S-Sense: A Sentiment Analysis Framework for Social Media Sensing

Choochart Haruechaiyasak, Alisa Kongthon,
Pornpimon Palingoon and Kanokorn Trakultaweekoon
Speech and Audio Technology Laboratory (SPT)
National Electronics and Computer Technology Center (NECTEC)
Thailand Science Park, Klong Luang, Pathumthani 12120, Thailand
choochart.har@nectec.or.th, alisa.kon@nectec.or.th
pornpimon.pal@nectec.or.th, kanokorn.tra@nectec.or.th

Abstract

Due to the explosive growth of social media usage in Thailand, many businesses and organizations including market research agencies are seeking for tools which could perform real-time sentiment analysis on the large contents. In this paper, we propose *S-Sense*, a framework for analyzing sentiment on Thai social media. The proposed framework consists of analysis modules and language resources. Two main analysis modules, *intention* and *sentiment*, are based on classification algorithm to automatically assign appropriate intention and sentiment class labels for a given text. To train classification models, language resources, i.e., corpus and lexicon, are needed. Corpus consists of a collection of texts manually labeled with appropriate intention and sentiment classes. Lexicon consists of both general terms from dictionary and clue terms which help identifying the intention and sentiment. To evaluate performance and robustness of the analysis modules, we prepare a data set from *Twitter* posts and *Pantip* web board in mobile service domain. The experiments are set up to compare the performance between two different lexicon sets, i.e., general and clue terms. The results show that incorporating clue terms into feature vectors for constructing the classification models yield significant improvement in terms of accuracy.

1 Introduction

Due to the enormous volume, social media has become recognized as a good example of *Big Data*. One of the challenging issues in handling big data is to perform real-time analysis on the contents. Today social media has been widely accepted as an active communication channel between companies and customers. Many companies regularly use social networking websites to promote new products and services, and post announcements to the customers. On the other hand, customers often post their comments to express some sentiments towards products and services. Many customers also post questions and requests to get answers and helps from the customer services. Due to the real-time nature of the social media, monitoring customers' comments has become a critical task in customer relation management (CRM). Sentiment analysis has received much attention among market research community as an effective approach for analyzing social media contents. Some highlighted applications of sentiment analysis include brand monitoring, campaign monitoring and competitive analysis.

Thailand is among the top countries having a large population on social networking websites such as *Facebook* and *Twitter*. The recent statistics show that the number of Facebook users in Thailand has reached 17 millions as of October 29, 2012[1]. Many companies in Thailand start to see the importance of using social media analysis to gain some insight on what people think about their brands, products and services. Although many commercial software tools for social media analysis are available, they do not support Thai language. In this paper, we propose *S-Sense*, a framework for analyzing sentiment on Thai social media contents. To provide a complete solution, our proposed framework consists of many components including tagging tool, language resources, analysis and visualizing modules.

Among all of the components in S-Sense, language resources are considered very essential for providing the infrastructure to train both inten-

[1]Facebook statistics, *http://en.wikipedia.org/wiki/Facebook_statistics*

IJCNLP 2013 Workshop on Natural Language Processing for Social Media (SocialNLP), pages 6–13,
Nagoya, Japan, 14 October 2013.

tion and sentiment analysis models. In our proposed framework, language resources consist of two components, corpus and lexicon. Corpus consists of a collection of texts manually labeled with appropriate intention and sentiment classes. Lexicon consists of two types of terms, general and clue. The general lexicon includes terms found in *LEXiTRON*[2], which is a well-known Thai-English electronic dictionary. In S-Sense, the general lexicon is modified by including new terms such as slangs, chat language, transliterated words, found in Thai Twitter corpus. The second lexicon consists of clue terms which help identifying the intention and sentiment. Example of clue terms for sentiment analysis are polar terms (such as "stylish", "beautiful" and "expensive"), which contain either positive or negative sentiment.

For the analysis modules, we apply classification algorithm to automatically assign appropriate intention and sentiment class labels for a given text. The performance of classification models generally depends on the choice of classification algorithms including parameter settings, the size of training corpus and the design of term feature sets. The current version of S-Sense applies the multinomial Naive Bayes algorithm. The reason we used Naive Bayes is its requirement of a small amount of training data to estimate the parameters for learning the models. Also Naive Bayes is a descriptive and probabilistic machine learning, therefore, the results could be easily analyzed and explained. The classification results are returned with a probability value which could be interpreted as the confidence level. In addition to the proposed framework, another contribution of this paper is the comparative study of using different lexicon sets for training the analysis models. We compare the performance of intention and sentiment analysis models by using two different sets of lexicons, general and clue terms. The evaluation corpus consist of *Twitter* posts and *Pantip* web board topics in mobile service domain. The experimental results will be presented along with the discussion on the error analysis.

The remainder of this paper is organized as follows. In next section, we review some related works on sentiment analysis and many different approaches for constructing language resources for sentiment analysis. In Section 3, we present the proposed S-Sense framework for Thai intention and sentiment analysis. Details on each components are given with illustration. In Section 4, we evaluate the framework by using a data set collected from Twitter and Pantip Thai web board. Examples of difficult cases are discussed along with some possible solutions. Section 5 concludes the paper with the future work.

2 Related work

Due to its potential and useful applications, opinion mining and sentiment analysis has gained a lot of interest in text mining and NLP communities (Ding et al., 2008; Jin et al., 2009; Tsytsarau and Palpanas., 2012). Much work in this area focused on evaluating reviews as being positive or negative either at the document level (Pang et al., 2002; Beineke et al., 2004) or sentence level (Kim and Hovy, 2004; Wilson et al., 2009). For instance, given some reviews of a product, the system classifies them into positive or negative reviews. No specific details or features are identified about what customers like or dislike. To obtain such details, a *feature-based* opinion mining approach has been proposed (Hu and Liu, 2004).

The problem of developing subjectivity lexicons for training and testing sentiment classifiers has recently attracted some attention. Although most of the reference corpora has been focused on English language, work on other languages is growing as well. Ku and Chen (2007) proposed the bag-of-characters approach to determine sentiment words in Chinese. This approach calculates the observation probabilities of characters from a set of seed sentiment words first, then dynamically expands the set and adjusts their probabilities. Later in 2009, Ku et al. (2009), extended their bag-of-characters approach by including morphological structures and syntactic structures between sentence segment. Their experiments showed better performance of word polarity detection and opinion sentence extraction. Haruechaiyasak et al. (2010), proposed a framework for constructing Thai language resource for feature-based opinion mining. The proposed approach for extracting features and polar words is based on syntactic pattern analysis.

Our main contribution in this paper is the proposed framework for analyzing intention, sentiment, and language usage from social media texts. We initially performed some evaluation on Thai texts to show the effectiveness of the proposed

[2]LEX*i*TRON, *http://lexitron.nectec.or.th*

components and modules. The proposed framework can be easily extended to support other languages, especially for unsegmented languages, by providing the plugged-in resources including lexicon and corpus.

3 The proposed framework

In this paper, we focus on both language resources and the analysis modules as a complete framework for Thai-language intention and sentiment analysis. The proposed framework could easily be extended to support other languages by constructing language-specific resources. Our framework is also designed for easy adaptation to businesses in different domains. Similar to language-specific support, to apply the proposed framework for a specific domain, one can use the provided tagging tool to prepare domain-specific resources, i.e., annotated corpus and lexicon.

3.1 Components and modules

The proposed S-Sense framework (shown in Figure 1) consists of the following components.

- **Text collecting & processing**: This component involves the process of crawling and collecting social media contents from different websites. The process includes basic text processing, i.e., sentence segmentation, tokenization and normalization. Term normalization is the process of converting a word as appeared in the text into a predefined term and cleaning extra repeated characters which are not part of the term. For example, a word "thnxsss" can be normalized to the term "thank".

- **UREKA**: The main task of UREKA (Utilization on REsource for Knowledge Acquisition) is to extract key feature terms or phrases from a given text. Terms or phrases which are statistically significant in the corpus can be presented as interesting issues to the users. Another task is to filter and classify a given text into a topic. When collecting texts from social networking websites, it is very common to see many collected texts are not relevant to the brands or products being monitored. Therefore, a classification model could be trained to filter out the irrelevant texts from the corpus. After obtaining the relevant texts, another classification model could be trained to classify each text into a predefined set of topics. For example, in mobile service domain, topics could include signal quality, promotion and customer service.

- **S-Sense**: This is the main analysis component under the framework. S-Sense consists of four analysis modules. Language usage analysis classifies each text based on two aspects, the use of obscene language and the use of chat or informal languages. Detecting obscenity is useful since many texts with strongly negative sentiment could sometimes contain obscene language. Intention analysis classifies each text into four classes: *announcement*, *request*, *question* and *sentiment*. Sentiment analysis further classifies each text based on its sentiment, i.e., positive or negative. Emotion analysis is set in our future work. The task of emotion analysis module is to perform an in-depth sentiment analysis regarding to the emotion or feeling such as sad, happy and angry. Other components of S-Sense include visualizing modules including adaptive emoticon and interactive dashboard. These modules are used for displaying the summarized reports for the analyzed texts.

- **Tagging tool and language resources**: Under the proposed framework, language resources include two components, annotated corpus with domain and language-specific lexicons. To construct language resources, we provide a tagging tool for linguists to work with. The tagging tool is a web-based application which consists of a DBMS and a GUI.

3.2 Analysis tasks

The current version of S-Sense framework focuses on two main analysis modules, intention and sentiment. The intention analysis include the following categories.

1. **Announcement**: This type of intention refers to messages or posts in which a company intends to communicate with their customers, e.g., advertisement of new products or event announcement.

2. **Request**: This intention is used for customers to ask for help when having trouble or problem with the company's products or services.

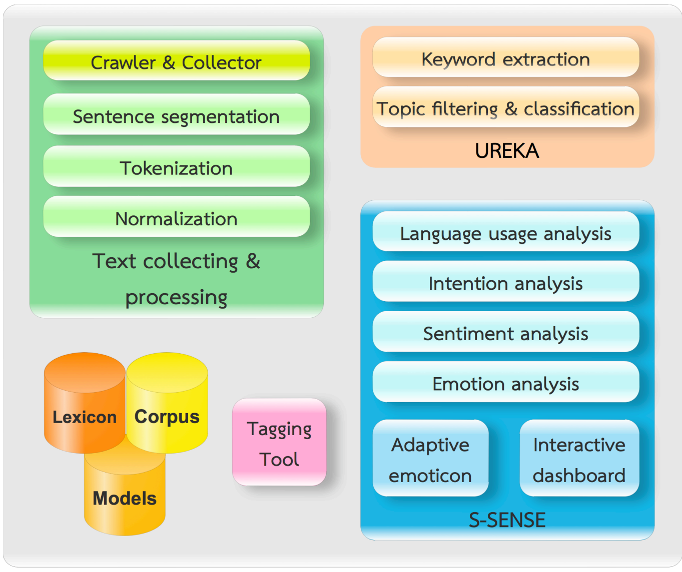

Figure 1: The proposed S-Sense framework.

Customers would expect immediate response from the company to solve the problem.

3. **Question**: This intention refers to messages or posts from customers asking for information related to products and services. The question is, for example, a customer's post asking for more details of a new mobile service promotion.

4. **Sentiment**: This intention is when customers express their opinions or sentiments towards the company's brand, products and services. Sentiment can be divided into positive, neutral and negative aspects.

It is important to analyze intention before performing sentiment analysis. Without intention analysis, a sentence containing positive polar words such as an advertisement would be identified as containing the sentiment intention. For example, a sentence "The new high-speed Internet is faster and cheaper. Apply today at the shop near you." is an advertisement, but could be incorrectly identified as having positive sentiment. Therefore, Identifying a sentence as announcement or advertisement would help improve overall performance of sentiment analysis.

3.3 Potential applications

S-Sense can be applied in many different applications. Some of the potential applications are as follows.

- **Brand monitoring**: With the widespread of social media, today customers have more freedom to express their sentiments towards products and services. Analyzing sentiments of the customers could help companies gain some insight on how they feel when using their products and services. More importantly, many companies are highly associated with their brands. Negative sentiments towards the company's brand could have negative impact on the product sales. Therefore, it is very important for companies to monitor or track the mentions and sentiments of the customers on social media.

- **Campaign Monitoring**: Many times throughout the year, the company would launch different campaigns involving new products and services. The goal of campaign monitoring (i.e, tracking) is to measure the customers' feedback on each campaign. The results could be analyzed in terms of number of mentions, positive and negative sentiments and the key product or service features in which customers feel positive or negative about.

- **Competitive Analysis**: This task is to monitor and analyze the activities including sentiments of customers towards the company's competitors. The analysis results could help gain some insight on strengths and weaknesses of the competitors in the market. For example, if a competitor has many complaints on certain product features, the company could grab the opportunity by advertising its own product features which are better than the competitor's.

- **Employee Engagement**: One of the main problems in many organizations today is the high turnover rate. One of the solutions is to monitor and analyze the employee engagement level. This task is to measure the employees' sentiments towards their jobs, colleagues and organization. The measure could reveal how much employees are willing to learn and perform at work, and to get involved in different activities initiated by the organization.

4 Experiments and discussion

To evaluate the proposed framework, we perform experiments using a corpus in the domain of mobile service. The corpus is obtained between March and June in 2013 from two sources, *Twitter*[3] and *Pantip*[4], one of the top visited web boards in Thailand. The total number of randomly selected texts in the corpus is *2,723*. The corpus was annotated in two aspects, intention and sentiment. Table 1 summarizes the number of tagged texts in four different intentions. The majority of intentions is sentiment which accounts for approximately 64% of the corpus. The reason is when using social networking websites or web boards, users often express their opinion and sentiment more than other intentions.

For the sentiment intention, we further annotated each text based on its sentiment, i.e., positive or negative. Table 2 summarizes the number of tagged texts in positive and negative sentiment. It can be observed that negative sentiment accounts for approximately 91%. This is not very surprising since users tend to complain when having problems using the mobile service. Major reported problems in mobile service industry include, for example, weak or unavailable signal, call drop, slow data transfer rate, impolite service and long waiting time for call center.

Table 3 shows some examples of annotated corpus in different intention and sentiment. In addition to annotating each text with an intention label,

Intention	# Texts
Announcement	94
Request	405
Question	456
Sentiment	1,768
Total	**2,723**

Table 1: Number of tagged texts in four different intentions.

Sentiment	# Texts
Positive	156
Negative	1,612
Total	**1,768**

Table 2: Number of tagged texts in positive and negative sentiments.

we collect clue terms which could help identify the intention. For example, from the announcement intention, the terms and phrases "new promotion", "best-deal" and "will start on" are collected into the clue lexicon. From the sentiment intention, we collected the terms "annoyed" and "impressive". Other clue terms are underlined for each example in the table.

Table 4 shows the statistics of lexicons used in the experiments. There are two types of lexicons: *general* and *clue* terms. General lexicon include two sets of terms, LEX*i*TRON[5] which are general words from Thai dictionary, and *Twitter* which contains newly found words from Thai Twitter corpus. Words obtained from Twitter include slangs and transliterated words from other languages. Clue lexicon include terms or phrases which could help identify intention and sentiment. One of the main objectives in the experiments is to observe the effect of incorporating clue lexicon in constructing classification models for intention and sentiment analysis. Therefore, we perform a comparative study on using different sets of lexicons.

To perform experiments, we apply the multinomial Naive Bayes algorithm to learn the classification models (McCallum and Nigam, 1998). The reason we use Naive Bayes is due to the small number of sample texts in the corpus, especially

[3]Twitter, *http://twitter.com*
[4]Pantip, *http://pantip.com*

[5]LEX*i*TRON, *http://lexitron.nectec.or.th*

Intention		Example
Announcement		อัตราค่าบริการ Happy Bonus ปรับปรุงใหม่จ๊ะ เริ่มใช้วันที่ 1 ค่ะ The new service fee for Happy Bonus will start on the 1st of this month.
		โปรใหม่!! ทรูมูฟ... ซิมสุดคุ้ม โปรวินาทีละ 1 ส.ต. ตลอด 24 ชั่วโมง New promotion!! True Move... Best-deal SIM, 1 satang / second all day and night.
Request		สมัครใช้บริการ Call Screening เองไม่ได้ CC ช่วยด้วยครับ I can't apply for Call Screening myself. CC (Call Center), please help me.
		รบกวนCC AISหน่อยค่ะ..เงินในโทรศัพท์หายไปไหนไม่รู้ (- -)?? AIS Call Center, please.. My pre-paid balance has gone missing without a clue ??
Question		โทรศัพท์หาย จะทำซิมใหม่เบอร์เดิมของ ais ต้องใช้เอกสารอะไรบ้างครับ I lost my phone. To get a new SIM card, what documents are required?
		โปรไหนของ one-2-call ที่รอรับสายได้นานสุดครับ Which promotion package of one-2-call allows the longest call waiting time?
Sentiment	Negative	น่ารำคาญมาก DTAC เมื่อไหร่จะปรับปรุงสัญญาณสักที โดยเฉพาะบนBTS Very annoyed. DTAC, when will you improve the signal? Especially on the BTS.
	Positive	ขอบคุณและชื่นชม เจ้าหน้าที่ AIS serenade call center ประทับใจครับ Thank you to the operator at AIS serenade call center. Very impressive.

Table 3: Example of annotated texts categorized by different intentions and sentiments.

Lexicon		# Terms
General	Lexitron	35,328
	Twitter	1,341
Clue	Announcement	86
	Request	177
	Question	454
	Polar (Negative)	1,675
	Polar (Positive)	1,237

Table 4: Two types of lexicons: *general* and *clue*

for the announcement intention. Naive Bayes only requires a small amount of training data to estimate the parameters for learning the models. Also Naive Bayes is a descriptive and probabilistic machine learning, therefore, the results could be easily analyzed and explained. The classification results are returned with a probability value which could be interpreted as the confidence level.

The first experiment is the intention analysis. For each intention listed in Table 1, we train a binary classification model with two classes, *related* and *other*. If a given text is analyzed as containing a particular intention, it will be assigned with the class label *related*. We prepare the data set by using the same amount of texts in each class. For example, in announcement intention, we use 94 announcement texts and randomly select another 94 texts from other intentions. To see the advantage of using clue terms as additional term feature, we compare the results between using only general lexicon and using both general and clue lexicons. The performance metric is *accuracy* which is defined as the number of correctly classified instances over the total number of test instances.

Table 5 shows the experimental results for intention analysis. The results are based on 10-fold cross validation. From the table, it can be observed that adding clue terms into the term feature helps improve the classification accuracy for all intentions. Especially for request, question and sentiment, the improvement is over 6%. For announcement, the improvement is approximately 2%. This is probably due to the difficulty in defining and collecting the clue terms for announcement intention. For example, some of the terms like "new" must be collocated with other term in a phrase, e.g. "new promotion". As the phrase becomes more specific, it will not be found in the test instances. Another observation is the request intention is the most difficult to analyze. This is due to often when users wish to request for something, there is no specific term or clue term in the message. The request intention is implicitly expressed with verbs or polar terms, therefore causing confusion to other intention classes.

Intention	Term feature	Accuracy (%)
Announcement	General	78.72
	General + Clue	**80.85**
Request	General	63.08
	General + Clue	**69.38**
Question	General	73.13
	General + Clue	**79.82**
Sentiment	General	67.47
	General + Clue	**73.61**

Table 5: Experimental results on intention analysis

The second experiment is the sentiment analysis. We train a binary classification model with two classes, *positive* and *negative*. The number of instances for each class is given in Table 2. Table 6 shows the experimental results on sentiment analysis. The results are based on 10-fold cross validation. From the table, we can observe that using clue terms as additional term features helps increase the accuracy by approximately 2%. The small amount in improvement is probably due to terms in general dictionary and from Twitter contain sentiment which already helps identify the polarity of the texts.

Term feature	Accuracy (%)
General	89.55
General + Clue	**91.64**

Table 6: Experimental results on sentiment analysis

To perform error analysis, we look at the test instances which are misclassified, i.e., classifying positive into negative and vice versa. We can summarize two major causes of errors as word sense ambiguity and sarcasm. The first problem occurs when a polar term contains both positive and negative senses depending on the contexts. For example, the word "strong", when appearing with the term "signal" will give positive polarity. However, when it appears with the term "employee", the term has the meaning of "impolite" and a negative polarity should be assigned. However, due to the small corpus size and simple feature vector which treats each term independent, sometimes, the terms cannot be learned properly. To solve this problem, we will explore the idea of incorporating contextual terms with the clue terms in our future work. Each clue term will be associated with some context terms to identify the polarity of the texts.

The second problem is sarcasm which is much more difficult to solve. This problem is still a difficult and challenging task in sentiment analysis of any languages (González-Ibáñez et al., 2004). While there are some research work to identify sarcasm in given texts, the performance is still poor. However, some of the sarcastic texts can still be identified by detecting some common slangs which are usually used in sarcastic texts. In Thai language, if users express a positive sentiment in an exaggerated way or in a contradicting way, then the message is most likely sarcastic. For example, "Today the download speed is faster than the speed of light. Thank you very much!" is considered as sarcastic.

5 Conclusion and future work

We proposed a framework called *S-Sense* (Social Media Sensing) for developing a social media analyzing tool. The current version focuses on intention and sentiment analysis. We applied the Naive Bayes as the classification algorithm to analyze four different intentions (announcement, request, question and sentiment) and two sentiments (positive and negative). The proposed framework was evaluated by using a social media corpus in the domain of mobile service obtained from *Twitter* and *Pantip* web board.

To study the effect of using different lexicon sets to train the models, we compared two approaches: using only general lexicon and using both general lexicon and clue terms. The results showed that adding clue terms into feature vector for training the classification models helps improve the accuracy for all intention and sentiment analysis models. For intention models of request, question and sentiment, the accuracy is increased by approximately 6%. For sentiment model, the accuracy is increased by approximately 2%.

From the error analysis, we found that two major problems are word sense ambiguity and sarcasm. For future work, we plan to improve the performance of both intention and sentiment analysis models by incorporating the contexts nearby the clue terms. Considering contexts could help reduce the disambiguation of the word sense. Another plan is to construct the lexicon and corpus for other different domains. In addition to mobile service, other business domains in Thailand often mentioned in the social media are automotive, consumer electronics, fashion, healthcare and tourism.

References

Philip Beineke, Trevor Hastie and Shivakumar Vaithyanathan. 2004. The sentimental factor: improving review classification via human-provided information. *Proc. of the 42nd Annual Meeting on Association for Computational Linguistics*, 263–270.

Xiaowen Ding, Bing Liu and Philip S. Yu. 2008. A holistic lexicon-based approach to opinion mining. *Proc. of the int. conf. on web search and web data mining*, 231–240.

Roberto González-Ibáñez, Smaranda Muresan and Nina Wacholder. 2011. Identifying sarcasm in Twitter: a closer look. *Proc. of the 49th ACL: Human Language Technologies*, 581–586.

Choochart Haruechaiyasak, Alisa Kongthon, Pornpimon Palingoon, Chatchawal Sangkeettrakarn. 2010. Constructing Thai Opinion Mining Resource: A Case Study on Hotel Reviews. *Proc. of the Eighth Workshop on Asian Language Resources*, 64–71.

Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. *Proc. of the 10th ACM SIGKDD international conference on Knowledge discovery and data mining*, 168–177.

Wei Jin, Hung Hay Ho and Rohini K. Srihari. 2009. OpinionMiner: a novel machine learning system for web opinion mining and extraction. *Proc. of the 15th ACM SIGKDD*, 1195–1204.

Soo-Min Kim and Eduard Hovy. 2004. Determining the sentiment of opinions. *Proc. of the 20th international conference on Computational Linguistics*, 1367–1373.

Lun-Wei Ku and Hsin-Hsi Chen. 2007 Mining opinions from the Web: Beyond relevance retrieval. *Journal of American Society for Information Science and Technology*, 58(12):1838–1850.

Lun-Wei Ku, Ting-Hao Huang and Hsin-Hsi Chen. 2009. Using morphological and syntactic structures for Chinese opinion analysis. *Proc. of the 2009 empirical methods in natural language processing*, 1260–1269.

Andrew McCallum and Kamal Nigam. 1998. A Comparison of Event Models for Naive Bayes Text Classification. *Proc. of the AAAI-98 Workshop on 'Learning for Text Categorization'* 41–48.

Bo Pang, Lillian Lee and Shivakumar Vaithyanathan. 2002. Thumbs up?: sentiment classification using machine learning techniques. *Proc. of the ACL-02 conf. on empirical methods in natural language processing*, 79–86.

Mikalai Tsytsarau and Themis Palpanas. 2012. Survey on mining subjective data on the web. *Data Mining and Knowledge Discovery*, 24(3): 478–514.

Theresa Wilson, Janyce Wiebe and Paul Hoffmann. 2009. Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Comput. Linguist.*, 35(3):399–433.