Stance Analysis for Debates on Traditional Chinese Medicine at Tianya Forum

Can Wang(✉) and Xijin Tang

Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, People's Republic of China
wangcan@amss.ac.cn, xjtang@iss.ac.cn

Abstract. Internet and social media devices have created a new public space for debates on societal topics. This paper applies text mining methods to conduct stance analysis of on-line debates with the illustration of debates on traditional Chinese medicine (TCM) at one famous Chinese BBS Tianya Froum. After crawling and preprocessing data, logistic regression is adopted to get a domain lexicon. Words in the lexicon are taken as features to automatically distinguish stances. Furthermore a topic model latent Dirichlet allocation (LDA) is utilized to discover shared topics of different camps. Then further analysis is conducted to detect the focused technical terms of TCM and human names referred during the debates. The classification results reveal that using domain discriminating words as features of classifier outperforms taking nouns, verbs, adjectives and adverbs as features. The results of topic modeling and further analysis enable us to see how the different camps express their stances.

Keywords: Stance analysis · Opinion mining · Latent Dirichlet allocation · Traditional Chinese medicine

1 Introduction

With the development of Internet, people can easily express and exchange their opinions through on-line forums or social media. It is widely recognized that mining public opinion from on-line discussions is an important task, which is related to a wide range of applications. There exist two streams of literature in this domain. One is distinguishing subjective expressions from factual information [1, 2]. The other is detecting the text polarity, positive or negative. The bulk of such works have focused on feature selection [3–5], classifiers optimization [6], and finally improving the precision of classifiers.

Despite the fair amount of studies in the opinion mining domain, there are several limitations of the existing literature. Firstly, opinion mining and sentiment analysis are usually used as synonyms, for both fields apply data mining and natural language processing (NLP) techniques to deal with textual information [7]. However, sentiments cannot truly represent stances [8]. Secondly, corpora are important for opinion mining. Many of previous studies used users' comments[1] or news[2] as corpora. Unlike those

[1] http://www.cs.cornell.edu/people/pabo/movie-review-data/.
[2] http://mpqa.cs.pitt.edu/.

H.T. Nguyen and V. Snasel (Eds.): CSoNet 2016, LNCS 9795, pp. 321–332, 2016.
DOI: 10.1007/978-3-319-42345-6_28

corpora, the debates on societal problems on Internet are more diverse and conversational. They are highly contextualized, depending on rich background of shared knowledge and assumptions. Thirdly, previous researches on opinion mining mostly depended on existing lexicons, or generated lexicons by seed words [9]. The lexicons or the seed words came from people's experiences. While one word may have opposite meanings within different contexts. Fourthly, some previous studies focused on automatically determining the stance of a debate participant [10–13]. There are limited researches on how people express their different perspectives towards an issue.

In this paper we focus on stance analysis of debates rather than sentiment analysis. There are two camps of people by their attitudes towards traditional Chinese medicine (TCM). Some people take the "abolishing TCM" stance. In their opinion TCM should be abolished from the national health system. Other people take the "preserving TCM" stance and insist that TCM should be preserved. The debate started since the modern medicine entered into China. The discussion on TCM is always 2-paralyzation that is correlated to culture, philosophy, history and economy. Now Internet provides a public space for people to voice and exchange their opinions on societal hot spots and the livelihood issues. We select on-line discussion on TCM as our corpus since it enables us to understand different perspectives of debates on TCM directly from the public. Considering the context of the debate, we use logistic regression to generate discriminating words relevant to TCM. Latent Dirichlet allocation (LDA) is utilized to generate topics of the two camps. We try different ways to capture how people from different camps express their viewpoints.

The rest of the paper is organized as follows. Section 2 describes related work. Section 3 discusses our corpus in more details and describes the preprocessing of data. Section 4 presents our stance classification experiments, including two policies of feature words selection. Section 5 describes topics of the two different camps. Section 6 describes further analysis to detect the focused technical terms of TCM and human names referred during the debate. Conclusions are presented in Sect. 7.

2 Literature Review

To some extent, stance analysis is related to arguing or debate. Somasundaran and Wiebe [10] from University of Pittsburgh explained that "arguing is a type of linguistic subjectivity, where a person is arguing for or against something or expressing a belief about what is true, should be true or should be done in his or her view of the world". They focused on automatically determining the stances of debate participants with respect to a particular issue. In their research, they used the MPQA (Multiple-Perspective Question Answering) corpus to get arguing lexicon for debate. They combined the arguing lexicon and sentiment lexicons as opinion features to discriminate the debate stances and improved the precision of the classifier. Anand et al. [12, 13] from University of California Santa Cruz, taking debates from open debating websites "ConvinceMe.net" and "4forums.com" as corpora, tried a variety of features to get one's stance within debate, such as repeated punctuation, syntactic dependency, posts per author, words per sentence, etc. Their research illustrated that subjective expressions varied across debates.

Stance classification, by recognizing politically oriented polarity in texts, can be widely applied in political domain. Tikves et al. [14, 15] from Arizona State University research on profiling Islamic organizations' ideology and activity patterns along a hypothesized radical/counter-radical scale. They utilized ranked perspectives to map Islamic organizations in UK and Indonesia on a set of socio-cultural, political and behavioral scales based on their web corpus. Gryc and Moilanen [16] focused on modeling blogosphere sentiments centered around Barack Obama during the 2008 U.S. presidential election. Lin et al. [17] used statistical models to identify perspectives about "Palestinian" or "Israeli" at the document and sentence levels.

3 Data Collection and Preprocessing

3.1 Debate on TCM at BBS

Because of anonymity, bulletin board systems (BBS, in this paper as "forum") are good platforms for Internet users to freely express their opinions. Tianya Forum is one of the most popular Chinese BBS sites and there are many hot posts on TCM at Tianya Forum. Some of these posts are listed in Table 1 [8]. In this paper we take the hottest post "2822432" as our corpus.

Table 1. Hot posts about TCM at Tianya Forum

Post-ID	Replies	Participants	Start time	End time
2822432	117318	4890	2012-10-16	2013-11-29
2121178	36592	5522	2011-03-21	2015-01-24
2317943	33547	6067	2011-11-12	2015-01-24

3.2 Preprocessing of Data

Firstly, we label the replies by user IDs' stances. There are 4890 authors (user IDs) who participate the debate. 267 authors who have replied more than 5 times are chosen and their stances are manually labeled. There are 84 authors who hold "abolishing TCM" stance and 183 authors who hold "preserving TCM" stance.

Secondly, we preprocess the labeled replies as follows:

(1) Remove replies with no texts.
(2) Filter out urls.
(3) Segment words with the ICTCLAS tool[3], keep the user ID names and technical terms of TCM as reserved words. We use a TCM terminology dictionary from Sougou Cell dictionary[4] which contains 28428 TCM technical terms.
(4) Remove stop words (such as "oh") from the bag of words and words with only one character.

[3] http://ictclas.nlpir.org/.

[4] http://pinyin.sogou.com/dict/detail/index/20664.

4 Stance Classification

4.1 Features and the Classifier

Lin et al. [17] observed that people from different perspectives seemed to use words with different frequencies. For example, a participant who talks about "child" and "life" at an abortion debate is more likely from an against-abortion side, while someone who talks about "woman", "rape" and "choice" is more likely from a for-abortion side. To automatically distinguish the stances of the participants, either support or oppose, in this paper we use logistic regression to get the stance feature words. The process is as follows.

(1) Calculate the frequencies of the words appeared within a reply;
(2) Create a term-document matrix of frequencies. In our research terms mean words, documents mean replies;
(3) Label the replies' stances with "1" and "−1", "1" means "preserving TCM" and "−1" means "abolishing TCM";
(4) Use the MATLAB implementation of the SLEP package[5] to run the logistic regression. The vector of labeled stances and the term-document matrix are inputs, and the vector of words' coefficients is the output;
(5) Filter words with a threshold of absolute coefficient 0. Words with positive coefficients are chosen as "preserving TCM" feature words, and words with negative coefficients are taken as "abolishing TCM" feature words;
(6) Take the selected words as features, use the "e1071" package[6] in R to train a support vector machine (SVM) model to predict replies' stances.

Adjectives words were employed as features in opinion mining, as many researches on subjectivity detection revealed a high correlation between adjectives and sentences subjectivity [18]. Benamara et al. [19] demonstrated that features with both adjectives and adverbs outperformed features with only adjectives. Subrahmanian and Reforgiato [20] added verbs to feature words besides adjectives and adverbs. Turney and Littman [9] proposed a new method to get the semantic orientation of words by using adjectives, adverbs, verbs and nouns. In this paper, we select words including all the nouns, adjectives, adverbs and verbs in the corpus as a baseline.

Pang et al. [6] employed three machine learning methods to determine whether a review was positive or negative. The results showed that SVM model outperformed Naive Bayes and maximum entropy classifier. So we approach the classification work by using SVM. Figure 1 shows the experimental process of the paper.

4.2 Results and Discussions

After preprocessing, 44940 replies are labeled "preserving TCM" and 28646 replies are labeled "abolishing TCM". To avoid the imbalance problem, we randomly sample

[5] http://www.yelab.net/software/SLEP/.
[6] http://cran.r-project.org/web/packages/e1071/.

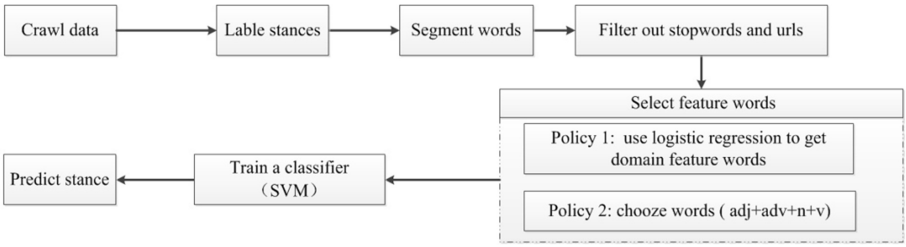

Fig. 1. The experimental process with different policies of feature selection

10000 "abolishing TCM" replies and 10000 "preserving TCM" replies. To guarantee enough text information in the replies, we select replies with more than 15 characters. We randomly split the data into training set and predicting set, each set respectively contains half of the sample data.

By logistic regression, each word has a coefficient contributing to stance towards TCM. With the threshold of absolute coefficient 0, we get 2879 discriminating words from 23441 words, including 1288 words with positive coefficients (related to "preserving TCM" stance) and 1491 words with negative coefficients (related to "abolishing TCM" stance). Table 2 lists top 15 discriminating words from both stances.

Table 2. Top 15 discriminating words in each camp

Preserving TCM			Abolishing TCM		
No.	Original Chinese words	English translation	No.	Original Chinese words	English translation
p1	反中医	Opposition to TCM	a1	郎中	Quack doctor
p2	正义	Justice	a2	中医粉	TCM fans
p3	儿童	Children	a3	博大精深	Vast
p4	废除	Abolish	a4	虫草	Cordyceps
p5	先生	Doctor	a5	巫医	Witch doctor
p6	小学	Primary school	a6	粪坑	Cesspit
p7	西医	Western medicine	a7	愚昧	Ignorance
p8	不要脸	Shameless	a8	禽流感	Avian influenza
p9	天理	Justice	a9	网友	Internet users
p10	根本	Fundamental	a10	特别	Special
p11	而已	Only/nothing more	a11	国人	Compatriots
p12	中西	Chinese and Western	a12	人中黄	Pulvis glycyrrhizae Praeparatus
p13	救死扶伤	Heal the wounded and rescue the dying	a13	质疑	Doubt
p14	一直	Always	a14	养生	Health preservation
p15	无效	Ineffectiveness	a15	郎中	Quack

Table 3 shows the experiments' results. 15669 words including adverbs, adjectives, verbs and nouns are selected. Using our domain discriminating words as features, the precision of the SVM model predicting the stances of replies is 63.13 %. Using the adverbs, adjectives, verbs and nouns as features, the precision of the SVM model is 51.18 %.

Table 3. The comparison of two feature selection policies for SVM classifier

Selection policies of feature words	Feature words	Precision
Adverbs, adjectives, verbs and nouns	15669	51.18 %
Domain discriminating words	2879	63.13 %

Shen et al. [21] attempted to identify perspectives of TCM. They collected Sina Weibo users whose tags contain their given TCM related words, crawled down these users' tweets and labeled the tweets "supporting TCM" and "opposing TCM". The differences between their research and ours are as follows. Firstly, their corpus is selected from Weibo posts and the length of the posts are limited in 140 characters. Our corpus is selected from Tianya Forum and there is no limitation of the length of the replies. So authors can fully express themselves. Secondly, their data are imbalanced, including 40,888 "supporting TCM" posts and 6,975 "opposing TCM" posts due to their biased data collection policy. We sample our data unbiased from the replies. Thirdly, there is no interaction between their subjects from Weibo while our subjects from Tianya Forum reply to the seed post or others' replies. Our corpus from Tianya Forum is more "discussion" oriented.

5 Topic Analysis Based on Camps

Latent Dirichlet allocation (LDA) is a topic model to generate topics of a group of documents based on the words of the documents [22]. We utilize LDA on the "preserving TCM" replies and the "abolishing TCM" replies to get more details to see how the opposite camps express their stances.

Table 4 shows the ten topics from the "preserving TCM" replies. We list the top 15 words of each topic. We label the topics by words distributed in the topic. There are mainly five groups of topics from the "preserving TCM" stance holders. Firstly, people in this camp doubt the motivation of the "abolishing TCM" stance holders. In their standpoint, the "abolishing TCM" stance holders are traitors of the traditional culture (e.g., Topics p1 & p3). Secondly, they mention TCM which can actually treat some diseases (e.g., Topics p7 & p10). Thirdly, they list some health preserving theories in TCM (e.g., Topics p4, p5 & p9). Fourthly, they use the national policy and the curriculum setup in colleges and universities to demonstrate the scientific nature of TCM (e.g., Topics p2 & p8). Additionally, rude Internet behaviors appear during the debate (e.g., Topic p6).

Table 5 shows the ten topics from the "abolishing TCM" replies. There are mainly five groups of topics from the "abolishing TCM" stance holders. Firstly, people in this

camp doubt the rationality of TCM, especially the theories of *yin-yang* and five elements, feeling the pulse and acupuncture points (e.g., Topics a6, a7 & a9). Secondly, they emphasize that some TCM contain abnormal materials even materials with toxicity (e.g., Topics a3 & a8). Thirdly, they list some pseudo TCM experts or some people related to illegal practice of medicine (Topic a1), in their opinion we should discard the dross of traditional things (e.g., Topic a2). Fourthly, they introduce the modern medicine, for example the virus theory (e.g., Topic a4). Additionally, rude Internet behaviors also appear in this camp. Some of the "abolishing TCM" stance holders write doggerel to express their opinions (e.g., Topics a5 & a10).

Table 4. Topics from "preserving TCM" replies

No.	Topics	Words related to topics
p1	Motivation of "abolishing TCM"	科学 西医 没有 人类 真理 生命 创新 先生 实验 技术 癌症 能够 转基因 神经 理论
p2	Curriculum setup in colleges	西医 知道 没有 骗子 治病 医学 废除 西药 骗人 东西 理论 大學 否定 反对 中醫藥
p3	Water army	人士 问题 理解 没有 智商 证明 不能 中药 回答 事实 认为 逻辑 青蒿 实践 天涯
p4	Health preserving theory	人體 方法 疾病 患者 沒有 可能 科盲 请问 血氣 网友 中醫 能力 血液 水平 能量
p5	TCM theory	理论 实践 科学 医学 中医 经络 研究 方法 人体 物质 发展 疾病 存在 系统 认识
p6	Rude Internet behaviors	中药 中医 智商 知道 证明 儿童 出来 砒霜 逻辑 东西 板蓝根 告诉 孙子 承认 无效
p7	TCM with good effectiveness	不能 治愈 患者 高血压 血压 大气 甘草 山药 饮食 人参 医学 大便 疗程 知母 下陷
p8	National policy	中医 中医药 国家 医学 中药 医药 医疗 发展 结合 文化 临床 全国 研究 我国 社会
p9	Health preserving theory	治疗 药物 西药 西医 抗生素 疾病 使用 人体 病人 病毒 作用 引起 出现 没有 导致
p10	Specific examples	医院 医生 病人 没有 患者 治疗 时间 检查 手术 知道 时候 问题 结果 情况 认为

6 Further Analysis Based on Camps

In Sect. 4, discriminating words are generated by logistic regression. In Sect. 5, topics from respective camps are generated by topic modeling. For the context of TCM, we focus on the TCM technical terms in this section. Taking the TCM terminology dictionary from Sougou Cell dictionary[7] with 28428 TCM technical terms as reserved words, and filtering out other words, we do logistic regression to get more details of the debate.

[7] http://pinyin.sogou.com/dict/detail/index/20664.

Table 5. Topics from "abolishing TCM" replies

No.	Topics	Words related to topics
a1	Pseudo TCM experts	医院 没有 国家 医疗 行医 中医药 医生 工作 大学 患者 记者 部门 误诊 医学 国际
a2	Reject the dross	问题 世界 东西 祖先 历史 病人 人类 作为 体系 废除 作用 选择 接受 结合 五行
a3	Toxicity of TCM	方法 实验 服用 问题 植物 临床 副作用 毒性 研究 标准 朱砂 含有 动物 毒副作用
a4	Virus theory	研究 原因 禽流感 药物 引起 预防 结果 细菌 医学 方法 检查 死亡 治愈 临床 抗生素
a5	Doggerel	天涯 问题 治病 地方 老子 教养 看到 喜欢 水平 说明 世界 缺乏 相信 支持 浪费
a6	Theory of *yin-yang* and five elements	解释 认为 方法 知识 思想 文化 发展 真理 基础 不能 存在 检验 事物 社会 错误
a7	Feeling the pulse and acupuncture points	告诉 逻辑 养生 证据 傻子 时候 人们 能够 算命 事实 是不是 可能 承认 请问 不能
a8	Strange prescriptions	阴阳 垃圾 方子 回答 月经 明白 人中黄 解释 乾坤 听说 试试 五行 狗屁 不止 看到
a9	Scientific nature of TCM	认为 臆想 概念 五脏 心脏 循环 研究 致病 脏腑 解剖 六淫 组成 胃肠 阴阳 作用
a10	Rude Internet behaviors	医院 网友 手术 粉丝 结合 针灸 没有 疗效 遇到 寿命 治愈 仪器 重病 医治 神医

With the threshold of absolute coefficient 0.2, we get 562 discriminating words from 1049 words, including 305 words with positive coefficients (related to "preserving TCM" stance) and 257 words with negative coefficients (related to "abolishing TCM" stance).

Table 6 lists 12 TCM technical terms with high absolute coefficients from "preserving TCM" perspective. "Preserving TCM" stance holders always mention the TCM theories and philosophies (e.g., Nos. tp1, tp4 & tp7) and the specific medicines which are well known can actually treat some disease (e.g., Nos. tp6 & tp12).

Table 7 lists 12 TCM technical terms with high absolute coefficients from "abolishing TCM" perspective. "Abolishing TCM" stance holders have mainly four groups of technical terms. Group one (e.g., Nos. ta1, ta7, ta8 & ta11) contains those specific abstract conceptions which are difficult to be explained and understood. Group two refers to medical prescriptions contains abnormal materials (e.g., Nos. ta6, ta12) or materials with toxicity (e.g., No. ta4). Group three (e.g., No. ta4) explains that patients may recover themselves. Group four (e.g., Nos. ta3 & ta9) mentions acute diseases which cannot be cured by TCM.

Similarly, we use human names appeared in the corpus to do logistic regression because people usually quote others' sayings or list some human names related to famous events to support their stance in debates.

With the absolute threshold of absolute coefficient 0.2, we get 100 discriminating words from 7459 names, including 48 human names with positive coefficients (related to "preserving TCM" stance) and 52 human names with negative coefficients (related to "abolishing TCM" stance).

Table 6. 12 discriminating technical terms from "preserving TCM" stance

No.	Original Chinese words	Note
tp1	辩证施治	TCM philosophy
tp2	卒中	Illness
tp3	风热	Illness
tp4	脏腑学说	TCM theory
tp5	体外	TCM conceptions
tp6	桂枝汤 (Guizhi Decoction)	TCM prescription
tp7	奇经八脉	TCM theory
tp8	猪脑	TCM prescription
tp9	胃气	Illness
tp10	球后	A specific acupuncture point
tp11	实热	Illness
tp12	金匮要略	An ancient book about TCM

Tables 8 and 9 show 8 human names from each side of the debate by decreasing rank of their absolute coefficients. These human names (e.g., Nos. na1, na2 & na6) are well known to the pubic because they are pseudo experts or related to illegal practice of medicine. The historical figures (e.g., Nos. na4 & na5) are famous TCM practitioners in ancient China. The man (No. na7) is a western medicine doctor who made contribution for conducting epidemic prevention work in the 1910s. In the "preserving TCM" camp, there are mainly two groups of human whose names are referred. Group one (e.g., Nos. np1, np2, np3 & np7) are government administers who support TCM. These people in group 2 (e.g., Nos. na5 & na6) are doctors. Some journalists' names (e.g., Nos. na3 & np8) outperform as their newspaper articles supporting opposite stance are mentioned for many times. Some user IDs of Tianya Forum are referred because they are active participants during the debate (e.g., No. np4).

Table 7. 12 discriminating technical terms from "abolishing TCM" stance

No.	Original Chinese words	Note
ta1	六淫	TCM conception
ta2	康复医学	Rehabilitation medicine
ta3	肋痈	A kind of acute disease
ta4	缓解期	Remission stage
ta5	阅读障碍	Illness
ta6	齿垢 (Denticola)	TCM prescription
ta7	清热解毒	TCM conception
ta8	肝肾阴虚	Illness
ta9	脑出血 (hemorrhage)	A kind of acute disease
ta10	轻粉	A specific TCM
ta11	口舌生疮	Illness
ta12	人中黄	TCM prescription

Table 8. 8 discriminating human names from "preserving TCM" stance

No.	Human names	Note
np1	王国强	One government administer
np2	钱信忠	One government administer
np3	李斌	One government administer
np4	施正义	One active User ID during the TCM debate
np5	倪建俐	One doctor
np6	王拥军	One doctor
np7	温家宝	Former Prime Minister
np8	魏敏	One journalist

Table 9. 8 discriminating human names from "abolishing TCM" stance

No.	Human names	Note
na1	张悟本	One pseudo health expert
na2	闫芳	One pseudo Tai Chi and Kung Fu expert
na3	孙国根	One journalist
na4	华佗	One famous TCM practitioner in ancient China
na5	孙思邈	One famous TCM practitioner in ancient China
na6	胡万林	One pseudo health expert
na7	伍连德	One western medicine doctor
na8	刘海若	One journalist of a television station

7 Conclusions

This study explores a stance mining problem about a debate on societal issue TCM. We select one hot post on TCM from one of the most influential Chinese BBS, Tianya Forum, and automatically determine the replies' stances about TCM. Our results show that logistic regression can effectively select domain feature words and identify replies' stance with precision of 63.13 %, outperforming the SVM model using adjectives, adverbs, verbs and nouns as features.

Secondly, our topic modeling by LDA reveal that the emphases of the two camps are different during the debate. The "preserving TCM" stance holders concern the motivations of the other camp, the effectiveness of the TCM, etc. The "abolishing TCM" stance holders doubt the scientific nature and the rationality of TCM, introduce the modern medicine, and condemn the illegal medical practice relevant to TCM.

Thirdly, our further analysis verifies meanings of specific discriminating words present during the debate by logistic regression. The details of the concerned technical terms and human names in the different camps let us see how people express their viewpoints and perspectives during the TCM debate.

This paper provides an example for future research designed to explore stances on societal issues. In the future, we will do more study on identifying stance by interactions within debate and how opposing perspectives and arguments are put forward during debates.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (Nos. 61473284 and 71371107).

References

1. Riloff, E.: Automatically generating extraction patterns from untagged text. In: 13th National Conference on Artificial Intelligence, Portland, pp. 1044–1049 (1996)
2. Riloff, E., Wiebe, J.: Learning extraction patterns for subjective expressions. In: Conference on Empirical Methods in Natural Language Processing, Sapporo, pp. 105–112 (2003)
3. Cui, H., Mittal, V., Datar, M.: Comparative experiments on sentiment classification for online product reviews. In: 21st National Conference on Artificial Intelligence, Boston, pp. 61–80 (2006)
4. Ng, V., Dasgupta, S., Arifin, S.M.N.: Examining the role of linguistic knowledge sources in the automatic identification and classification of reviews. In: International Conference on Computational Linguistics and Meeting of the Association for Computational Linguistics, Sydney, pp. 381–393 (2006)
5. Gamon, M.: Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis. In: 23rd International Conference on Computational Linguistics, Beijing, pp. 841–847 (2010)
6. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? sentiment classification using machine learning techniques. In: Conference on Empirical Methods in Natural Language Processing, Philadelphia, pp. 79–86 (2009)
7. Liu, B.: Opinion mining and sentiment analysis. Found. Trends Inf. Retrieval **2**(1–2), 1–135 (2008)
8. Zhao, Y.L., Tang, X.J.: In-depth analysis of online hot discussion about TCM. In: 15th International Symposium on Knowledge and Systems Science, pp. 275–283. JAIST Press, Sapporo (2014)
9. Turney, P.D., Littman, M.L.: Measuring praise and criticism: inference of semantic orientation from association. ACM Trans. Inf. Syst. **21**(4), 315–346 (2003)
10. Somasundaran, S., Wiebe, J.: Recognizing stances in ideological on-line debates. In: NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, Los Angeles, pp. 116–124 (2010)
11. Abbott, R., Walker, M., Anand, P., et al.: How can you say such things?!?: recognizing disagreement in informal political argument. In: Workshop on Languages in Social Media, Portland, pp. 2–11 (2011)
12. Anand, P., Walker, M., Abbott, R., et al.: Cats rule and dogs drool!: classifying stance in online debate. In: 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, Portland, pp. 1–9 (2011)
13. Walker, M.A., Anand, P., Abbott, R., et al.: That is your evidence? Classifying stance in online political debate. Decis. Support Syst. **53**(4), 719–729 (2012)
14. Tikves, S., Banerjee, S., Temkit, H., et al.: A system for ranking organizations using social scale analysis. Soc. Netw. Anal. Min. **3**(3), 313–328 (2013)
15. Tikves, S., Gokalp, S., Temkit, M., et al.: Perspective analysis for online debates. In: International Conference on Advances in Social Networks Analysis and Mining, Istanbul, pp. 898–905 (2012)
16. Gryc, W., Moilanen, K.: Leveraging textual sentiment analysis with social network modeling: sentiment analysis of political blogs in the 2008 U.S. Presidential Election. In: Workshop on from Text to Political Positions, Amsterdam (2010)

17. Lin, W.H., Wilson, T., Wiebe, J.: Which side are you on? Identifying perspectives at the document and sentence levels. In: 10th Conference on Computational Natural Language Learning, New York, pp. 109–116 (2006)
18. Hatzivassiloglou, V., Wiebe, J.M.: Effects of adjective orientation and gradability on sentence subjectivity. In: International Conference on Computational Linguistics, Mexico, pp. 299–305 (2003)
19. Benamara, F., Cesarano, C., Picariello, A., et al.: Sentiment analysis: adjectives and adverbs are better than adjectives alone. In: Veselovská, K., Hajic, J., Šindlerová Bojar, O., Žabokrtský, Z. (eds.) International Conference on Weblogs and Social Media, Boulder (2007)
20. Subrahmanian, V.S., Reforgiato, D.: AVA: Adjective-verb-adverb combinations for sentiment analysis. IEEE Intell. Syst. **23**(4), 43–50 (2008)
21. Shen, J., Zhu, P., Fan, R., Tan, W., Zhan, X.: Sentiment analysis based on user tags for traditional Chinese medicine in Weibo. In: Li, J., et al. (eds.) NLPCC 2015. LNCS, vol. 9362, pp. 134–145. Springer, Heidelberg (2015). doi:10.1007/978-3-319-25207-0_12
22. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)