Journal of Ambient Intelligence and Humanized Computing (2023) 14:13451–13463
https://doi.org/10.1007/s12652-022-03799-9

ORIGINAL RESEARCH

ABET: an affective emotion-topic method of biterms for emotion recognition from the short texts

Anima Pradhan[1] · Manas Ranjan Senapati[1] · Pradip Kumar Sahu[1]

Received: 23 August 2021 / Accepted: 7 March 2022 / Published online: 2 April 2022

Abstract
Nowadays, online users write short messages to share their feelings on social networking sites, such as discussion forums, question answering websites, etc., making these sites very popular. The increase in these short-term messages causes a huge data sparsity, making emotion recognition a challenging task. Therefore, a word co-occurrence pattern called biterms is generated from a large-scale dataset to prevent severe data sparsity issues. The topic modeling algorithms and acceleration algorithms are implemented to extract more reliable topics from the group of terms. Based on biterm technique, in this paper, a new algorithm called "Affected biterm emotion topic" is proposed for emotion recognition from a short text. For the experimental purpose, two popular short text datasets, SemEval and International Survey on Emotion Antecedents and Reactions (ISEAR), are used to investigate the performance of the proposed algorithm with the benchmark methods light latent dirichlet allocation (LLDA), biterm topic model (BTM), emotion-topic model (ETM), contextual sentiment topic model (CSTM), Sentiment latent topic model (SLTM) and siasme network based supervised topic model (SNSTM). The proposed algorithm is evaluated using the benchmark methods for mean, variance, and accuracy. The experimental result shows that the proposed algorithm is effective in analyzing emotions.

Keywords BTM · Topic model · Alias method · Metropolis–Hastings · Acceleration algorithm · Emotion classification · Short text analysis

1 Introduction

Nowadays, people are inclined towards social network websites such as Twitter, Facebook, etc. share their opinions or emotions. Though there is an unlimited range of message sizes, users are more comfortable expressing their comments on distinct topics such as politics, news, events, sports, etc., in a short length. A message consists of a few words to communicate. Therefore, it is different from other text information. However, due to its limited number of words in the message, words are usually once in each message. Classifying emotions such as joy, surprise, sadness, etc., is the most challenging task on short text because it contains low information compared to other text. Another big challenge on short text for researchers is feature sparsity. Because two different short texts may have different words, they may semantically correlate. Each word conveys multiple meanings based on its context (Erik et al. 2017).

Emotion detection using topic-level modeling is one of the solutions where each document is a mixture of topics. In the corpus, each word is semantically correlated to the other. Each topic consists of emotions of unlabelled documents. Bao et al. developed the emotion-topic model, generating more informative and coherent topics that come under different emotion labels (Bao et al. 2012). A labeled latent Dirichlet allocation model proposed defines a one-to-one mapping between the topics and emotion labels that neglect the latent topic features (Ramage et al. 2009). Other joint emotion topic models such as multi-label supervised topic model (MSTM), Sentiment latent topic model (SLTM) (Rao et al. 2014a), The affective topic model (ATM) (Rao et al. 2014b), proposed an intermediate layer in LDA. Words are extracted from both contextual and background themes in

Manas Ranjan Senapati
manassena@gmail.com

Anima Pradhan
animap2011_phdit@vssut.ac.in

Pradip Kumar Sahu
pksahu_it@vssut.ac.in

1 Department of Information Technology, Veer Surendra Sai University of Technology, Burla, India

the contextual sentiment topic model (CSTM) proposed by Rao (2016). Contextual theme indicates information on the contextual theme, whereas background theme indicates non-discriminative information. Ere classification of emotional labels is context-independent topics. These models beliefs in bag-of-words assumptions and do not follow word order in the sentence. To address this issue, another topic model is the Hidden Topic Markov Model (HTMM) (Gruber et al. 2007) approach, considering the structure of the sentence and order of the words to generate the topics. Though PLSA and LDA are the more popular and successful models in text mining, they cannot create proper topical knowledge over short text, leading to word co-occurrence sparsity. To handle data sparsity over short text, authors (Yan et al. 2013; Cheng et al. 2014) proposed a binary topic model if two co-occurred words belonging to the same topic learn more accurate topics from short messages in each document. However, BTM is simple and easy to implement but time-consuming to model on large datasets. Predicting emotions from the generated topic features of the BTM model is difficult without knowledge of labeled documents.

According to our knowledge, not much work has been done regarding this problem. Motivated by the concern mentioned above, we construct a weighted label topic model (WLTM) and affected biterm emotion topic (ABET) method to detect the emotions of the labeled document. This paper proposes two supervised topic models, weighted labeled topic model (WLTM) and Affective Biterm emotion-topic (ABET). WLTM model is the probability distribution of biterms of unlabelled documents. ABET, a multi-labeled topic model, is a probability distribution of topic of emotions to predict emotions distributions based on training data. BTM concept was adopted to solve the problem of data sparsity and learn more latent topics. Due to its high time complexity, inspired by highly efficient models LightLDA (Yuan et al. 2015) and AliasLDA (Li et al. 2014), Alia's method and Metropolis–Hasting (MH) algorithm have been applied to reduce the sampling complexity of Gibb's sampling algorithm.

The main contribution of our work is as follows:

1. A set of Biterm model of total corpus is generated, and Alias table Method Metropolis-Hasting (MH) algorithm are applied to reduce sampling complexity from O(k) to O (1).
2. The generative process of ABET is followed to generate a probability distribution for each label to multiple topics.
3. Probability distribution of each biterm of WLTM model created.
4. Radial Basis function employed to predict emotions on both the WLTM and ABET model.
5. Two public short text datasets ISEAR and SemEval were used to conduct experiment.

The organization of this paper is as follows. In Sect. 2, related work on the analysis of short text and emotion prediction on short text is presented. In Sect. 3, the detail of our research model WLTM and ABET is explained. The result is evaluated in Sect. 4 and the Conclusion of the paper is presented in Sect. 5.

2 Related article

Research on sentiment analysis and emotion in texts has attracted researchers due to its wide applications. The applications are mainly focused on stock prediction (Bollen et al. 2011), advertisement or product recommender systems (Bougie et al. 2003), marketing of a company based on consumers' emotions strategies (Mohammad and Yang 2013), etc. The fundamental method of emotion prediction is mainly divided into three categories: lexicon-based, supervised learning and unsupervised learning. The lexicon-based method creates dictionaries in concept-level, word-level, or emotional/sentiment level to detect emotions. It is domain-dependent training set is not required. Lexicon-based approaches depend on the dataset. Therefore, their accuracy depends on the availability of the word-emotion pairs in the respective lexicon.

Data sparsity problem is the main issue in a short text. Various solutions have been proposed to tackle it on short messages, including the prevalence of tweets, news headlines, Q&A websites, status messages, etc. Several methods proposed, aggregate the short text based on some dependent information to increase the length of a message before training model (Zhao et al. 2011; Weng et al. 2010). Hong and Davison (2010) proposed another solution: aggregate the tweet's message containing standard terms and train the author-topic and standard-topic models (Rosen-Zvi et al. 2004). In recent years, BTM achieved great success, combining the co-occurrence of word-pair in the document for the topic model (Yan et al. 2013; Cheng et al. 2014). For short text with V number of total words creates biterm, unorder co-occurrence of word-pair assuming each biterm share same topic in the document. Predicting the emotions of the labeled document via a supervised model is another challenging issue. Gibbs sampling was employed to efficiently estimate the parameters on many topics model (Yuan et al. 2015), but it consumes much time with increasing the size of the document or topic and combining the alias method and Metropolis-Hasting algorithms in parameter estimation to solve the issue.

The principle of the topic-based method described first (Bao et al. 2009, 2012) for emotion classification that the text of emotion correlated to the topic than the words. A popular topic modeling is known as latent Dirichlet allocation (LDA) (Blei et al. 2003), used in the task of text mining such as information retrieval. An intermediate layer was added into the LDA model to associate the similarity between emotion and topics.

Emotion-LDA (ELDA) (Rao et al. 2014c) was proposed to deal with social emotions at word-level and topic-level. However, ELDA is not a supervised model. A single word in a document has a different meaning. The probability of the social emotion is estimated by maximum likelihood estimation for each topic.

The topic-over-time (TOT) model proposed by Wang and Andrew (2006) and Blei and McAuliffe (2007) studied a supervised topic model (STM) are single-label topic modeling that is not suitable for the emotion classification task.. Two popular emotion-topic models, multi-label supervised topic model (MSTM) and sentiment latent topic model (SLTM) (Rao et al. 2014a), introduced LDA as an emotional layer to predict emotion. They collected 4570 news articles from Sina news to evaluate the experiment. Another multi-label topic model affective topic model (ATM) evaluated for emotion classification (Rao et al. 2014b). Siasme network-based supervised topic model (SNSTM) (Huang et al. 2018) model is developed by joining documents and emotion labels where the weight matrices are considered as a conditional distribution.

The short text documents are collected from various sources and the labels them manually from reader's emotion. A supervised topic model, universal affective model (UAM) (Liang et al. 2018) combines with word-emotion dictionary to solve the data sparsity problem. It consists of two sub models: word-level and topic-level. Another topic-level model combines the results of the unsupervised topic model into a Maximum Entropy classifier to solve the issue of data sparsity. The model's performance is evaluated on real-world dataset and the accuracy of the model is quite accurate.

To automatically recognize emotions from the different contexts, such as companions, locations, and activities of 32 participants, use machine learning techniques (Salido Ortega et al. 2020). Information regarding humans' emotions from other facial expressions (angry, happy, or sad), head movement (frequency and direction), eye gaze (averted or direct) extracted, and soft computing techniques are employed to get cognitional and emotional states (Zhao et al. 2013). Machine learning and sensor techniques are applied to recognize the emotions from the facial expression of people having issues with an autism spectrum disorder (Sivasangari et al. 2019).

Some classification algorithms such as support vector machine (Pang et al. 2002), Naive Bayes (Kim et al. 2006), maximum entropy (Li et al. 2016), deep memory network (Tang et al. 2016), H-Sentic LSTM, and Sentic LSTM (Ma et al. 2018), are supervised based learning methods used to detect sentiments and emotions from the text document. Detection of Emotion or Sentiment orientation based on unsupervised learning algorithms is by counting co-occurrence frequency between words. However, these methods are suitable for short text but not for standard text.

3 Affective biterm emotion-topic (ABET)

This section first proposed an effective supervised topic model for dealing with emotions over short texts. Our objective is to correctly model the links between emotion and words to help predict a document's sentiment. WLTM model and ABET have been proposed to achieve our goal. To make more efficient accelerated algorithms are being developed.

3.1 Problem definition

Summarization of terms, notations, variables to effectively represent our model is presented in Table 1. Let a corpus containing L number of labeled short texts $\{s_1, s_2, \ldots, s_L\}$ associated with words w_L and emotion e. For each labeled text s, Ls words denoted as ws = {w1, w2, …, wL} and Ne number emotions represented as $\{E_s =\}$ Es = {e1, e2, …., Ne}. Here Ls is the length of the document and V is total vocabulary size. We create a group of biterms for each document in corpus represented as $B = \{b_i\}_{j=1,} N_B$ with $b_i = \{w_{i,1}, wi_{,2}\}$. For example, a short text can be generated as follows, $\{w_1, w_2, w_3\} => \{(w_1, w_2), (w_2, w_3), (w_1, w_3)\}$. Common labels are joy, anger, sad, surprise, etc. A list of presence/absence of binary indicator $\Lambda(d) = \{w_1, w_{2,} \ldots w_v\}$ where each I ε (0,1).

Our objective is to model links between emotion and words correctly to help in improving the performance of predicting the sentiment of a document. The first weighted label topic model (WLTM) has been modeled to map each

Table 1 Notations

Symbol	Description
T	Number of topics
V	Number of unique words/vocabularies
N_e	Number of emotion labels
N	Number of texts
B	Number of biterms
ε_i	emotion labels
α	Dirichlet prior of θ and δ
β	Dirichlet prior of ϕ
θs	L × T, Multinomial distribution of document to topic
ϕ_t	V × T, Multinomial distribution of topic to words
φ	E × T, Multinomial distribution of topics to emotions

emotion of training text into topics to achieve our goal. Therefore, label-to-topic projection is many-to-many.

3.2 Biterm topic model (BTM)

Each document d of the corpus is a multinomial distribution over topics and each topic as a multinomial distribution over words described by LDA. The generative process of LDA is given as follows:

For whole biterm set B, draw topic distribution,

$$\theta_d \sim \text{Dirichlet}(\alpha), \tag{1}$$

where Dirichlet (α) is the distributions of Dirichlet parameter.

For each word, draw a word distribution,

$$\phi_t \sim \text{Dirichlet}\ (\beta), \tag{2}$$

where Dirichlet (β) is the distribution of Dirichlet parameter for topic t.

For each biterm, draw topic t from the multinomial distribution θ_d,

$$\text{T} \sim \text{Multinomial}(\theta). \tag{3}$$

For biterm b, draw two words from the multinomial distribution ϕ_t,

$$w_i, w_j \sim Multinomial(\phi). \tag{4}$$

Gibb's sampling is applied to infer topics because it is more efficient than maximum posterior estimation and variational inference. Conditional probability equation for biterm b_i is given by,

$$p\left(T|_{-b}, B, \alpha, \beta\right) \propto \frac{\left(n_{w_i|t} + \beta\right)\left(n_{w_j|t} + \beta\right)}{\left(\sum_{w=1}^{V} n_{w|t}^{-b} + V\beta\right)^2}, \tag{5}$$

where B denotes set of biterm group, $b = \left(w_i, w_j\right)$. T_{-b} is denoted as biterm assigned to topic t where biterm b is not included. n_t^{-b} is some biterms assigned to topic t where biterm b is not including. $n_{w_i}^{-b}$ and $n_{w_j}^{-b}$ are the number of words w_i assigned to topic t where biterm b is excluded respectively. V is the size of the vocabulary of corpus. Then, probability of each biterm b_i is computed with given parameters θ and ϕ is presented as,

$$P\left(\text{b}_\text{i}|\theta, \phi\right) = \sum_{\text{t}=1}^{\text{T}} \theta_\text{t}\phi_{\text{t},\text{w}_{\text{i},1}}\phi_{\text{t},\text{w}_{\text{i},2}}. \tag{6}$$

After that likelihood function of total biterms of the whole corpus is computed as follows:

$$P(B|\Theta, \Phi) = \prod_{i=1}^{B}\sum_{t=1}^{T}\theta_t\phi_{t,w_{i,1}}, \phi_{t,w_{i,2}}, \tag{7}$$

where T denotes the number of topics of the whole corpus. θ is T-dimensional multinomial distribution and Φ is T* V matrix. The probability of topic t is indicated by θ_t. ϕ_t is V-dimensional multinomial distribution and probability of word w with given condition t is denoted as $\phi_{t,w}$. Then with the given number of iterations, the occurrences of biterms assigned to topic t are recorded, indicated by n_t and the occurrences of word w assigned to topic t of vocabulary denoted as $n_{w|t}$. Then probabilities of topics over corpus θ and probabilities of words conditioned to ϕ are computed as follows:

$$\theta_t = \frac{n_t + \alpha}{B + T\alpha}, \tag{8}$$

$$\phi_{t,w} = \frac{n_{w|t} + \beta}{\sum_{w=1}^{V} n_{w|t} + V\beta}. \tag{9}$$

For each document d, each biterms generated through, the topic proportion is computed via Gibbs sampling algorithm used in WLTM (He et al. 2017). However, it is unable to model the documents directly in BTM, the topic proportion of a document $P(t|d)$ is derived from the posterior probability topic of biterms $b_i^{(d)} = (w_{i,1}^{(d)}, w_{i,2}^{(d)})$ assuming each topic t is conditionally independent of d can be calculated with the given equation:

$$P(t|d) = \sum_{i=1}^{N} P(t|b_i^{(d)})P(b_i^{(d)}|d). \tag{10}$$

$P(t|b_i^{(d)})$ is computed using Baye's formula as follows:

$$P\left(t|b_i^{(d)}\right) = \frac{\theta_t\phi_{t,w_{i,1}^{(d)}}\phi_{t,w_{i,2}^{(d)}}}{\sum_{t^{-i}}\theta_{t^{-i}}\phi_{t^{-i},w_{i,1}^{(d)}}\phi_{t^{-i},w_{i,2}^{(d)}}}. \tag{11}$$

3.3 Affective biterm emotion topic (ABET)

The generative process of ABET is shown as follows:

For emotion $e \in \left[1, N_E\right]$, draw

$$\delta_e \sim Dirichlet(\alpha); \tag{12}$$

For each topic $t \in \left[1, N_t\right]$, draw

$$\phi_t \sim Dirichlet(\beta); \tag{13}$$

For each document $d \in D, do.$
For each biterm $b_i \in d$, do.
Draw

$$e_i \sim Multinomial(\gamma_d); \tag{14}$$

Draw

$$t_i \sim \text{Multinomial}(\psi_i); \tag{15}$$

Draw

$$w_{i,1}, w_{i,2} \in b_i \sim Multinomial(\phi_{t_i}). \tag{16}$$

Here, emotion and topic for biterm are denoted as $e_i \in E$ and $t_i \in T$ respectively. For each document in the corpus, the value of emotion ε is normalized and summed to 1 is sampled as the parameter for multinomial distribution.

Based on the generative process, the joint probability of all the samples in each document is as follows:

$$\begin{aligned} &P(\gamma, \varepsilon, t, w, \psi, \varphi; \alpha, \beta) \\ &= P(\psi;\alpha)P(\phi, \beta)P(\gamma) \\ &\times P(\varepsilon|\gamma)P(t|\varepsilon, \psi)P(w|t, \phi). \end{aligned} \tag{17}$$

Posterior probability distribution on emotion conditioned to the topic for each biterm

$$P(\varepsilon_i = e|\gamma, \varepsilon_{-b}, t, w : \alpha, \beta) \propto \frac{\alpha + nt^{-b}_{e,t_i}}{|T|\alpha + \sum_t nt^{-b}_{\varepsilon_i,t}} \times \frac{\gamma_{d_i,e}}{\sum_{e'} \gamma_{d_{i,e'}}} \tag{18}$$

Sampling topic conditioned to the biterm is given as:

$$\begin{aligned} &P(t_i = t|t_{-b}, \gamma, \varepsilon, w;\alpha, \beta) \propto \frac{\alpha + nt^{-b}_{\varepsilon_i,t}}{|T|\alpha + \sum_t nt^{-b}_{\varepsilon_i,t}} \\ &\times \frac{\left(\beta + nw^{-b}_{t,w_i}\right)\left(\beta + nw^{-b}_{t,w_j}\right)}{|V|\beta + \sum_w nw^{-b}_{t,w}}. \end{aligned} \tag{19}$$

where t and ε are served as candidate topic and emotion respectively. w_i is current word extracted from document d. The number of times topic t is assigned to emotion e is denoted as,$nt_{\varepsilon,t}$ and the occurrence of word *w* is assigned to topic t is denoted as, $nw_{t,w}$. The *-b* of *nt* indicates topic assignment for all topic *t* where current biterm is not included and *-b* for *nw* indicates word *w* assignment for topic t where biterm *b* is not included.

Then the posterior probabilities of ψ, φ sampled from topics and emotions is estimated as follows:

$$\psi_{\varepsilon,t} = \frac{\alpha + n_{t|\varepsilon}}{T\alpha + \sum_t n_{t|\varepsilon}} \tag{20}$$

and

$$\varphi_{t,w} = \frac{\beta + n_{w|t}}{V\beta + \sum_t n_{w|t}} \tag{21}$$

With all given parameters, to predict the probability of word w given emotion ε, the latent topic t is integrated as follows:

$$P(w|\varepsilon) = \sum_t \psi_{\varepsilon,t}\varphi_{t,w}. \tag{22}$$

Finally, emotion distribution in each document d is estimated via Bayes theorem as:

$$\begin{aligned} P(\varepsilon|d) &= \frac{P(d|\varepsilon)P(\varepsilon)}{P(d)} \propto \mathrm{P}(\mathrm{d}|\varepsilon)\mathrm{P}(\varepsilon) \\ &= \mathrm{P}(\varepsilon)\prod_{w\in d} P(w|\varepsilon)^{\delta_{d,w}}. \end{aligned} \tag{23}$$

A brief procedure of ABET is shown in Algorithm 3.

3.3.1 Acceleration algorithms

The acceleration algorithm for WLTM and ABET has been employed to reduce complexity through Metropolis-Hasting sampling (Geweke and Tanizaki 2001) and Alias method (Walker 1977).

3.3.1.1 Alias method Generally, if anyone wants to sample n number of discrete distributions, it will take at least O(n) number of operations. The alias method gives an algorithm to extract n number of samples from sample distributions in O (1) operations if a discrete distribution is a uniform. Alias method creates an alias table by simulating uniform samples. For n times of sampling, it can finish the sampling in O (1) amortized time, though take O(n) operations for creating alias table. An example of probabilities table and alias table built using discrete probabilities distribution shown in Fig. 1. A detailed description of the method is given in Algorithm 1 and Algorithm 2.

3.3.1.2 Metropolis–Hastings In Gibbs sampling algorithm for WLTM and ABET, extracting topics in each iteration of ABET consumes much time when the total number of biterms is too large. The total cost to complete the above task is relatively high,

which is also a waste of storage space. For Gibbs sampling, if only alias table and probability table build up, total B × T size takes to save these tables for total biterms, B. Inspired by LightLDA (Yuan et al. 2015), the alias method and MH sampling method are employed together to estimate the parameters, as it is cheap.

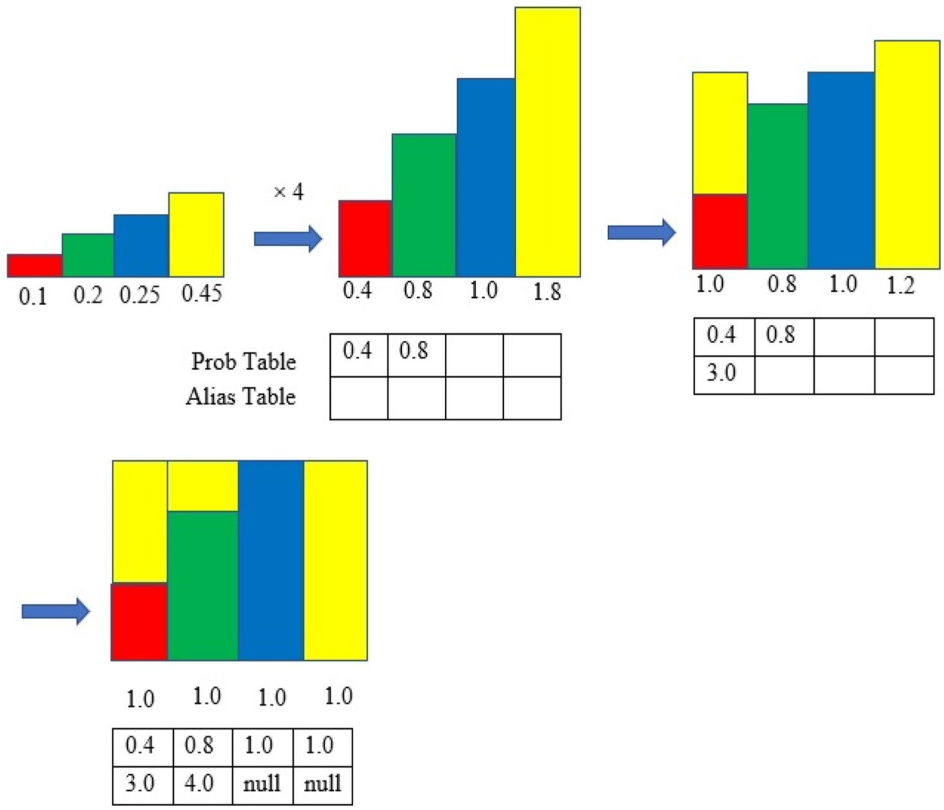

Fig. 1 An example of probability table and alias table

Algorithm 1: Generation of Alias Table

```
Input:   a set of discrete probabilities p1, p2, …, pn
Output:  Alias Table and Prob Table
Initialize two list Alias Table and Prob Table
Initialize two list LowList and HighList
Pi = n× pi, where i= 1 to n
for i=1 to n, do
if pi>1,
        Add i in HighList
else:
        Add i in LowList
end if
end for
while LowList and HighList not empty, do
l= LowList.pop (1) and h=HighList.pop (1)
ProbTable[l]= pl
AliasTable[h]=ph
ph = ph + pl -1
if ph<1 then,
        add l to LowList
else
        add h to HighList
end if
end while
while HighList not empty do
h=HighList.pop (1)
ProbTable[h]=1
end while
while LowList not empty do
h=LowList.pop (1)
ProbTable[l]=1
end while
```

Algorithm 2: Sample of Alias table

```
Input: AliasTable and Prob Table
Output: Sampling values
b=RandInt (n)
m=random (0,1)
if m< ProbTable[b] then
return b
else
return AliasTable [b]
end if
```

3.3.1.3 Parameter estimation

In WLTM, the conditional distribution of BTM decomposes into three parts:$\left(n_t+\alpha\right)$, $\frac{\left(n_{w_i|t}+\beta\right)}{\left(\sum_{w=1}^{V} n_{w|t}^{-b}+V\beta\right)}$ and $\frac{\left(n_{w_j|t}+\beta\right)}{\left(\sum_{w=1}^{V} n_{w|t}^{-b}+V\beta\right)}$.These parts are known as proposal distribution as per MH sampling. $\left(n_t+\alpha\right)$ is considered as corpus proposal $p_c(t)$ and $\frac{\left(n_{w_i|t}+\beta\right)}{\left(\sum_{w=1}^{V} n_{w|t}^{-b}+V\beta\right)}$ as word proposal$p_{w_i}(t)$.

Corpus proposal distribution is given as follows,

$$p_c(t) \propto \left(n_t+\alpha\right) \tag{24}$$

The acceptance probability is $\min(1, \pi_c)$, when topic t_1 translates to topic t_2, $\pi_c^{t_1\to t_2}$ is given as:

$$\pi_c = \frac{(n_{t_2}^{-b}+\alpha)\left(n_{w_i|t_2}^{-b}+\beta\right)}{(n_{t_1}^{-b}+\alpha)\left(n_{w_i|t_1}^{-b}+\beta\right)} \cdot \frac{(n_{w_j|t_2}^{-b}+\beta)}{\left(n_{w_j|t_1}^{-b}+\beta\right)} \cdot \frac{\left(\sum_w^V n_{w|t_1}^{-b}+V\beta\right)^2 (n_{t_1}+\alpha)}{\left(\sum_w^V n_{w|t_2}^{-b}+V\beta\right)^2 (n_{t_2}+\alpha)}. \quad (25)$$

In the corpus proposal, $p_c(t)$ is decomposed into two parts, n_t and α. Topics assigned for each biterm b_i, is stored as T_{b_i} which equals to the length of the number of biterms B of the corpus. First, topics T_{b_i} are randomly assigned for each i[th] biterm b_i from T_b, the current topic is considered a translation topic t_1.

$$\mathrm{p}(\mathrm{T_i}) = \frac{\sum_{\mathrm{i=1}}^{\mathrm{B}} \mathrm{T_{b_i}}}{\mathrm{B}} \quad (26)$$

Probability of topic for biterm b_i from T is given in Eq. (26). Here assigned topic of T_{b_i} is considered as uniform distribution of n_t. Therefore, sampling topic from T_{b_i} in O (1) time. Drawing topic from the second term is also in O (1) times due to the constant value of α for all biterms. T_{b_i}, α, both can draw in O (1) from corpus proposal without a built-up alias table. The values are randomly assigned in the range x = [0, B + T α]. If x is less than, x = int(x) is set else x = int(x-B).

3.3.1.4 Word proposal Word proposal distribution is given as,

$$\mathrm{p_{w_j}(t)} \propto \frac{(\mathrm{n_{w_j|t}}+\beta)}{\sum_{\mathrm{w=1}}^{\mathrm{V}} \mathrm{n_{w|t}}+\mathrm{V}\beta} \quad (27)$$

The acceptance probability is min(1, π_w), when a topic t_1 translates to topic t_2, $\pi_{w_j}^{t_1\to t_2}$ is given as:

$$\pi_{w_j} = \frac{(n_{w_i|t_2}^{-b}+\beta)\left(n_{w_j|t_2}^{-b}+\beta\right)(\sum_{w=1}^V n_{w|t_1}^{-b}+V\beta)^2}{(n_{w_i|t_1}^{-b}+\beta)(n_{w_j|t_1}^{-b}+\beta)(\sum_{w=1}^V n_{w|t_2}^{-b}+V\beta)^2} \times \frac{(n_{t_2}^{-b}+\alpha)(n_{w_j|t_1}^{-b}+\beta)(\sum_{w=1}^V n_{w|t_2}+V\beta)}{(n_{t_1}^{-b}+\alpha)(n_{w_j|t_2}^{-b}+\beta)(\sum_{w=1}^V n_{w|t_1}+V\beta)} \quad (28)$$

O(K) operations were sampled to extract topics from word proposal that is more costly than Gibbs sampling. Therefore, to minimize the cost to O (1) operations, an alias table is constructed (Li et al. 2014) and (Yuan et al. 2015) for computing p_{w_i}.

Parameter estimation for Algorithm 2, Eq. (19) is decomposed into three parts, $\frac{\alpha+nt_{\varepsilon_i,t}^{-b}}{|T|\alpha+\sum_t nt_{\varepsilon_i,t}^{-b}}$, $\frac{(\beta+nw_{t,w_i}^{-b})}{|V|\beta+\sum_w nw_{t,w}^{-b}}$ and $\frac{(\beta+nw_{t,w_j}^{-b})}{|V|\beta+\sum_w nw_{t,w}^{-b}}$. The first part is an emotional proposal, while the second and third parts are a word proposal.

The emotion proposal is given as follows:

$$p_{t|\varepsilon_i} \propto \frac{n_{t|\varepsilon_i}^{-b}}{|T|\alpha+\sum_t nt_{\varepsilon_i,t}^{-b}} \quad (29)$$

Acceptance probability is min$(1, \pi_{\varepsilon_i}^{t_1\to t_2})$, when a topic t_1 translates to topic t_2, $\pi_{\varepsilon_i}^{t_1\to t_2}$ is given as:

$$\pi_{\varepsilon_i} = \frac{(n_{t_2|\varepsilon_i}^{-b}+\alpha)\left(n_{t_1}+\alpha\right)(\sum_{w=1}^V n_{w|t_1}^{-b}+V\beta)^2}{(n_{t_1|\varepsilon_i}^{-b}+\alpha)(n_{t_2}+\alpha)(\sum_{w=1}^V n_{w|t_2}^{-b}+V\beta)^2} \times \frac{(n_{w_i|t_2}^{-b}+\beta)\left(n_{w_j|t_2}^{-b}+\beta\right)(|T|\alpha+\sum_t nt_{\varepsilon_i|t_2}^{-b})}{(n_{w_i|t_1}^{-b}+\beta)\left(n_{w_j|t_1}^{-b}+\beta\right)(|T|\alpha+\sum_t nt_{\varepsilon_i|t_1}^{-b})} \quad (30)$$

Word proposal distribution is given as,

$$p_{w_j}(t) \propto \frac{\left(n_{w_j|t}^{-b}+\beta\right)}{\sum_{w=1}^V n_{w|t}+V\beta} \quad (31)$$

Topic translates from t_1 to t_2, acceptance probability $\min\left(1, \pi_{w_j}^{t_1\to t_2}\right)$ is calculated as follows,

$$\frac{(n_{t_1|\varepsilon_i}^{-b}+\alpha)\left(n_{w_i|t_2}^{-b}+\beta\right)\left(n_{w_j|t_2}^{-b}+\beta\right)\left(\sum_{w=1}^V n_{w|t_1}+V\beta\right)}{(n_{t_2|\varepsilon_i}^{-b}+\alpha)(n_{w_i|t_1}^{-b}+\beta)\left(n_{w_j|t_1}^{-b}+\beta\right)\left(\sum_{w=1}^V n_{w|t_2}+V\beta\right)} \times \frac{(n_{w_j|t_1}^{-b}+\beta)(\sum_{w=1}^V n_{w|t_1}^{-b}+V\beta)^2}{(n_{w_j|t_2}^{-b}+\beta)(\sum_{w=1}^V n_{w|t_2}^{-b}+V\beta)^2}. \quad (32)$$

The MH-sampling method is applied to infer topics based on emotion ε_i, that depends on the emotion label of the dataset used in the dataset.

Algorithm 3: ABET

```
Input: emotional proposal, word proposal
Output: t, new topic
for i in MH step, do
    draw topic from emotional proposal
    compute the acceptance probability (p) of state transition t1 to t2
    m= random (0,1)
    if m<p, then
        t1=t2
    end if
    draw topic from word proposal w_i
    compute the acceptance probability (p) of state transition t1 to t2
    m=random (0,1)
    if m<p, then
        t1=t2
    end if
    draw topic from word proposal w_j
    compute the acceptance probability (p) of state transition t1 to t2
    m=random (0,1)
    if m<p, then
        t1=t2
    end if
end for
```

4 Experiment

Here, the result of our experiment on the proposed model has presented. The performance of the models is analyzed to achieve emotional prediction from the proposed model.

4.1 Dataset

ISEAR: This dataset consists of 7666 sentences, where every 1099 sentences belong to each emotion category. There are seven emotions: anger, fear, guilt, joy, disgust, sadness, and shame. 60% of the dataset was selected randomly for the training set, 20% for the validation set, and 20% for the testing set.

SemEval*:* There are 1246 news headlines in the dataset used in the 14th task of the 4th International Workshop on Semantic Evaluations (SemEval-2007). The training set consists of 1000 documents, and the testing set includes 246. Feelings of six basic emotions, joy, surprise, disgust, sadness, anger and fear, are contained in emotion labels (Katz et al. 2007).

During pre-processing, the stop-words, non-Latin characters are removed and converted into a lower case of each dataset document. For, ISEAR dataset, 1,571,829 biterms, and for SemEval, 5123 terms are created. Then, datasets are split into the training and testing sets and evaluated using fivefold cross-validation. Two methods, WLTM and ABET were implemented that incorporate accelerated algorithms. Six topic-level baseline methods modeling, LLDA (Ramage et al. 2009), BTM (Cheng et al. 2014), ETM (Bao et al. 2012), CSTM (Rao 2016), SLTM (Rao et al. 2014a), and SNSTM (Huang et al. 2018) have been implemented for comparison.

4.2 Experimental result

The performance of our model is evaluated using a fine-grained metric, the average Pearson's correlation coefficients (AP) (Rao 2016). AP is given as follows:

$$AP(m,n) = \frac{\sum_l \left(m(l) - m'\right)\left(n(l) - n'\right)}{\sqrt{\sum_l (m(l) - m')^2}\sqrt{\sum_l (n(l) - n')^2}} \tag{33}$$

where m, n, two vectors with an element l, m' and n' are mean of m and n, respectively. The range of AP, − 1 to 1, indicates more correlation coefficient with perfect prediction.

For emotion prediction, the Radial basis function (RBF) is applied on WLTM, BTM and LLDA. Five-fold cross-validation is performed on the training data for ISEAR and SemEval. The value of hyperparameters α and β is set to 0.1

and 0.01, respectively. For the SemEval dataset, the number of iterations was set to 500 on Gibbs sampling, whereas 200 iterations were run for ISEAR dataset as the average number of words is large. We run MH-sampling two times for WLTM and ABET algorithms to get a more effective acceptance rate.

To predict emotion by estimating the probability $P(\varepsilon|d)$, the emotion-term and topic-emotion model has been applied. The accuracy of the evaluation metric evaluates the performance of predicted emotion is given as:

$$Accuracy_d@N == \begin{cases} 1\, if\ \varepsilon_p \in E_{top}N@d \\ 0,\ n\,\text{even} \end{cases}. \tag{34}$$

Accuracy @ N for the testing set D is

$$Accuracy@N = \sum_{d \in D} \frac{Accuracy_d@N}{|D|}. \tag{35}$$

Emotion-term and emotion-topic models can be applied to emotion prediction by estimating the probability. In this section, their prediction performance has been evaluated. The parameter N is set to 1, 2, 3 (Erik et al. 2017).

4.2.1 Influenced of number of topics

In this part, we focus on selecting topic numbers that indicate the number of latent aspects that may affect the performance of our proposed model. Topics varied from 2 to 200 (a total 30 number of topics were tested) were used to evaluate the number of topics for Gibb's sampling. The performance of the WLTM method was evaluated with a different number of topics, which is measured by the log-likelihood function. Based on the log-likelihood function, the top 15 topics were considered for better accuracy for the WLTM algorithm. Figure 2a, b presented log-likelihood values of top topics over ISEAR and SemEval, respectively. Mean Accu@1, Accu@2 and Accu@3 metric is presented for different models to measure the performance of our proposed algorithm based on the topic numbers. The proposed model was compared with the popular six baseline methods LLDA (LLDA (Ramage et al. 2009), BTM (Cheng et al. 2014), ETM (Bao et al. 2012), CSTM(Rao 2016), SLTM (Rao et al. 2014a) and SNSTM (Huang et al. 2018).

4.2.2 Comparison with baselines

Experiments are conducted to analyze the mean and variance of the model in terms of AP. The top values of mean and variance of AP are reported in Table 2a, b in the boldface on ISEAR and SemEval dataset, respectively.

On the SemEval dataset, $AP_{document}$ and $AP_{emotion}$ performance is measured with baseline models such as LLDA, BTM, ETM, CSTM, SLTM, and SNSTM. Our proposed model WLTM outperformed in terms of $AP_{emotion}$ than other models. Compared to LLDA, BTM, ETM, CSTM, SLTM, SNSTM, the mean of $AP_{emotion}$ improves 0.0032, 0.0024, 0.0031, 0.0023, 0.0027 and 0.0031 respectively and ABET placed top 3 rank with the 0.1998 value. For the variance, WLTM placed rank top 3 and ABET in rank 4. In terms of $AP_{document}$, the performance of ABET gives better mean value as compared to LLDA, BTM, ETM, CSTM, SLTM, SNSTM that improves 0.3092, 0.1229, 0.0856, 0.0123, 0.1378, 0.0656 respectively and WLTM performed slightly worse for SemEval dataset. The possible reason is, 28 words do not appear in the 246 training documents, whereas available in 1000 testing documents. Due to missing samples in tuning parameters, SVR may underfit the emotion prediction in the document level of WLTM, BTM, and LLDA. According to variance values, WLTM achieves top rank 3 and ABET in top 4. Hence, WLTM reliable model for SemEval dataset.

Performance of experimental results over ISEAR dataset indicate that WLTM outperformed on both $AP_{document}$ and $AP_{emotion}$. WLTM outperformed baselines LLDA, BTM, ETM, CSTM, SLTM, SNSTM improves 0.4159, 0.0974, 0.0831, 0.2190,0.3344, 0.0796 on $AP_{emotion}$ respectively. The variance result of the WLTM is placed in rank top 2 on $AP_{document}$ and gives better performance on $AP_{emotion}$ with variance value 9.31E−05 which is more stable than baselines. According to the experimental result, ABET model yields competitive performance on both $AP_{document}$ and $AP_{emotion}$ with the value 0.2978 and 0.3427, respectively. The variance result is in top 4 and top 3 in $AP_{document}$ and $AP_{emotion}$ respectively. Although the experimental result of ABET in Pearson Correlation Coefficient cannot achieve best results on both SemEval and ISEAR dataset, still indicates significantly stable model. Based on the experimental result of WLTM on ISEAR yield better performance on both $AP_{document}$ and $AP_{emotion}$ that prove WLTM is more efficient than ABET and baselines.

For the metrics Accuracy@1, Accuracy@2 and Accuracy@3, on both ISEAR and SemEval datasets presented in shown in Table 3a, b respectively. On SemEval dataset, WLTM outperformed other models that improves 32.66%, 7.86%, 11.02%, 5.61%, 16.25%, 4.06 in Accuracy@3 metric with baselines LLDA, BTM, ETM, CSTM, SLTM, SNSTM respectively. Our proposed ABET model shows competitive performance which is in top2 rank according to Accuracy@3 metric result. Compared to baselines LLDA, BTM, ETM, CSTM, SLTM, SNSTM improves 30.87%, 6.07%, 9.23%, 3.82%, 14.46%, 2.27%, respectively. On the ISEAR dataset, the performance of WLTM model is quite better, that is placed in the rank top 4 and ABET is in the top 2 ranks in Accuracy@3. The performance of ETM is better than both WLTM and ABET on the ISEAR dataset, as topic sampling in ETM is constrained by one label. So, ETM can be mapping most of the samples to their actual emotion label.

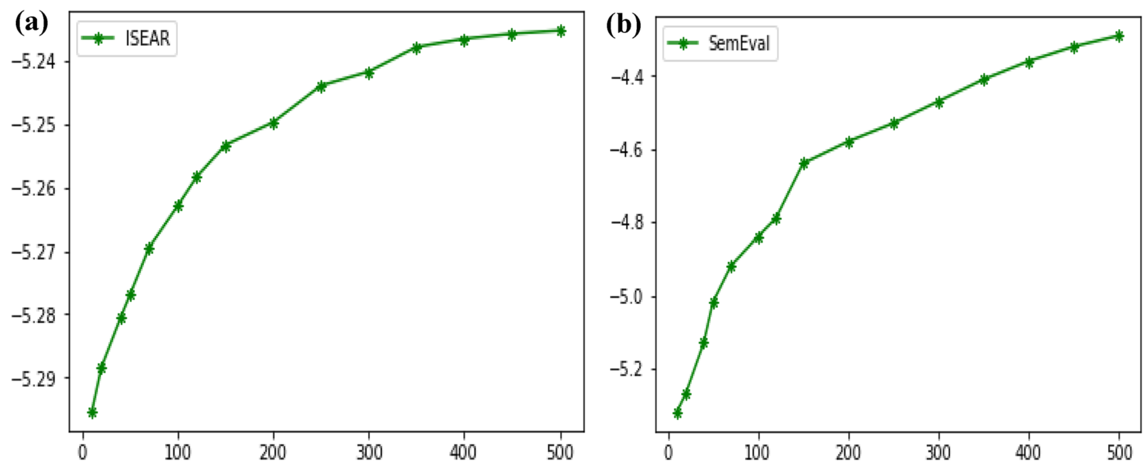

Fig. 2 Log-likelihood values over the top 15 topics for **a** ISEAR and **b** SemEval dataset

The T test is conducted to compare the performance of a paired model of WLTM and ABET model. The conventional significant level i.e., p value = 0.05. The result based on t test, WLTM outperformed the ABET, BTM, and LLDA and randomly significantly with a p value equal to 5.32E−8, 3.31E−11 respectively. The performance of between ABET is not statistically significant with a p value equal to 0.2917.

Table 2 Performance of AP on (a) ISEAR and (b) SemEval dataset

Models	$AP_{document}$		$AP_{emotion}$	
	Mean	Variance	Mean	Variance
(a) ISEAR				
WLTM	**0.4301**	**4.12E-05**	**0.4512**	**9.31E-05**
ABET	0.2978	1.93E-05	0.3427	**0.0001**
LLDA	0.0142	**3.41E-05**	0.0239	**7.90E-05**
BTM	0.3327	0.0014	0.3590	0.0015
ETM	0.3470	**6.87E-05**	0.4149	0.0002
CSTM	0.2111	0.0004	0.2269	0.0007
SLTM	0.0957	0.0012	0.0896	0.0010
SNSTM	0.3323	0.0001	0.3716	0.0002
(b) SemEval				
WLTM	0.1974	**0.0006**	**0.2416**	**0.0002**
ABET	**0.3124**	**0.0007**	0.1998	0.0004
LLDA	0.0032	0.0032	0.0082	**6.77E-05**
BTM	0.1895	0.0011	0.2262	0.0008
ETM	0.2268	0.0009	0.0666	**0.0001**
CSTM	0.3001	**0.0001**	0.1114	0.0009
SLTM	0.1746	0.0044	0.0205	0.0005
SNSTM	0.2468	0.0011	0.1495	0.0001

The top values of mean and variance of AP in are shown in boldface on ISEAR and SemEval datasets respectively

Figure 3a represents the accuracy results for ISEAR dataset. As shown in the figure for accuracy1, ABET model achieves an accuracy of 39.48%, whereas it is 5.43% for BTM model and 23.86% for LLDA model. Similarly, the accuracy2 for ABET model is 58.27%, for BTM model is 0.92% and LLDA model is 29.02% and the accuracy3 for ABET model is 69.78%, for BTM model it is 1.75% and LLDA model is 22.77%. Figure 3b shows the comparison of accuracy results of SemEval dataset in terms of percentage on different models. For accuracy1, the ABET model achieves the highest accuracy of 36.12% compared to 3.04% for the BTM model and 19.78% for the LLDA model. Similarly, the accuracy2 for ABET model is 57.02%, for BTM model is 4.61% and LLDA model is 24.98% and the accuracy3 for ABET model is 76.07%, for BTM model it is 4.09% and LLDA model is 30.79%.

4.2.3 Samples of the emotion Lexicons

After analyzing deeply, the topics generated in the WLTM model, the top terms of the topics assigned strong emotion labels are shown in Table 4. It shows that most of the terms in topic 6, topic 13, topic 21, topic 28 of the corpus with high probability values are strongly correlated to emotions like Fear, Sadness, Anger, guilt. After a close analysis, the emotion topic model can identify successfully that Topic 6 related to death news-related terms topic 21 relates to crime activity terms. Topic 13 and topic 28 share the same term "fail" associated with different emotion labels such as sadness and guilt. Though topic 6 belongs to fear associated with sadness, able to identify by the emotion-topic model. Topic 13 and topic 28 contain the similar term "fail" associated with both Sadness and Guilt emotion. After study thoroughly the documents, it happens to depend on the types of documents. For example, the sentence "I failed to complete

Table 3 Experimental result of accuracy in terms of percentage over (a) ISEAR and (b) SemEval on different models

Models	Accuracy@1	Accuracy@2	Accuracy@3
(a) ISEAR			
WLTM	40.76%	57.04%	68.10%
ABET	39.48%	58.27%	72.67%
LLDA (Ramage et al. 2009)	14.29%	28.57%	42.86%
BTM (Cheng et al. 2014)	32.35%	55.77%	70.65%
ETM (Bao et al. 2012)	**48.50%**	**66.22%**	**77.86%**
CSTM (Rao 2016)	28.98%	46.61%	61.76%
SLTM (Rao et al. 2014a)	20.40%	36.48%	51.12%
SNSTM (Huang et al. 2018)	45.40%	61.19%	72.21%
(b) SemEval			
WLTM	36.61%	**59.31%**	**77.80%**
ABET	36.12%	57.02%	76.07%
LLDA (Ramage et al. 2009)	20.20%	26.80%	45.20%
BTM (Cheng et al. 2014)	31.60%	53.30%	70.00%
ETM (Bao et al. 2012)	25.41%	50.19%	66.84%
CSTM (Rao 2016)	29.77%	53.84%	72.25%
SLTM (Rao et al. 2014a)	20.84%	42.53%	61.61%
SNSTM (Huang et al. 2018)	**38.90%**	56.30%	73.80%

The bold values show the percentage values of Accuracy@1, Accuracy@2 and Accuracy@3, on both ISEAR and SemEval datasets respectively

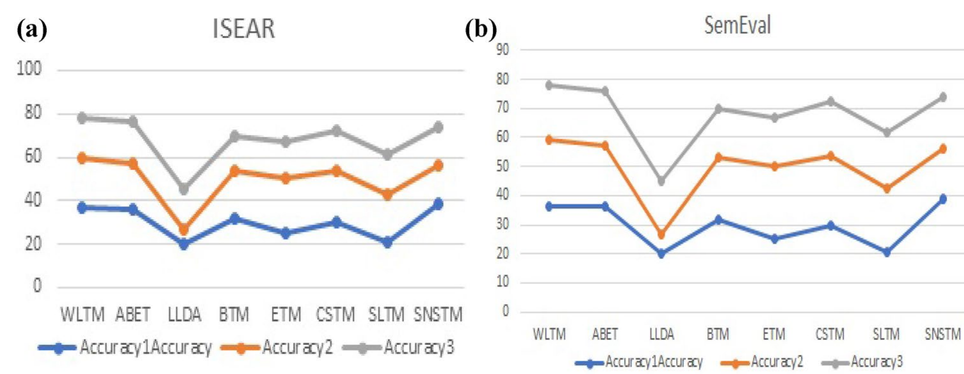

Fig. 3 Comparison of accuracy result in terms of percentage over **a** ISEAR and **b** SemEval on different models

Table 4 Emotion Lexicon samples from WLTM over *ISEAR*

Topic	Top 5 representative words	Emotion	Probability
6	Death, exam, fall, knife, traffic	Fear	0.9878
13	Fail, cancer, funeral, leave, depress	Sadness	0.9684
21	Drink, stole, complain, lost, forced	Anger	0.9262
28	Hate, lie, damaged, shouted, drunk	guilt	0.8956

a working task with the greed time", and "I fail in the exam." The first sentence expresses guilt emotion whereas the second sentence belongs to sadness. We can see that the same term may express different emotions based on topics identified by our model.

The probability distribution of seven different emotions for each topic is generated. Some samples have shown in Table 5. It indicates that each topic connected heavily to one emotion label with probability rate. We can see that Topic 7 as probability 97.77% related to the emotion label "Fear" and Topic 16 related to the emotion "joy".

5 Conclusion

Predicting emotions from short text is a challenging task in text mining. Here, two algorithms, WLTM and ABET, are presented to set the connection between topics and emotions. Our algorithms can also handle the feature sparsity issue in detecting emotion over short messages. Alias method and MH- algorithms, two accelerated methods, are proposed to

Table 5 Emotion Lexicon samples from ABET over ISEAR

Topic	Top 5 representative words	Joy	Anger	Fear	Sadness	Disgust	Shame	Guilt
7	Traffic-signal, road, accident, car, drive	3.4E−05	0.02	**0.9777**	0.02	0.0011	3.4E−05	0.0011
16	Friend, child, party, mother, cake	**0.9578**	0.1123	0.0125	0.0015	0.0015	0.0125	0.1123
35	Drink, steal, quarrel, broke, insult	0.1972	0.3213	0.1972	0.0036	0.0036	**0.87**	0.3213
41	Hate, lie, damaged, shouted, drunk	0.2754	0.5217	0.5217	**0.2781**	0.78	0.2754	0.2781

The top probability distribution values for each topic are shown in boldface on ISEAR datasets

reduce complexity in parameter estimation. An experiment has been conducted to evaluate the effectiveness of the proposed methods. After being compared with baseline methods, the experimental result indicated that the performance of our approach was competitive.

Acknowledgements The authors acknowledge the support given by the faculties and staff of the Department of Information Technology, VSSUT, Burla

Author contributions All the authors have equally contributed to preparing the manuscript.

Funding Not applicable.

Availability of data and material Not applicable.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

Bao SH, Xu SL, Zhang L, Yan R, Su Z, Han DY, Yu Y (2012) Mining social emotions from affective text. IEEE Trans Knowl Data Eng 24(9):1658–1670

Bao SH, Xu SL, Zhang L, Yan R, Su Z, Han DY, Yu Y (2009) Joint emotion topic modeling for social affective text mining. In: Proceedings of the 9th IEEE international conference on data mining (IDM), Miami, pp 699–704

Blei D, Mcauliffe J (2007) Supervised topic models. In: Advances in neural information processing systems, vol 20

Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. J Mach Learn Res 3:993–1022

Bollen J, Mao H, Zeng X (2011) Twitter mood predicts the stock market. J Comput Sci 2(1):1–8

Bougie R, Pieters R, Zeelenberg M (2003) Angry customers don't come back, they get back: the experience and behavioral implications of anger and dissatisfaction in services. J Acad Mark Sci 31(4):377–393

Cheng X, Yan X, Lan Y, Guo J (2014) Btm: topic modeling over short texts. IEEE Trans Knowl Data Eng 26(12):2928–2941

Erik C, Soujanya P, Alexander G, Mike T (2017) Sentiment analysis is a big suitcase. IEEE Intell Syst 32(6):74–80

Geweke J, Tanizaki H (2001) Bayesian estimation of state-space model using the metropolis-hastings algorithm within gibbs sampling. Comput Stat Data Anal 37(2):151–170

Gruber A, Weiss Y, Rosen-Zvi M (2007) Hidden topic markov models. In: Proceedings of the conference on Artificial intelligence and statistics (PMLR), pp 163–170

He X, Xu H, Li J, He L, Yu L (2017) FastBTM: Reducing the sampling time for biterm topic model. Knowl Based Syst 132:11–20

Hong L, Davison B (2010) Empirical study of topic modelling in Twitter. In: Proceedings of the first workshop on social media analytics. Association for Computing Machinery, New York, pp 80–88

Huang M, Rao Y, Liu Y, Xie H, Wang FL (2018) Siamese network-based supervised topic modeling. In: Proceedings of the 2018 conference on empirical methods in natural language processing (EMNLP). Association for Computational Linguistics, Brussels, pp 4652–4662

Katz P, Singleton M, Wicentowski RH (2007) SWAT-MP: the SemeVal-2007 systems for task 5 and task 14. In: Proceeding 4th international workshop semantic evalaluation. Association for Computational Linguistics, Prague, pp 308–313

Kim S-B, Han K-S, Rim H-C, Myaeng SH (2006) Some effective techniques for naive Bayes text classification. IEEE Trans Knowl Data Eng 18(11):1457–1466

Li AQ, Ahmed A, Ravi S, Smola AJ (2014) Reducing the sampling complexity of topic models. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. Association for Computing Machinery, New York, pp 891–900

Li J, Rao Y, Jin F, Chen H, Xiang X (2016) Multi-label maximum entropy model for social emotion classification over short text. Neurocomputing 210:247–256

Liang W, Xie H, Rao Y, Lau RY, Wang FL (2018) Universal affective model for Readers' emotion classification over short texts. Exp Syst Appl 114:322–333

Ma Y, Pen H, Khan T, Cambria E, Hussain A (2018) Sentic LSTM: a hybrid network for targeted aspect-based sentiment analysis. Cogn Comput 10(4):639–650

Mohammad SM, Yang T W (2013) Tracking sentiment in mail: How genders differ on emotional axes. In: Proceedings of the 2nd workshop on computational approaches to subjectivity and sentiment analysis (WASSA 2.011). Association for Computational Linguistics, Portland, pp 70–79

Pang B, Lee L, Vaithyanathan S (2002) Thumbs up? Sentiment classification using machine learning techniques. In: Proceeding ACL conference empirical methods natural language processing (EMNLP), pp 79–86

Ramage D, Hall D, Nallapati R, Manning C D (2009) Labeled LDA: a supervised topic model for credit attribution in multi-labeled corpora, In: Proceedings of the 2009 conference on empirical methods in natural language processing (EMNLP). Association for Computational Linuistics, Singapore, pp 248–256

Rao Y (2016) Contextual sentiment topic model for adaptive social emotion classification. IEEE Intell Syst 31(1):41–47

Rao Y, Li Q, Mao XD, Wenyin L (2014a) Sentiment topic models for social emotion mining. Inform Sci 266:90–100

Rao Y, Li Q, Wenyin L, Wu QY, Quan XJ (2014b) Affective topic model for social emotion detection. Neural Netw 58:29–37

Rao YH, Lei JS, Wenvin L, Li Q, Chen ML (2014c) Building emotional dictionary for sentiment analysis of online news. World Wide Web J 17:723–742

Rosen-Zvi M, Griffiths T, Steyvers M, Smyth P (2004) The author-topic model for authors and documents. In: Proceedings of the 20th conference on uncertainty in artificial intelligence. AUAI Press, Arlington, Virginia, pp 487–494

Salido Ortega MG, Rodríguez LF, Gutierrez-Garcia JO (2020) Towards emotion recognition from contextual information using machine learning. J Ambient Intell Human Comput 11(8):3187–3207

Sivasangari A, Ajitha P, Rajkumar I, Poonguzhali S (2019) Emotion recognition system for autism disordered people. J Ambient Intell Human Comput. https://doi.org/10.1007/s12652-019-01492-y

Tang D, Qin B, Liu T (2016) Aspect level sentiment classification with deep memory network. In: Proceeding of conference empirical methods natural language processing (EMNLP). Association for Computational Linguistics, Austin, pp 214–224

Walker AJ (1977) An efficient method for generating discrete random variables with general distributions. Trans Math Soft 3(3):253–256

Wang X, Andrew M (2006) Topics over time: a non-markov continuous-time model of topical trends. In: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining. Association for Computing Machinery, New York, pp 424–433

Weng J, Lim E, Jiang J, He Q (2010) Twitterrank: finding topic-sensitive influential twitterers, In: Proceedings of the third ACM international conference on web search and data mining (WSDM), New York, pp 261–270

Yan X, Guo J, Lan Y, Cheng X (2013) A biterm topic model for short texts. In: Proceedings of the 22nd international conference on World Wide Web. ACM, pp 1445–1456

Yuan J, Gao F, Ho Q, Dai W, Wei J, Zheng X, Xing EP, Liu TY, Ma WY (2015) Lightlda: Big topic models on modest computer clusters. In: Proceedings of the 24th international conference on World Wide Web, Switzerland, pp 1351–1361

Zhao W, Jiang J, Weng J, He J, Lim E, Yan H, Li X (2011) Comparing Twitter and traditional media using topic models. European conference on information retrieval. Springer, Berlin, pp 338–349

Zhao Y, Wang X, Goubran M, Whalen T, Petriu EM (2013) Human emotion and cognition recognition from body language of the head using soft computing techniques. J Ambient Intell Human Comput 4(1):121–140