Emotion Recognition System for Affective Text

K. Thendral, S. Chitrakala
1,2Department of CSE, Anna University, Chennai
Email: 1thendral.khce@gmail.com, 2ckgops@gmail.com

Abstract: Mining social emotions deals with a new aspect for categorizing the document based on the emotions such as happy, sad, sympathy, etc. They help online users to select related documents based on their emotional preferences. Aiming to find the connections between social emotions and affect terms and thus predict the social feeling from text content mechanically, Joint emotion-topic model by augmenting Latent Dirichlet Allocation with an additional layer for emotion modelling initially generates a group of latent topics from emotions, followed by generating affect terms from every topic. The techniques involved in this are emotion term model, topic model and emotion topic model. The emotion-topic model utilizes the complementary advantages of both emotion term model and topic model, and is not only effective in extracting the meaningful latent topics, but also improves the performance of social emotion prediction.

Keywords – Affective Text, Emotion-topic model, Latent Dirichlet Allocation.

I. INTRODUCTION:

An emotion is a Meta communicative pictorial representation of a facial expression which in the absence of body language and probably draw to serve a receiver’s attention to the tenor or temper of a sender’s nominal verbal communication, changing and improving its interpretation. It expresses-usually by means of punctuation marks – a person’s feelings or mood and can include numbers and letters.

The interrelation of text and emotions has been a captivating topic for centuries. What makes people feel what they read? How is the writer’s emotion conveyed in a text? How can we write to communicate an emotional message more clearly? A number of researchers have attempted to obtain answers to these questions for a long time and there is an enormous amount of literature on techniques and devices for emotion detection. (Bloom, Garg, & Argamon, 2007.)

Two attempts to measure emotions are based on two different models: dimensional and categorical. In the categorical model emotions are labelled, say that a person is “happy” or “sad” and people get a sense of what others mean. In the dimensional model the representation is using multidimensional scaling (e.g. “pleasant-unpleasant”, “excitement”, and “yielding-resisting”).

In the affective computing domain, supervised learning techniques are preferred due to strong performance. However, a challenge to using supervised techniques is the need for corpora with text that has been annotated with emotion labels. These are time consuming and expensive to produce. Unsupervised techniques do not have these requirements but are often less precise.

II. RELATED WORK

Many methods have been proposed to mine emotions from the text and social networks. Affective text mining deals with mining emotions from affective words. SemEval introduced a task named “affective text” in 2007 [2], aiming to annotate short headline texts with a predefined list of emotions and/or polarity orientation (positive/negative). There is a large body of previous work on mining affective content from text documents, product reputation mining [10], customer opinion extraction/summarization[11], [12], and sentiment classification [13]. However, none of these studies explores the connection between social emotions and affective terms.

An online system Mood Views has also been developed for tracking and searching emotion annotated blog posts [12], [13], [14], [15]. The posts are published with an indicator of the “current mood” of the blogger, at the time of posting the blog. Mood-Views is a platform for collecting, analyzing, and displaying aggregate moods in the blog space. Launched in mid-2005, Mood Views continuously collects these emotion indications, as well as the blog posts themselves, and provides a number of services. Despite the success of previous work on emotion prediction, existing approaches usually model documents under the “bag-of-word” assumption, so that the relationship across words is not taken into account. This also prevents us from further understanding the connections between emotions and contents in the topic level, because it is arguable that emotions should be linked to specific document topics.

D.M. Blei, A.Y. Ng, and M.I. Jordan [8] proposed Latent Dirichlet Allocation generative probabilistic model for collections of discrete data such as text corpora. LDA is a three-level hierarchical Bayesian model, in which each item of a collection is modelled as a finite mixture over an underlying set of topics. Each
topic is, in turn, modelled as an infinite mixture over an underlying set of topic probabilities. In the context of text modelling, the topic probabilities provide an explicit representation of a document.

Joint latent topic model for text and citations [8]. The Pairwise-Link-LDA model combines the ideas of LDA [4] and Mixed Membership Block Stochastic Models [1] and allows modelling arbitrary link structure. However, the model is computationally expensive, since it involves modelling the presence or absence of a citation (link) between every pair of documents. The second model solves this problem by assuming that the link structure is a bipartite graph. As the name indicates, Link-PLSA-LDA model combines the LDA and PLSA models into a single graphical model.

I. Titov and R. McDonald [8] proposed statistical model which is able to discover corresponding topics in text and extract textual evidence from reviews supporting each of these aspect ratings – a fundamental problem in aspect-based sentiment summarization. Achieve high accuracy, without any labelled data except the user opinion ratings.

Rosen-Zvi et al. [3] merged author factors with document generation to jointly estimate document contents as well as author interests. From the perspective of model generation, their author variable shares some similarity with the emotion variable in this model. The key difference lies in different sampling distributions. Their author variable is chosen uniformly from a set of authors while emotion variable is sampled from multinomial distributions by the emotions contributed by web users.

III. PROPOSED SYSTEM

An online text collection D is associated with a vocabulary W, and a set of predefined emotions E. Comparing the extracted and optimized content with the already founded latent topics that relating to the extracted and optimized content with the already founded latent topics that relating to each emotion. Based on the result we are finding which emotion the particular content represents. Based on the user emotion requests the categorized content will display.

Objective is to accurately model the connections between words and emotions, and improve the performance of its related tasks such as emotion prediction. Both the emotion–term model and emotion-topic model can be applied to emotion prediction by estimating their probability to evaluate their prediction performance. In this paper, proposing a joint emotion-topic model for social affective text mining, which introduces an additional layer of emotion modelling into Latent Dirichlet Allocation (LDA).

Proposed model follows a three-step generation process for affective terms, which first generates an emotion from a document-specific emotional distribution, then generates a latent topic from a Multinomial distribution conditioned on emotions, and finally generates document terms from another Multinomial distribution based on latent topics. Because its exact inference is intractable, developing an approximate inference method based on Gibbs sampling. For social emotion prediction, the proposed model outperforms the emotion term model, term-based SVM model, and topic-based SVM model significantly.

3.1 EMOTION TERM MODEL

Emotion-term model, follows the Naive Bayes method by assuming words are independently generated from social emotion labels. It generates each word $w_i$ of document $d$ in two sampling steps, i.e., sample an emotion $e_i$ according to the emotion frequency count $d$, and sample a word $w_i$ given the emotion under the conditional probability $P (w|e)$. The model parameters can be learned by maximum likelihood estimation. It can be formally derived based on the word and emotion frequency counts. To use the emotion-term models for predicting emotion on a new document $d$, apply the Bayes theorem (1) under the term independence assumption.

$$P(e|d) = \frac{\alpha P(d|e)P(e)}{P(d)}$$  \hspace{1cm} (1)

where $P(e)$ is the a priori probability of emotion $e$. It can again be calculated by maximum likelihood estimation (MLE) from the emotion distribution of the entire collection.

3.2 TOPIC MODEL

Many topic models have been proposed and well-studied in previous work, of which, LDA [8] is one of the most successful models. LDA addresses the over fitting problem faced by other models like pLSI by introducing a Dirichlet prior over topics and words. Although LDA can only discover the topics from document and cannot bridge the connection between social emotion and affective text, for the ease of understanding in the following description, a simple review of LDA is here. In the first study of LDA, proposed a convexity-based variation inference method for inference and parameter estimation under LDA.

$$P(z_i=j|z, w) \propto n_{ij}^{(w)} + \beta n_{ij}^{(d)} + \alpha n_{i-1j}^{(w)}$$  \hspace{1cm} (2)

where $n_{ij}$ means the count that does not include the current assignment of $z_i$, $n_{ij}^{(w)}$ is the number of times word $w$ has been assigned to topic $j$, and $n_{ij}^{(d)}$ is the number of times a word from document $d$ has been assigned to topic $j$.  

ISSN (Online): 2347 - 2812, Volume-2, Issue -11,12 2014 38
3.3 EMOTION TOPIC MODEL

Emotion-term model simply treats terms individually and cannot discover the contextual information within the document. While topic model utilizes the contextual information within the documents, it fails to utilize the emotional distribution to guide the topic generation. In this paper, proposing a new approach called emotion topic model. The importance of this latent topic generation in the affective text mining is very much. Likewise, different latent topics are discovered based on the emotions involved in it. Those latent topics should be collected together as a whole so that whenever needed it can be referred. After collecting each and every topic, it should be categorized on the basis of the different emotions such as love, happy, sad, sympathy, worry etc. They are used to select the document based on the preference assigned to the emotions. Relate the social emotions with an affective term that predict the emotions automatically from the text. After collecting and categorizing each latent topic based on different emotions, are stored to check with the extracted content. Then the topics are compared with the extracted content as a result of which it will generate topics and get processed.

For each word the posterior distribution on emotion “$E$” and topic “$z$” based on the following conditional probabilities which can be derived by the following equations (3).

$$P(E_i = e | \gamma, E, z, w; \alpha, \beta) = \frac{\alpha + \sum_{e}^{z} e_{i} \gamma d_{i}, e}{\alpha + \sum_{e}^{z} e_{i} \gamma d_{i}, e} \gamma d_{i}, e$$ (3)

Where $e$ and $z$ are the candidate emotion and topic for sampling, $d_i$ $\epsilon$ D indicates the document from which current word $w_i$ is sampled, $nz_{e, z}$ is the number of times topic $z$ has been assigned to emotion $e$.

IV. EXPERIMENTAL RESULTS

This section presents the experimental results on both joint emotion topic modelling and its application to emotion prediction. News articles were collected from the news portal and the input data’s are pre-processed to remove stem and stop words and perform tagging to extract the explicit words. Word frequency, document frequency were calculated. Emotion term model performs calculating word frequency and the emotion frequency count and the corresponding terms and emotion were obtained.
(b) Topic model

Emotion topic model bridges the connection between words and the emotion with the associated topic.

(c) Emotion topic model

The standard parameters which are used for experimental evaluation are precision, recall and accuracy. Precision is defined as number of retrieved relevant documents divided by total number of retrieved documents and the recall is the number of retrieved relevant document divided by total number of relevant documents in the database. Accuracy can be calculated as relevant document retrieved in top T returns divided by T.

\[
\text{Precision} = \frac{\text{Number of retrieved relevant document}}{\text{Total number of retrieved documents}}
\]

\[
\text{Recall} = \frac{\text{Number of retrieved relevant document}}{\text{Total number of relevant documents}}
\]

\[
\text{Accuracy} = \frac{\text{Relevant documents retrieved in top T}}{T}
\]

V. CONCLUSION

This paper presents and analyse a new problem called social affective text mining, which aims to discover and model the connections between online documents and user-generated social emotions. To this end, proposing a new joint emotion-topic model by augmenting Latent Dirichlet Allocation with an intermediate layer for emotion modelling. Rather than emotion term model that treats each term in the document individually and LDA topic model that only utilizes the text co-occurrence information, emotion-topic model allows associating the terms and emotions via topics which is more flexible and has better modelling capability.

REFERENCES


