Contextual Sentiment Topic Model for Adaptive Social Emotion Classification

Yanghui Rao, *Sun Yat-sen University*

The contextual sentiment topic model (CSTM) distinguishes context-independent topics from both a background theme, which characterizes nondiscriminative information, and a contextual theme, which characterizes context-dependent information across different collections.

Measuring the opinions of the general public about social events, company strategies, marketing campaigns, and product preferences has raised steady interest in various scientific communities.[1] Traditional methods required questioning a large number of people about their feelings using polls, which, although accurate, can be expensive or time-consuming. With the growth of the social Web and the availability of reviews, statistical polling data, and other user-generated content, an increasingly large volume of significant information about user opinions is being stored online.[2] In addition, interactive services spread the communication of user emotions through news websites, blogs, microblogs/tweets, and so forth. In fact, many news websites now provide a service that lets users convey their emotions after browsing a news article,[3] and then records the emotional responses shared by readers from a set of predefined emotion labels incrementally. The aggregation of such emotional responses is known as *social emotions*.[3] Figure 1 shows an example of the emotional responses of 3,064 readers after browsing a news article published on Sina, a popular news website in China. Among eight emotion labels, the category of "amusement" is the highest rated (55.8 percent).

Given labeled data in a certain context, adaptive social emotion classification is concerned with a reader's emotional classification of unlabeled data in another context. Online data is found in many domains and contexts such as economics, entertainment, or sports; adaptive social emotion classification uses a model trained on one source context to build a model for another target context. This is challenging because topics that evoke a certain emotion in readers are often context-sensitive[2]—for example, the topic of price reduction could cause readers in the context of selling to feel depressed. On the other hand, in a buying context, a positive mood could be triggered in response to the same topic.

Although several cross-domain sentiment classification algorithms[4–6] have been proposed to tackle the context-adaptation problem, they're designed primarily from the perspective of writers (see the "Related Work in Sentiment Analysis and Cross-Domain Sentiment Classification" sidebar).[7] This

1541-1672/16/$33.00 © 2016 IEEE
Published by the IEEE Computer Society

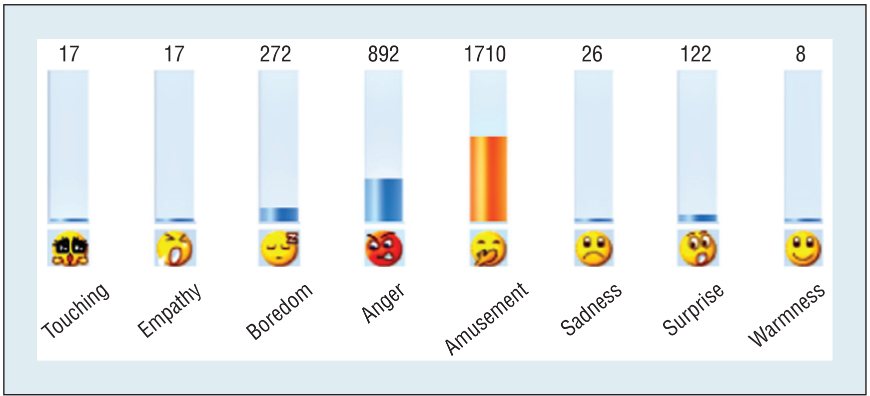

Figure 1. An example of social emotions. In the emotional responses of 3,064 readers of a news article, the category of "amusement" is the highest rated (55.8 percent).

article addresses the problem of adaptive social emotion classification from the reader's perspective, in which the emotional distribution of readers isn't always consistent with that of writers. A sentence with no emotional words—a neutral sentence from the writer's perspective—can evoke a positive or negative emotion in readers.[8] In light of these considerations, our team at Sun Yat-sen University developed a contextual sentiment topic model (CSTM) to classify reader emotions across different contexts.

Contextual Sentiment Topic Model

We propose a multilabeled sentiment topic model—the CSTM—for adaptive social emotion classification.

Problem Formulation

We formulated the problem of adaptive social emotion classification as follows. Given two document collections with specific contexts D_{src} and D_{tar}, where D_{src} and D_{tar} are the labeled and unlabeled documents in a source context and a target context, respectively, the task of adaptive social emotion classification is to learn an accurate classifier from D_{src} to predict the evoked emotions of unseen documents in D_{tar}. Because a classifier trained from one context might not perform as well in another, we designed the CSTM by distilling useful information from such collections. We denoted the set of unique words contained in all documents as V, the number of topics as T, the amount of emotion labels as $|E|$, and the union set of D_{src} and D_{tar} as D. The main difference when applying other models versus our model to documents written in different languages is the segmentation of words. Unlike English words, consecutive Chinese words in a document aren't separated by spaces. To deal with this, Chinese word segmentation tools can be used to extract words from documents.[3] To further describe the model, we also defined the notations in Table 1.

Proposed Framework

The CSTM's objective is to distinguish explicitly generalized topics that are context-independent from both a background theme, which characterizes nondiscriminative information and a contextual theme, which characterizes context-dependent information across different collections. As Figure 2 shows, the context-independent topics $z_1, z_2, \ldots, z_T$ are controlled by $\pi_{d,j}$, that is, the probability of using topic z_j when generating words in document d. The parameters λ_B and λ_C control the strength of generating words from θ_B and θ_C, respectively, satisfying $0 < \lambda_B + \lambda_C < 1$.

Weighting parameter λ_B reflects the degree of common information in different contexts, which is used to "absorb" noninformative words from topics and can be estimated by word co-occurrence. The larger the ratio of words co-occurring in D_{src} and D_{tar}, the higher the value of λ_B. Thus,

$$\lambda_B = \frac{|\{w \mid w \in D_{src} \cap w \in D_{tar}\}|}{|\{w \mid w \in D\}|}. \qquad (1)$$

Weighting parameter λ_C indicates our emphasis on the specialty of a certain context in comparative collections. The value of λ_C depends on the variety of D_{src} and D_{tar}:

$$\lambda_C = \frac{|\{w \mid w \in D_{src} \cap w \notin D_{tar}\}|}{|\{w \mid w \in D\}|}. \qquad (2)$$

To demonstrate the computation of λ_B and λ_C, consider the following simple example. Assume that the sets of words contained in D_{src} and D_{tar} are $\{w_1, w_2, w_3\}$ and $\{w_2, w_4\}$, respectively. For the set of words contained in all documents, $\{w_1, w_2, w_3, w_4\}$, the degree of common information across different contexts $\lambda_B = |\{w_2\}|/|\{w_1, w_2, w_3, w_4\}| = 0.25$, and the extent of variety $\lambda_C = |\{w_1, w_3\}|/|\{w_1, w_2, w_3, w_4\}| = 0.5$.

Each word w in document d is generated by the following mixture model:

$$p_d(w) = \lambda_B p(w \mid \theta_B) + \lambda_C p(w \mid \theta_C) + (1 - \lambda_B - \lambda_C) \sum_{j=1}^{T} (\pi_{d,j} p(w \mid z_j)), \qquad (3)$$

where $p(w|\theta_B)$ is the probability of word w conditioned to the background theme θ_B, and $p(w|\theta_C)$ is the probability of word w conditioned to the contextual theme θ_C. When we use the CSTM to generate a word, it intuitively first decides whether to use θ_B or θ_C according to λ_B and λ_C, respectively. If the model decides not to use θ_B or θ_C, then it needs to decide which topic to use; this is represented by $p(w|z_j)$, the probability of word w

Related Work in Sentiment Analysis and Cross-Domain Sentiment Classification

Before discussing works related to social emotion detection, let's first review research related to sentiment analysis and cross-domain sentiment classification.

Sentiment Analysis

The Internet contains important information on its users' sentiments and opinions. The extraction of such unstructured Web data is known as *sentiment analysis* and *opinion mining*. Sentiment analysis identifies and extracts the attitude of a subject—an opinion holder—toward either a topic or a document's overall tone. Preliminary works on sentiment analysis aimed to classify entire documents as containing overall positive or negative polarity, or rating scores of reviews.[1–3] The limitation of these studies, however, was that a classifier trained for use in one domain might not perform as well in a different one.[4]

Cross-Domain Sentiment Classification

Sentiment analysis is very domain-specific in supervised learning classification because users can use domain-specific words to express certain sentiments in different domains.[5] For example, in the domain of electronic product reviews, the word "compact" is often used to express a positive sentiment. In the books domain, the word "exciting" expresses a primarily positive sentiment. Cross-domain sentiment classification[4–6] could tackle this domain-specific problem. Unfortunately, it's mainly suitable for annotating emotions from the writer's perspective.[7]

Social Emotion Detection

Research into social emotion detection began with the "affective text" in SemEval-2007 tasks,[8] which focused on exploiting reader emotions with individual words. The limitation of early word-level studies was that different senses could be evoked from the same word. To deal with this, an emotion LDA (ELDA) model[9] was developed by associating emotions with topics. Such a topic was generated by Latent Dirichlet Allocation (LDA).[10] However, unlike other joint-labeled topic models, ELDA was designed essentially as a feature reduction method.

The emotion-topic model (ETM),[11] the affective topic model (ATM),[12] the multilabel supervised topic model (MSTM), and the sentiment latent topic model (SLTM)[13] are joint emotion-topic models that introduced an intermediate layer into LDA. ETM is homologous to labeled LDA[7] in its generative process. Although labeled LDA is expressive enough to model multi-label corpora, it's inappropriate for directly detecting social emotions. This is because labeled LDA represents each label as a binary presence/absence indicator, which is insufficient to model numerous user ratings over each emotion label. ETM was proposed to utilize emotional distribution to guide topic generation by extending labeled LDA.

Despite the growing amount of work in this research area, adaptive social emotion classification remains a problem. Most works on social emotion detection, including ETM, ATM, MSTM, and SLTM, have focused mainly on associating emotions with topics specific to one context. The context sensitivity of topics in evoking social emotions, however, hasn't been well addressed by the previous models. We targeted the problem of adaptive social emotion classification and developed the contextual sentiment topic model to solve it.

References

1. B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," *Foundations and Trends in Information Retrieval*, vol. 2, 2008, pp. 8–10.
2. P.D. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," *Proc. 40th Annual Meeting Assoc. Computational Linguistics*, 2002, pp. 417–424.
3. Y. Zhao, B. Qin, and T. Liu, "Creating a Fine-Grained Corpus for Chinese Sentiment Analysis," *IEEE Intelligent Systems*, vol. 30, no. 1, 2015, pp. 36–43.
4. D. Bollegala, D. Weir, and J. Carroll, "Using Multiple Sources to Construct a Sentiment Sensitive Thesaurus for Cross-Domain Sentiment Classification," *Proc. 49th Annual Meeting Assoc. Computational Linguistics*, 2011, pp. 132–141.
5. S. Pan et al., "Cross-Domain Sentiment Classification via Spectral Feature Alignment," *Proc. Int'l Conf. World Wide Web*, 2010, pp. 751–760.
6. R. Xia et al., "Feature Ensemble Plus Sample Selection: Domain Adaptation for Sentiment Classification," *IEEE Intelligent Systems*, vol. 28, no. 3, 2013, pp. 10–18.
7. D. Ramage et al., "Labeled LDA: A Supervised Topic Model for Credit Attribution in Multi-label Corpora," *Proc. Conf. Empirical Methods in Natural Language Processing*, 2009, pp. 248–256.
8. R.W.P. Katz and M. Singleton, "SWAT-MP: The SemEval-2007 Systems for Task 5 and Task 14," *Proc. 4th Int'l Workshop Semantic Evaluations*, 2007, pp. 308–313.
9. Y.H. Rao et al., "Building Emotional Dictionary for Sentiment Analysis of Online News," *World Wide Web J.*, vol. 17, 2014, pp. 723–742.
10. D.M. Blei, A.Y. Ng, and M.I. Jordan, "Latent Dirichlet Allocation," *J. Machine Learning Research*, vol. 3, 2003, pp. 993–1022.
11. S.H. Bao et al., "Mining Social Emotions from Affective Text," *IEEE Trans. Knowledge and Data Eng.*, vol. 24, no. 9, 2012, pp. 1658–1670.
12. Y.H. Rao et al., "Affective Topic Model for Social Emotion Detection," *Neural Networks*, vol. 58, 2014, pp. 29–37.
13. Y.H. Rao et al., "Sentiment Topic Models for Social Emotion Mining," *Information Sciences*, vol. 266, 2014, pp. 90–100.

conditioned to topic z_j. The value of $p(w|\theta_B)$ is estimated using the combined document collection as

$$p(w \mid \theta_B) = \frac{\sum_{d \in D} c(w,d)}{\sum_{w' \in V} \sum_{d \in D} c(w',d)}, \quad (4)$$

where w' is one of the words occurring in D, and $c(w,d)$ is the occurrence count of word w in document d. The larger the count of word w occurring in D, the higher the value of $p(w|\theta_B)$.

Without loss of generality, we can identify and prune some very common, short function words that occur in many collections that don't help much in topic discrimination, such as "the," "is," "on," and "that" in English documents, and "人 (person)," "说 (speak)," "后 (after)," and "记者 (journalist)" in Chinese documents. These nondiscriminative

Table 1. Notations of frequently used variables.

Notations	Description
w_d	Words in document d
e_i	The ith emotion label
θ_B	The background theme
θ_C	The contextual theme
z_j	The jth context-independent topic

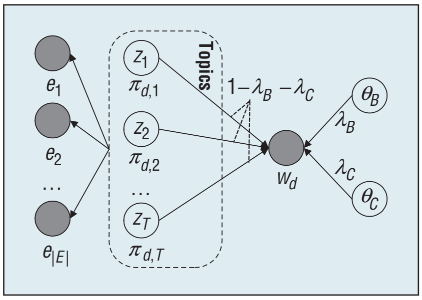

Figure 2. The contextual sentiment topic model (CSTM) framework. CSTM is designed for adaptive social emotion classification by distinguishing context-independent topics from both background and contextual themes.

words are mainly generated from the background theme, which characterizes common information across different collections, thus the corresponding values of $p(w|\theta_B)$ are high according to Equation 4.

We further estimate the value of $p(w|\theta_C)$ by

$$p(w \mid \theta_C) \propto \frac{\sum_{d \in D_{src}} c(w,d)}{\sum_{d' \in D} c(w,d')}, \qquad (5)$$

where d' is one of the documents contained in D, and $p(w|\theta_C)$ is summed to 1 for all words. The larger the count of word w occurring in D_{src} and the smaller the count of w occurring in D_{tar}, the higher the value of $p(w|\theta_C)$. The values of $p(w|\theta_C)$ range from 0 to 1, with the lowest being w only occurring in D_{tar} and the highest being w only occurring in D_{src}. The purpose of modeling the contextual theme is to "force" classification based on context-independent topics—for example, the topic of "Madonna's new baby" is specific to the context of entertainment and might not be helpful in classifying social emotions in other contexts. On the other hand, topics about "new life" are generalized and useful to social emotion classification across different contexts.

After separating generalized topics from background and contextual themes, we can classify social emotions of unlabeled documents in D_{tar} by these context-independent topics as follows:

$$p_d(e_i) = \sum_{w \in V} p(e_i \mid w) \sum_{j=1}^{T} (\pi_{d,j} p(w \mid z_j)), \qquad (6)$$

where $p(e_i|w)$ is the probability of emotion e_i conditioned to word w, which can be estimated by the existing method based on maximum likelihood estimation.[9]

Model Estimation

To compute the model's parameters, we can use the expectation maximization (EM) algorithm. According to the EM algorithm, the E-step is

$$p(z_j \mid w,d) = \frac{\pi_{d,j}^{(n)} p^{(n)}(w \mid z_j)}{\sum_{j'=1}^{T} \pi_{d,j'}^{(n)} p^{(n)}(w \mid z_{j'})},$$

$$p(\theta_B \mid w,d) = \frac{\lambda_B p(w \mid \theta_B)}{\lambda_B p(w \mid \theta_B) + \lambda_C p(w \mid \theta_C) + (1-\lambda_B-\lambda_C)\sum_{j=1}^{T}(\pi_{d,j} p(w \mid z_j))},$$

and

$$p(\theta_C \mid w,d) = \frac{\lambda_C p(w \mid \theta_C)}{\lambda_B p(w \mid \theta_B) + \lambda_C p(w \mid \theta_C) + (1-\lambda_B-\lambda_C)\sum_{j=1}^{T}(\pi_{d,j} p(w \mid z_j))}.$$

The M-step is

$$\pi_{d,j}^{(n+1)} = \frac{\sum_{w \in V} c(w,d)(1-p(\theta_B \mid w,d) - p(\theta_C \mid w,d))p(z_j \mid w,d)}{\sum_{j'=1}^{T}\sum_{w \in V} c(w,d)(1-p(\theta_B \mid w,d) - p(\theta_C \mid w,d))p(z_{j'} w,d)}$$

and

$$p^{(n+1)}(w \mid z_j) = \frac{\sum_{d \in D} c(w,d)(1-p(\theta_B \mid w,d) - p(\theta_C \mid w,d))p(z_j \mid w,d)}{\sum_{w' \in V}\sum_{d \in D} c(w',d)(1-p(\theta_B \mid w',d) - p(\theta_C \mid w',d))p(z_j \mid w',d)},$$

where

$$(1-p(\theta_B \mid w,d) - p(\theta_C \mid w,d))p(z_j \mid w,d) = \frac{(1-\lambda_B-\lambda_C)\pi_{d,j}^{(n)} p^{(n)}(w \mid z_j)}{\lambda_B p(w \mid \theta_B) + \lambda_C p(w \mid \theta_C) + (1-\lambda_B-\lambda_C)\sum_{j=1}^{T}(\pi_{d,j} p(w \mid z_j))}.$$

The two parameters in the M-step, $\pi_{d,j}$ and $p(w|z_j)$, are used for adaptive social emotion classification until convergence, as shown in Equation 6.

Experiments

To test the model's effectiveness, we collected 4,570 news articles and user ratings across eight emotions (touching, empathy, boredom, anger, amusement, sadness, surprise, and warmness) from Sina,[10] with the articles' published dates ranging from January to April 2012. To ensure that user ratings had stabilized, we gathered the data at least six months after the publication date. As a preprocessing step, we put this dataset through several measures.

First, we integrated the news headline and body of each news article into a single document. This process is appropriate for adaptive social emotion classification because both title and content can evoke a reader's emotions.

Second, we used a Chinese lexical analysis system (ICTCLAS; www.ictclas.org) to perform Chinese word segmentation. ICTCLAS is an integrated Chinese lexical analysis system based on a multilayer hidden Markov model. We analyzed a total of 1,975,153 generated word tokens and 325,434 user ratings jointly for adaptive social emotion classification.

Third, we removed those documents with no ratings and then normalized user ratings and summed them to 1 for each document—that is, we computed the percentage of users who voted for each emotion label *e* by dividing the number of ratings over *e* by the total number of ratings over all emotion labels. Because the publication dates vary, this treatment avoided the potential scenario of having too many user ratings for older news articles.

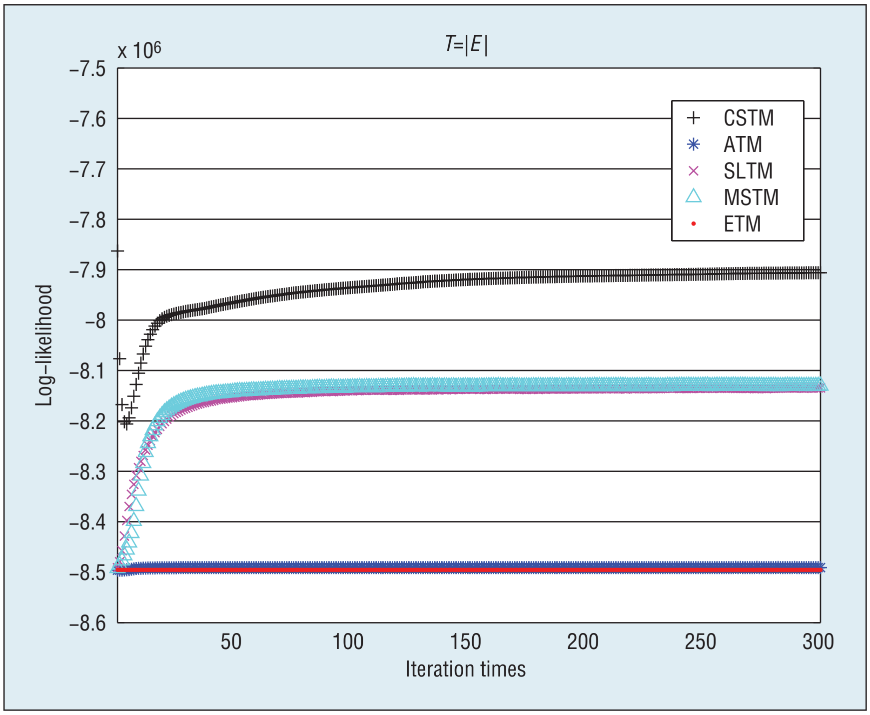

Figure 3. Log-likelihood of different models. The results show that the proposed CSTM stabilized within about 50 iterations and consistently outperformed the baseline models by a large margin.

Because adjacent news articles can have a similar context, we split the dataset into training and testing sets by putting documents into chronological order. We used a total of 2,342 documents published from January to February 2012, and 2,228 documents published from March to April 2012 for training and testing, respectively. We compared the existing supervised unigram model (SWAT),[11] emotion-term (ET) and emotion-topic model (ETM),[3] affective topic model (ATM),[10] multilabel supervised topic model (MSTM), and sentiment latent topic model (SLTM).[12] We set all the hyperparameters of ETM, ATM, MSTM, and SLTM at default and estimated the values of weighting parameters λ_B and λ_C in the CSTM using Equations 1 and 2, respectively.

We employed three evaluation metrics as indicators of performance: log-likelihood, *Acc*@1, and *AP*. Log-likelihood measures how well the model predicts unseen data. The higher the log-likelihood the model assigns to the testing set, the better its generalizability and predictive power. The coarse-grained metric *Acc*@1 stands for the accuracy at top 1, which we also used to evaluate the classification performance of ET and ETM.[3] According to *Acc*@1, a predicted ranked list of emotion labels is correct if the list's first item is identical to the actual ranked list's first item. If two emotion labels in the actual ranked list have the same number of votes, then their positions are interchangeable. *Acc*@1 is computed by dividing the number of correctly predicted documents by the total number of documents. The fine-grained metric *AP* (the averaged Pearson's correlation coefficient over all documents) is conducted because *Acc*@1 doesn't take emotional distributions into account. For each document, *AP* measures the correlation between the predicted probabilities and the truth votes over all emotion labels. The value of *AP* ranges from −1 to 1, where 1 indicates a perfect positive correlation.

Evaluation of Predictive Power

To evaluate the different models' generalizability, we set the number of topics to be the amount of emotion labels (|*E*|) and computed the log-likelihood each model assigns to new documents.

Figure 3 presents the log-likelihood curves of topic-level models with a growing number of iterations. The results show that the proposed CSTM stabilized within about 50 iterations and consistently outperformed the baseline models by a large margin. This verifies the effectiveness of distinguishing context-independent topics from background and contextual themes, leading to a more informative and discriminative model across different collections.

It's worth noting that the CSTM uses probabilistic modeling to generate topics, but the baselines of ATM, SLTM, MSTM, and ETM employ Latent Dirichlet Allocation (LDA). The Dirichlet hyperparameters α and β were introduced in models incorporating

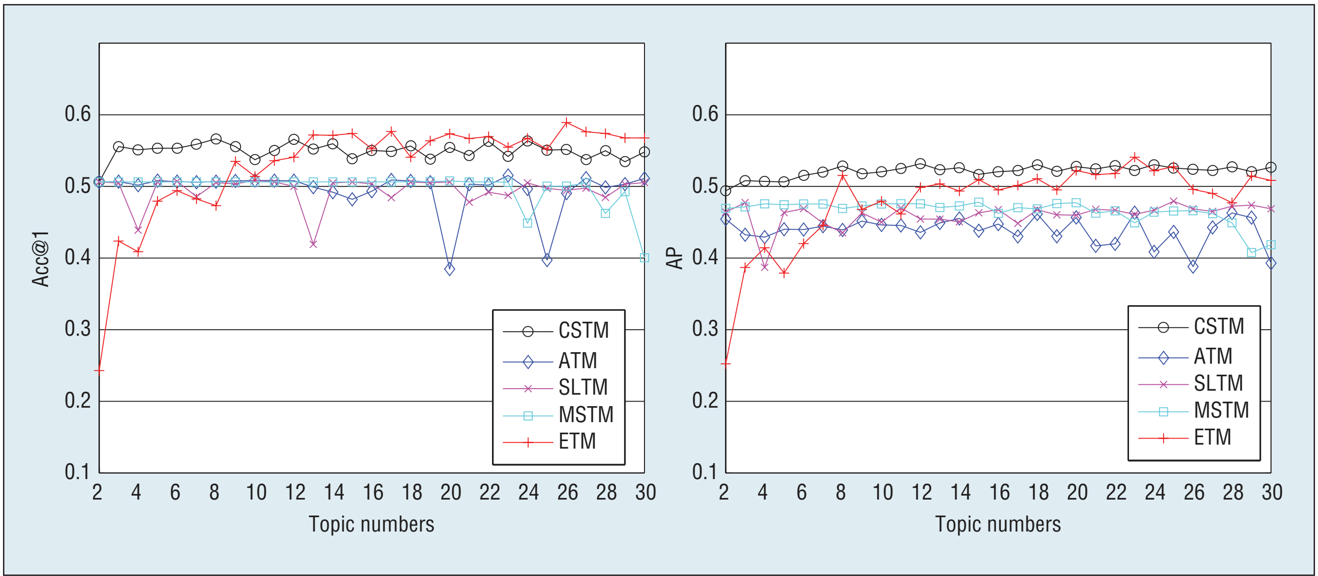

Figure 4. Performance with different topic numbers. Compared to the best-performing baseline of ETM, the proposed CSTM achieved stable performance when using different numbers of topics.

Table 2. Statistics for the number of topics according to different models.

Models	*Acc@1*		AP	
	Mean (%)	Variance	Mean	Variance
Contextual sentiment topic model (CSTM)	54.95	0.0001	0.52	0.0001
Affective topic model (ATM)	49.58	0.0009	0.44	0.0004
Sentiment latent topic model (SLTM)	49.50	0.0004	0.46	0.0003
Multilabel supervised topic model (MSTM)	49.83	0.0005	0.47	0.0003
Emotion-topic model (ETM)	52.90	0.0052	0.48	0.0036
Supervised unigram model (SWAT)	50.63	—	0.47	—
Emotion-term (ET)	42.91	—	0.43	—

LDA to control the densities of multinomial parameters. Although learning α and β from the data can be used to increase model quality, it's computationally intensive.[13] Different from those baselines, the probabilistic model CSTM included no Dirichlet hyperparameters. To make an appropriate comparison between models, all hyperparameters in the baselines incorporating LDA were set at default.

Influence of Number of Topics

The number of topics indicates how many latent aspects can be derived from the dataset, which can have an effect on ATM, SLTM, MSTM, ETM, and CSTM performance. Note that the performances of SWAT and ET didn't change with a different number of topics because they don't consider any latent topics. To evaluate the influence of the number of topics, we varied the number of topics from 2 to 30.[3] Figure 4 presents the performance of CSTM, ATM, SLTM, MSTM, and ETM when using different numbers of topics.

With respect to the number of topics, Table 2 shows the mean and variance of different models in terms of *Acc*@1 and *AP*. To conduct a comprehensive investigation into models at topic and word levels, we also included the performances of two word-level baselines (SWAT and ET). Compared to the best-performing baseline of ETM, the mean *Acc*@1 and *AP* of CSTM improved 3.9 percent and 9.0 percent, respectively.

To evaluate the differences in *Acc*@1 and *AP*, we performed two statistical tests on the CSTM and each baseline model. The first evaluated performance stability in terms of variances, and the second compared performance in terms of means. We used the conventional significance level (p value) of 0.05 (see Table 3).

First, we employed the analysis of variance F-test to test the assumption of homoscedasticity (the homogeneity of variance). Because SWAT and ET don't exploit latent topics, their performance is independent of topic number. The F-test was thus conducted on the CSTM and the baselines of ATM, SLTM, MSTM, and ETM (see Table 3). The results show that the differences in variances are statistically significant, with p values much less than 0.05. This suggests that the CSTM's performance is significantly more stable than that of ATM, SLTM, MSTM, and ETM when using a different number of topics.

Second, we conducted t-tests to test the assumption that the difference in performance between paired models had a mean value of zero. The results indicate that the proposed CSTM outperforms the baselines of ATM, SLTM, MSTM, SWAT, and ET significantly, with p values much less than 0.05. The difference in performance between the CSTM and the baseline ETM isn't statistically significant, with a p value equal to 0.07 in terms of *Acc*@1.

Although the baseline ETM also performed well on average in terms of accuracy, it had a strong preference for a larger number of topics (see Figure 4). This is because without removing nondiscriminative words, ETM can fit noise when using small numbers of topics. In terms of efficiency, large numbers of topics have higher time complexity.[9] The proposed CSTM used a background theme to alleviate this problem, thus achieving stable performance when using different numbers of topics. This type of profile is useful for modeling documents under both limited and diverse subjects, where small and large numbers of topics, respectively, are needed to best account for corpus structure.

Table 3. The p values of statistical tests on the CSTM and baselines.

Models	*Acc*@1		*AP*	
	F-test	*t*-test	F-test	*t*-test
ATM	3.4E-6	4.0E-11	2.7E-5	1.6E-23
SLTM	0.0037	1.2E-16	0.0002	1.0E-20
MSTM	0.0005	9.0E-14	0.0004	4.1E-20
ETM	1.7E-15	0.0701	0.0000	0.0003
SWAT	—	4.8E-18	—	8.7E-24
ET	—	3.9E-30	—	1.6E-31

In the future, we plan to apply the model to construct context-independent sentiment dictionaries and extend our research into the measurement of dynamic social emotions using long-period materials. With the rapid development of social media services, we also plan to extend our model by incorporating the context of users and their interactions. It follows that social emotion classification at the user level deserves further research. ■

Acknowledgments

This research was supported by the National Natural Science Foundation of China (no. 61502545), and the Fundamental Research Funds for the Central Universities (no. 46000-31610009).

THE AUTHOR

Yanghui Rao is an assistant professor in the School of Mobile Information Engineering at Sun Yat-sen University. His research interests include social emotion detection, natural language processing, and topic modeling. Rao received a PhD in computer science from the City University of Hong Kong. Contact him at raoyangh@mail.sysu.edu.cn.

References

1. E. Cambria et al., "New Avenues in Opinion Mining and Sentiment Analysis," *IEEE Intelligent Systems*, vol. 28, no. 2, 2013, pp. 15–21.
2. C. Bosco, V. Patti, and A. Bolioli, "Developing Corpora for Sentiment Analysis: The Case of Irony and Senti-TUT," *IEEE Intelligent Systems*, vol. 28, no. 2, 2013, pp. 55–63.
3. S.H. Bao et al., "Mining Social Emotions from Affective Text," *IEEE Trans. Knowledge and Data Eng.*, vol. 24, no. 9, 2012, pp. 1658–1670.
4. S. Pan et al., "Cross-Domain Sentiment Classification via Spectral Feature Alignment," *Proc. Int'l Conf. World Wide Web*, 2010, pp. 751–760.
5. R. Xia et al., "Feature Ensemble Plus Sample Selection: Domain Adaptation for Sentiment Classification," *IEEE Intelligent Systems*, vol. 28, no. 3, 2013, pp. 10–18.
6. D. Bollegala, D. Weir, and J. Carroll, "Using Multiple Sources to Construct a Sentiment Sensitive Thesaurus for Cross-Domain Sentiment Classification," *Proc. 49th Annual Meeting Assoc. Computational Linguistics*, 2011, pp. 132–141.
7. D. Ramage et al., "Labeled LDA: A Supervised Topic Model for Credit Attribution in Multi-label Corpora," *Proc. Conf. Empirical Methods in Natural Language Processing*, 2009, pp. 248–256.
8. B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," *Foundations and Trends in Information Retrieval*, vol. 2, 2008, pp. 25–26.
9. Y.H. Rao et al., "Building Emotional Dictionary for Sentiment Analysis of Online News," *World Wide Web J.*, vol. 17, 2014, pp. 723–742.
10. Y.H. Rao et al., "Affective Topic Model for Social Emotion Detection," *Neural Networks*, vol. 58, 2014, pp. 29–37.
11. R.W.P. Katz and M. Singleton, "SWAT-MP: The SemEval-2007 Systems for Task 5 and Task 14," *Proc. 4th Int'l Workshop Semantic Evaluations*, 2007, pp. 308–313.
12. Y.H. Rao et al., "Sentiment Topic Models for Social Emotion Mining," *Information Sciences*, vol. 266, 2014, pp. 90–100.
13. D.M. Blei, A.Y. Ng, and M.I. Jordan, "Latent Dirichlet Allocation," *J. Machine Learning Research*, vol. 3, 2003, pp. 993–1022.