement on humor identification, then investigating use of conventional classifying methods and related paradigms adopted in text categorization field on humorous text identification. We begin with the description of the data collecting procedure.

A. Data Collection and Pre-processing

Kinds of unexpectedness, collecting sufficient data for training and test is not trivial, even humor expression is very common in Sina Weibo. The main cause is that humorous text does not have evident and consistent language cues that we can rely on to filter and accumulate a large amount of data. The strategy we adopt is depicted in the following:

For the training data, we focus on the number. The first kind of source for humorous text is some special Sina Weibo accounts, and we select nine famous accounts in Sina Weibo that own tens of thousands to millions of followers. We should note that the text from those accounts are noisy, besides humorous text, other types are advertisement (some of them are very difficult to recognize), opinions or comments on some hot social topics, interaction with others users and so on. Even in the humorous text, there are many messages merely talking about the interesting pictures accompanying with the original text, more worse, humorous text with pictures are also common. In other words, we do not have good way to differ humorous and non-humorous text posted by those accounts, that is the reason we use these data only for training. The

[2] http://www.datatang.com/data/42306/

TABLE I. STATISTICS ON DATASET

Class	Number of training	Number of test	Number of development
Humor	40,355	568	568
Non-humor	34,268	626	625

TABLE II. REVIEWS AND THEIR SENTIMENT PHRASES

Text	Features extracted
眼镜蛇高度近视，和大象初次约会， 客套一番后,眼镜蛇对着大象的鼻子说 哎，来就来吧，还牵着这么大 一头猪来，你真实太客气了！ (One cobras with high myopia dates with one elephant after a short exchange of pleasance, the cobras says to the nose of that elephant ’you are so polite to bring me such a big pig’)	眼镜蛇 高度近视 大象 初次 客套 鼻子 还牵 太客气 大 猪 (cobras, high myopia, elephant nose,bring, so polite, big, pig)
程序员是可以做一辈子的职业， 即使将来不做程序员， 第一份工作从程序员做起， 也是不错的选择：写代码可以锻炼一个人的 逻辑思维能力，分析能力， 专注度，和一丝不苟的态度。 这些能力对将来 从事任何一个职业都是必须的 (Programmer can be a lifetime career, even you will be not a programmer in the future, it is good to choose programmer as your first job: Writing code can exercise a person’s logical thinking ability analytical ability, make you become focused, and have meticulous attitude. These capabilities are necessary when you engage in any one occupation in the future.)	程序员 职业 不错 代码 锻炼 逻辑 思维 能力 分析 一丝不苟 态度 必须 将来 (programmer, caree, good, code exercise, thinking, ability analytical, focused attitude, necessary, future)

other source for humorous text is several websites that list humorous stories, there are similar noise phenomenon. For these stories, we only reserve the text with the length less then 140 Chinese characters. For non-humorous training data, we randomly selected messages provided in Weibo Pameng [3] which is a crowdsourcing crawling project in China.

For the test data, we emphasize the cleanses and consistencies, meanwhile, we do not want to put constrains on the content or text format. Firstly, we issue the special hashtag # 幽默微小说# (humorous hint fiction) into the Sina Weibo search interface and crawled the retried results. For the hashtag is very particular, the data we collect is clean. Then we find several websites that list hint fictions, we crawled all the text that are listed in the humorous type. Just we described in section I, there are different types of humor in Sina Weibo, in order to get more types of humorous text, we also collected all the humorous Weibo text listed by the websit called “micro quotation”[4] which collects popular Weibo messages of different types. At the same time, we also collect hint fictions of other types (terrifying and moving) and popular messages of other classes in that “micro quotation” website, and take them as non-humor type. For these test data we further divide them into test data and development data for tuning related hyper-parameters.

The size of training, test data and development data are listed in the Table I. We have to confess there are two uncertainties that could have side influence on our task. One is the training data is noisy, there are false training samples; the other one is the non-humor training data is not consistent with the non-humor test data in genre. We will try to adjust them in the future.

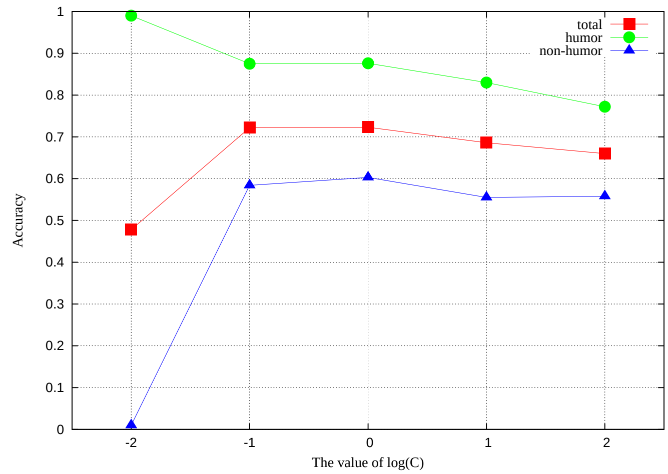

Fig. 2. The impact of hyper-parameter C

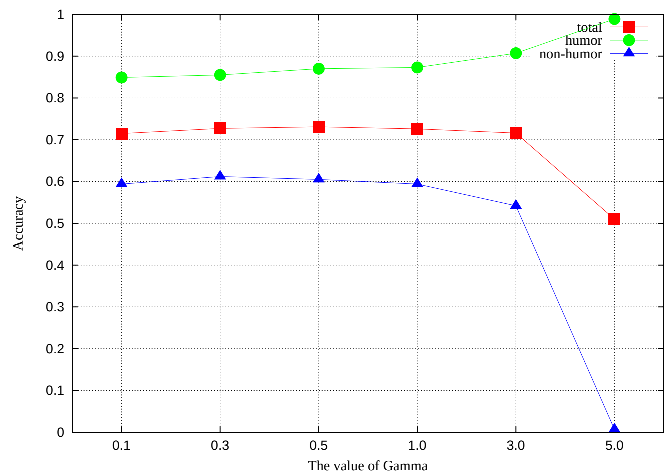

Fig. 3. The impact of hyper-parameter gamma

In our work, besides nouns, verbs, adjectives, we also add phrases with specified POS patterns as classifying features (prior to words and phrases extraction, Chinese text processing tool ICTCLAS2013 [5] are used to pre-process each text). The Table II lists two examples of features extracted from messages in the dataset we used in our experiment. The humorous text is the instance appearing in the section I.

We used SVM in scikit-learn [6] and Label Propagation in GraphChi [7] as the implementations of SVM and LP respectively in our experiments.

B. Classifying with Supervised Algorithm

One of the most famous hyper-parameters in SVM is C which controls the trade-off between training error and margin. The performance of humor recognition is highly sensitive to the choice of hyper-parameter C, which can be evidently seen in Figure 2. One advantage of SVM is the kernel trick [18]

[3] http://cnpameng.com/
[4] http://www.izz.cc/
[5] http://ictclas.nlpir.org/
[6] http://scikit-learn.org/stable/
[7] https://code.google.com/p/graphchi/

TABLE III. COMPARISON ON SEMANTIC EXPANDING

Method	Accuracy	Humor precision	Non-humor precision
Original text	0.727	0.873	0.594
MT	0.679	0.652	0.704
MT + synonyms(English)	**0.733**	0.877	0.622
Original text + synonyms	0.703	0.859	0.560

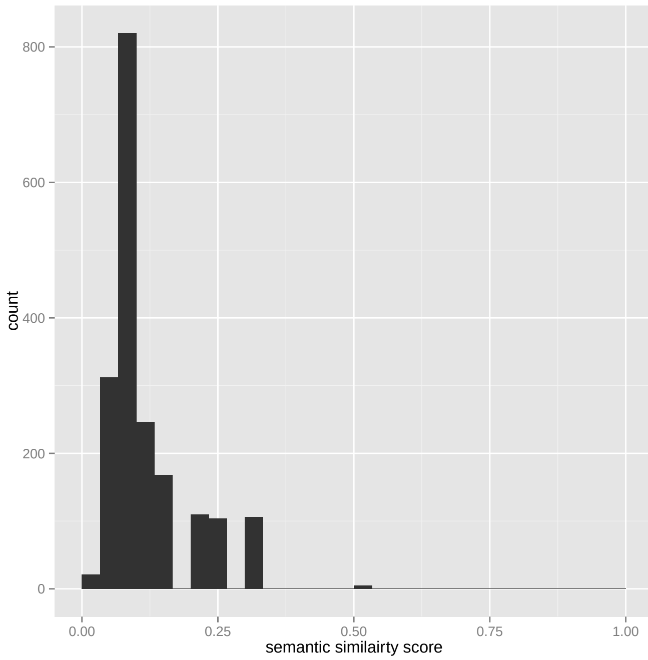

Fig. 4. Semantic similarity score among humorous Text

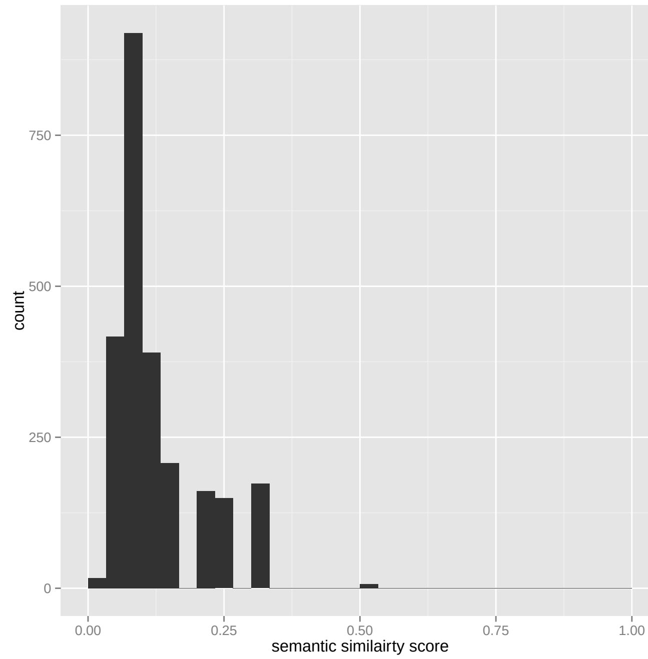

Fig. 5. Semantic similarity score among non-humorous text

TABLE IV. FEATURE SELECTION EFFECT

% features saved	Accuracy	Humor precision	Non-humor precision
100	0.727	0.873	0.594
90	0.734	0.868	0.613
70	0.749	0.877	0.634
50	**0.752**	0.896	0.624
30	0.737	0.866	0.621

which is believed to have positive role in text classification, we investigate RBF kernel in this task. RBF kernel brings in a new hyper-parameter called gamma. The impact of hyper-parameter gamma can be seen in Figure 3.

After exploring conventional semi-supervised and supervised learning methods, we can find the result is not very promising. In the following, we focus on improving the performance. All the performance enhancing trying are based on the development data, and all evaluations are devised using SVM with optimal hyper-parameters setting (RBF kernel, C is set to 1.0 and gamma is set to 0.5).

C. Classifying using Machine Translation

The result of using MT is shown in Table III. Conforms to our intuition, the performance become worse for the humor is related to language or culture. In reality, it is not rare to find that audience are aloof and indifferent to the amusement come from different nations.

D. Classifying using Synonyms Expanding

We can find in Table III the positive role of this native synonyms expanding strategy is not evident. Possible explanation locates in the semantic similarity (computed using the path similarity in WordNet, only for English translation) among those text including the synonyms words expanded. We can see the semantic score histogram clearly, in Figure 4 and Figure 5, that the distribution between humor and non-humor histogram are quite resemble, which means it could not work well in humor/non-humor distinguishing task.

E. Classifying using Feature Selection

The impact of feature selection is present in Table IV, and we can find that feature selection can play positive role if we applied it appropriately. We will explore more sophisticated feature selecting paradigms in our future work.

VI. MANUAL EVALUATION

Manual examination is important for us to find whether human being can distinguish humorous text from non-humorous counterparts. We give texts in the format of two-lines (249 for each class) from training data to two Chinese graduate students and ask them to judge wether they are humorous or not. This work also helpful for us to learn the quality of training data. The Cohen's kappa score is 0.5, which shows moderate agreement.

In order to find the confounding factors and the potential way in identifying humorous text, we qualitatively analyzed the text whose labels are not agreed between those two raters. The controversial humorous and non-humorous text are listed in following Table V and Table VI respectively (for better understanding, only English translation are shown).

TABLE V. CONTROVERSIAL HUMOROUS INSTANCES

ID	Text
1	Zhan Zhao studies painting; a group of black appears in the paper. Gong Suce says Zhan is wasting paper. Zhan replies coldly "I am paining portrait for Chief Bao"
2	One mosquito comes into the city, it is very hungry. The mosquito finds a lady with tall breasts, and then it plunges into the breasts and fiercely bites, finally its mouth is full of silicone. It sighs: "Oh, the food safety becomes a serious problem! Where is the safe milk ah?"
3	Doctor: Your X-rays shows your rib fractures. Patient: then how to do? Doctor: It does not matter; I have helped you with the PS repairing...
4	Some people say that science girls look like gentlewoman, but when they start to resolve the questions their real identity will be exposed, they have the habit of stroking the hair up and showing the big forehead. The reason is CPU in high speed operation need good heat dissipation.
5	A young man hastily runs to the sixth floor, met a middle-aged man, and says: "Your daughter was hit by a car!" The middle-aged man suddenly dizzies: "ah? My God! How ah? "Then he rushes downstairs. On the fourth floor, he realizes something

TABLE VI. CONTROVERSIAL NON-HUMOROUS INSTANCES

ID	Text
1	According to some experts, Shanghai dialect seriously hinders Shanghai becoming an international metropolis. Whether you agree to prohibit Shanghai dialect or not?
2	Programmer can be a lifetime career, even you will be not a programmer in the future, it is good to choose programmer as your first job: Writing code can exercise a person's logical thinking ability, analytical ability, make you become focused, and have meticulous attitude. These capabilities are necessary when you engage in any one occupation in the future.
3	For example, according to customer value to subdivide the users, the goal is clear. It is to find some rules that can be used to distinguish user groups according to different demographic information, Why not use CART for such kind of task?
4	If it is not for rent, is the training school be lent free for the owner is moved by idealistic businessman? Dose the deceptive zero-based classes become the teacher's problem? The Board of Education officials are not afraid of bother; dare to offend hooligans, which are commendable. The last resort of rogue is physical threaten, dose the countdown need?
5	After long time hesitance, I finally become the registered students in the college and the school in yesterday and today respectively. After registration this afternoon, I can claim to be one doctor candidate. The reason for registration is very vulgar ; it is helpful in enjoying the discount of bus tickets and meal card so that I can save money...

First of all, just we have emphasized in section I, humor understanding requires background knowledge, if the readers do not know Chief Bao (one history celebrity in China)is dark skinned, they cannot identify the comic spot in the first text. In addition, if the readers know the relationship among those figures appearing in the text, it would be better for them to interpret the story. Secondly, topic distortion is common in those text, for example, "mosquito" and "silicone" in the second text, "Doctor" and "PS" in the third one, and "forehead" and "CPU" in the forth story. We believe capturing the topic or concept incongruity would be a promising way. Moreover, humor identification also has demand on the inference capability, the last story is one representation, the rater who recognize it as a amusing text should realize the "middled-age" guy has a quicker way to take the elevator downstairs.

One the other side, humor is a quite subtle and subjective matter for human bing, it is the interacting process between the author and reader through the words. Some people may feel the text is joke even the author is serious, and such kind of accidents sometimes occur in our life. One possible cause is the readers do not agree or believe the opinion of authors, the beginning three text belong to this categorization. Comic effect can be brought about by the expression, for example, the forth text. Finally, sort of un-serious words such the fifth story can also make readers laugh. We believe model the unseen interaction between authors and readers would helpful for understanding the humor generation, however, this is still an under-exploring task.

The humor theory we may benefit is incongruity theory. Arthur Schopenhauer [19] believes that incongruity plays critical role in producing comic effect. This thought is further improved in Semantic Script Theory of Humor (SSTH) [20] which advocates that both semantic relatedness and incongruity are of great importance to humor generation. Simply speaking, suck kinds of theory suppose there are two or more related compositions; in the first sight, people do not recognize the implicated meaning, gradually people notice the odd shot (usually at the end of the story), then humorous feeling come into place. From the simplified statement, we can find SSTH is consistent with the topic distortion we mentioned. We will classify humorous and non-humorous text with the consideration of topic information in the future. Whereas, devising the way to encode the knowledge into our algorithms and grant inference ability to computer would be necessary and challenging road to the goal for humor cognition.

VII. CLASSIFYING USING SEMI-SUPERVISED ALGORITHMS

Just we have emphasized in section III, SSL is a good choice when we have obstacles to access excellent training data. We select the text whose belongs are agreed by two raters (147 for each class) as labeled seeds for label propagation, the other training data are treated as unlabeled data here.

Specifically speaking, the work is divided into three steps:

Step 1: We extract from each text *features*, which are words/phrases, and then represent the message with a (tf-idf weighted) vector of extracted features (*feature vectors*).

Step 2: We construct a similarity graph by regarding the text (feature vectors) as vertices. The edge (weight) between two vertices (messages) represents a degree of similarity between their labels. We measure the similarity by using cosine similarity between two vectors.

Step 3: Having a few vertices labeled as seeds, each vertex iteratively propagates its label to its neighboring vertices according to their similarity; seeds (initially labeled vertices) thereby behave like sources that push out labels to unlabeled vertices.

We have more than 70 thousands text totally, the number of resulting edges is more than 100 million. We make use of GraphChi [21] that supports large scale graph computation on one PC with the help hard disk or SSD. The final accuracy in employing LP is 0.62.

VIII. Conclusion and Future Work

In this work we formulate humor identification as a bi-classification problem, and adopt the widely-used supervised classification algorithms (SVM) and semi-supervised learning method (LP) to resolve it. We should also mention that graph-based SSL method is not the only campaign in SSL area, SVM also has a semi-supervised version called transductive SVM(TSVM)[22]. We will explore the use of TSVM in our future work.

With the hope to improve the performance we have explored various technics including hyper-parameters tuning, feature selection, classifying with the help of Machine Translation and synonyms expanding. We conclude that humor preconization in Sina Weibo is rather difficult task for it involves in inference culture background and reader experience beyond the words contained in the text.

In the future work, first of all, we will carefully examine the quality of training data and the potential impact of the quality. Then we plan to exploit advanced feature selection and semantic expanding method and examine the helpfulness of topic information in humor identification. Finally, we will also try to enlarge the dataset and conduct large scale experiment, and make use of our methodology on perception of other types such as terrifying, moving, encouraging and so forth.

References

[1] K. Binsted, A. Nijholt, O. Stock, C. Strapparava, G. Ritchie, R. Manurung, H. Pain, A. Waller, and D. O'Mara, "Computational humor," *Intelligent Systems, IEEE*, vol. 21, no. 2, pp. 59–69, 2006.

[2] G. K.Binsted, "Computational rules for generating punning riddles," in *Humor*, vol. 10, 1997.

[3] G. Ritchie, "The jape riddle generator: technical specification," *Institute for Communicating and Collaborative Systems*, 2003.

[4] O. Stock and C. Strapparava, "Getting serious about the development of computational humor," in *INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE*, vol. 18. LAWRENCE ERLBAUM ASSOCIATES LTD, 2003, pp. 59–64.

[5] J. Taylor and L. Mazlack, "Computationally recognizing wordplay in jokes," *Proceedings of CogSci 2004*, 2004.

[6] R. Mihalcea and C. Strapparava, "Making computers laugh: Investigations in automatic humor recognition," in *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2005, pp. 531–538.

[7] R. Mihalcea, C. Strapparava, and S. Pulman, "Computational models for incongruity detection in humour," in *Computational Linguistics and Intelligent Text Processing*. Springer, 2010, pp. 364–374.

[8] I. Labutov and H. Lipson, "Humor as circuits in semantic networks," in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2*. Association for Computational Linguistics, 2012, pp. 150–155.

[9] H. Liu and P. Singh, "Conceptnet-a practical commonsense reasoning tool-kit," *BT technology journal*, vol. 22, no. 4, pp. 211–226, 2004.

[10] I. Swartjes and M. Theune, "A fabula model for emergent narrative," in *Technologies for Interactive Digital Storytelling and Entertainment*. Springer, 2006, pp. 49–60.

[11] D. Davidov, O. Tsur, and A. Rappoport, "Semi-supervised recognition of sarcastic sentences in twitter and amazon," in *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*. Association for Computational Linguistics, 2010, pp. 107–116.

[12] V. N. Vapnik, *The nature of statistical learning theory*. Springer-Verlag New York, Inc., 1995.

[13] O. Chapelle, B. Schölkopf, and A. Zien, Eds., *Semi-Supervised Learning*. Cambridge, MA: MIT Press, 2006. [Online]. Available: http://www.kyb.tuebingen.mpg.de/ssl-book

[14] X. Zhu and Z. Ghahramani, "Learning from labeled and unlabeled data with label propagation," Tech. Rep., 2002.

[15] X. Wan, "Co-training for cross-lingual sentiment classification," in *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1*. Association for Computational Linguistics, 2009, pp. 235–243.

[16] G. A. Miller, "Wordnet: a lexical database for english," *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, 1995.

[17] Y. Yang and J. O. Pedersen, "A comparative study on feature selection in text categorization," in *MACHINE LEARNING-INTERNATIONAL WORKSHOP THEN CONFERENCE-*. MORGAN KAUFMANN PUBLISHERS, INC., 1997, pp. 412–420.

[18] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. Mullers, "Fisher discriminant analysis with kernels," in *Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop*. IEEE, 1999, pp. 41–48.

[19] A. Schopenhauer, *The World as Will and Representation*, 1818, vol. 1.

[20] V. Raskin, *Semantic mechanisms of humor*. Springer, 1984, vol. 24.

[21] A. Kyrola, G. Blelloch, and C. Guestrin, "Graphchi: Large-scale graph computation on just a pc," in *Proceedings of the 10th conference on Symposium on Opearting Systems Design & Implementation*, 2012.

[22] T. Joachims, "Transductive inference for text classification using support vector machines." Morgan Kaufmann, 1999, pp. 200–209.