PROMO for Interpretable Personalized Social Emotion Mining

Jason (Jiasheng) Zhang✉ and Dongwon Lee

The Pennsylvania State University, USA
{jpz5181, dongwon}@psu.edu

Abstract. Unearthing a set of users' collective emotional reactions to news or posts in social media has many useful applications and business implications. For instance, when one reads a piece of news on Facebook with dominating "angry" reactions, or another with dominating "love" reactions, she may have a general sense on how social users react to the particular piece. However, such a collective view of emotion is unable to answer the subtle differences that may exist among users. To answer the question "which emotion who feels about what" better, therefore, we formulate the *Personalized Social Emotion Mining (PSEM)* problem. Solving the PSEM problem is non-trivial in that: (1) the emotional reaction data is in the form of ternary relationship among user-emotion-post, and (2) the results need to be *interpretable*. Addressing the two challenges, in this paper, we develop an expressive probabilistic generative model, PROMO, and demonstrate its validity through empirical studies.

Keywords: Personalized Social Emotion Mining · ternary relationship Data · probabilistic generative models.

1 Introduction

It has become increasingly important for businesses to better understand their users and leverage the learned knowledge to their advantage. One popular method for such a goal is to mine users' digital footprints to unearth latent "emotions" toward particular products or services. In this paper, we focus on such a problem, known as *social emotion mining* (**SEM**) [24], to uncover latent emotions of social media posts, documents, or users. Note that SEM is slightly different from the conventional *sentiment analysis* (**SA**).

First and foremost, SEM is to unearth the latent emotion of a *reader* to a social post while SA is to detect underlying emotion of a *writer* of a social post or article. Therefore, for instance, knowing viewers' positive emotion toward a YouTube video for its frequent thumbs-up votes is the result of SEM, while knowing an underlying negative emotion of a news editorial is the result of SA. Second, in objectives, the goal of SEM is to find *which emotion people feel about what*, while SA aims to find *which emotion an author conveys through what* (referring to text in this work); Third, in methods, SEM learns the correlation between topics of text and emotions while SA needs to search for or learn

emotional structures, such as emotional words in text. For example, SEM is to predict the aggregated emotional reactions to a news article based on its content while an SA task may attempt to predict the emotions of customers based on the reviews that they wrote.

While useful, note that current SEM methods capture the emotional reactions of crowds as aggregated votes for different emotions. This brings two difficulties in understanding people's emotions in a finer granularity. On one hand, different people have their own ways to express similar emotions. For example, reading a sad story, some will react with "sad," but others may choose "love," both for expressing sympathy. When such emotions are aggregated, the cumulative emotions could be noisy. On the other hand, different people do not have to feel the same toward similar content. For instance, for the same Tweet, republican and democratic supporters may react opposite. In this case, no single emotion can represent the emotion of crowds well [3].

To close this gap, by observing personal difference as their common source, we formulate the novel *Personalized Social Emotion Mining* (**PSEM**) problem. Compared with SEM, here our goal is to find the hidden structures among the three dimensions, *users-emotions-posts*, in the emotional reactions data. The hidden structures are the clusters of different dimensions (i.e., groups of users, topics of posts) and the connections (i.e., the emotional reactions) among them. In short, we want to know *which emotion who feels about what*. This additional "who" helps reveal the differences in expression among users and also conserves the diversity in opinions. Such new knowledge in personalized emotion may help, for instance, business to provide personalized emotion based recommendation.

Note that it is nontrivial to model the ternary relationship of users-emotions-posts. More common relationship is binary (e.g., network, recommendation), where the closeness between the two entities (e.g., a user assigns 5 stars to an item) is modeled. The post dimension can be further expanded to the textual content of each post, which helps answer not only which post but also which topic triggers certain emotions. Moreover, because the task is to manifest the hidden structures of the data, the model applied should be well interpretable, which excludes the direct adoption of existing ternary relation learning methods [10, 19, 18, 31, 33, 29, 27, 12]. To the best of our knowledge, no existing models can satisfy the two requirements simultaneously.

To solve these challenges, in this work, we propose the **PROMO** (PRobabilistic generative cO-factorization MOdel), that learns the hidden structures in emotional reaction data by co-factorizing both ternary relationships and post-word binary relationships with a probabilistic generative model. PROMO is both *interpretable* and *expressive*. We especially showcase the interpretability of PROMO by the hidden structures of the emotional reactions revealed from two real-world dataset, which is not available by existing methods. In addition, the empirical experiments using two supervised learning tasks validate PROMO by its improved performance against state-of-the-art competing methods.

2 Related Work

As PSEM is a newly formulated problem, in this section, we review related literature in three aspects: (1) the social emotion mining problem (SEM) is related to PSEM as a real-world problem; (2) the ternary relationship data modeling problem is one of the challenges of the PSEM problem; and (3) probabilistic generative models are commonly applied to textual data when interpretability is concerned.

Social Emotion Mining (SEM). Previous works on SEM [17, 23, 25, 30, 4, 22, 38, 35, 5] attempt to find a mapping from online post content content (i.e., textual) to the cumulative votes for different emotions from the users. Different forms of the representation of the emotional reactions are studied. For example, [5] uses the normalized cumulative votes for different emotions as the emotion distribution of a post, [17, 25, 4] focus on the dominating emotion tags for posts, which leads to a classification problem; [23, 35] treat the emotional reactions as a ranking given the emotional tags and solve the label ranking problem with the imbalance challenge [35]. However, none of the existing works have added personalized views to SEM.

A few works have studied personalized emotion perception [2, 37, 34, 28, 36], which is to predict readers' perception of the emotional expressions in content. For example, given an image of flowers as content, the task is to predict whether a reader will tag keywords, expressing love or happiness. In such a case, the perception of emotion of objective content is less subjective compared with news content, such as political news in PSEM. As a result, methods used in personalized emotion perception [2, 37, 34, 28, 36], which does not explicitly consider the users-emotions-posts ternary relationship, are not a good fit to solve PSEM.

Ternary Relationship Data Modeling. One challenge of the PSEM problem is to model the ternary relationship data. Most previous methods are based on tensor factorization, such as Tucker decomposition [31] and CP (CANDECOMP/PARAFAC) decomposition [8]. Some are based on intuition from knowledge base, such as TransE [7], which still can be transformed into a factorization model [33]. A more advanced model based on neural network, the neural tensor network, is proposed in [29]. Such multi-relational learning methods have been applied to the tag recommendation [10], context-aware collaborative filtering [19], drug-drug interaction [18] and knowledge graph completion [26, 7, 29].

The addition of side information of post textual content brings more methods into our scope. For example, factorization machines [27] is a powerful method to model sparse interactions. When each user is treated as a task, multi-task learning methods [1, 12] are also investigated for personalized sentiment classification. However, the models currently used in the multi-relational learning problem are *not* interpretable enough to manifest the hidden structures of the data to answer the PSEM problem.

Probabilistic Generative Models. Since the classical topic models PLSA [16] and LDA [6], probabilistic generative model becomes a popular tool to analyze text data. There exist other closely-related works using topic model to enhance recommendation system [11, 32, 14]. For example, [14] uses topic model to analyze the legislator-vote-bill network to find the ideal point of legislators. The difference with this work is that the vote relation is still one-dimensional. Generally, no existing probabilistic generative models are designed to model ternary relationships.

3 Problem Formulation

In a PSEM problem, data is in the form of tuples <user,emotion, post>, where posts are typically news represented by short textual messages, such as headlines. Thereafter, the post is also referred to as the document. In this work, the bag-of-word representation is taken for documents.

Formally, there are four sets of nodes, U users $\mathcal{U} = \{\mu \in [U]\}$, where $[U] = \{1, 2, ..., U\}$, D documents $\mathcal{D} = \{d \in [D]\}$, E emotion labels $\mathcal{E} = \{e \in [E]\}$, and V distinct words $\mathcal{V} = \{v \in [V]\}$. There are two kinds of relationships. $R_e \subseteq \mathcal{U} \times \mathcal{E} \times \mathcal{D}$ is the collection of user emotional reactions to documents, where each document has M_d emotional reactions $\{\epsilon_{dm} | m \in [M_d]\}$ from different users $\{u_{dm} | m \in [M_d]\}$; R_w is the document-word relationship. $R_w \subseteq \mathcal{D} \times \mathcal{V}$, where each document has N_d words $\{w_{dn} | n \in [N_d]\}$. The problem framework is visualized in Fig.1. The annotation used is summarized in Table.1

Problem 1 (PSEM (Personalized Social Emotion Mining)) *Given users $\mathcal{U}$, documents $\mathcal{D}$, emotion labels $\mathcal{E}$ and vocabulary $\mathcal{V}$, find the hidden structures among them from relationships R_e and R_w.*

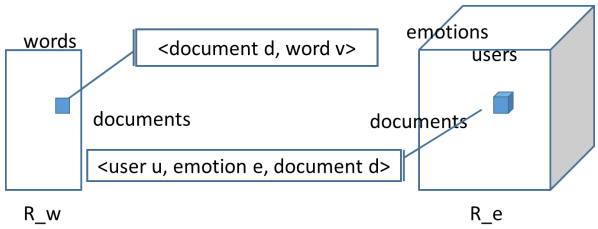


Fig. 1: PSEM data structure

4 Methodology

We propose a probabilistic generative co-factorization model (PROMO) to model the emotional reaction data. The probabilistic generative model itself provides a

Symbol	Description
R_e	the user-emotion-document relationship data
R_w	the document-word relationship data
$\mathcal{U}$	user set, of size $U = \|\mathcal{U}\|$, indexed by μ
$\mathcal{D}$	document (post) set, of size $D = \|\mathcal{D}\|$, indexed by d
$\mathcal{E}$	emotion label set, of size $E = \|\mathcal{E}\|$, indexed by e
$\mathcal{V}$	vocabulary, of size $V = \|\mathcal{V}\|$, indexed by v
N_d	number of words in the document d
M_d	number of emotional reactions to the document d
u_{dm}	a user, as a U-dimension one-hot vector
ϵ_{dm}	an emotional reaction, as an E-dimension one-hot vector
w_{dn}	a word, as a V-dimension one-hot vector
K	hyperparameter as the number of topics
G	hyperparameter as the number of groups
θ	corpus-wise topic distribution
ϕ	topic-word distribution
ψ	user-group distribution
η	group-topic-emotion distribution
z_d	topic indicator, as a K-dimension one-hot vector
x_{dm}	group indicator, as a G-dimension one-hot vector
α	hyperparameter for the Dirichlet prior of θ
β	hyperparameter for the Dirichlet prior of ϕ
ζ	hyperparameter for the Dirichlet prior of ψ
γ	hyperparameter for the Dirichlet prior of η

Table 1: Annotation summary

straightforward way to manifest the hidden structure of data, which meets the interpretability requirement. The two modules of PROMO to model the two relationships, R_w and R_e, are described separately, followed by the complete model description. In addition, the real-world interpretation of PROMO is discussed. After model construction, the inference algorithm is derived using stochastic variational inference. To valid the model, we show how to apply PROMO to two supervised learning tasks. Finally, the relationship between PROMO and existing models is discussed.

4.1 A Module for Short Documents

The most typical posts in the PSEM problem are news message posted by news channels in social media, such as CNN[1] in Facebook. These messages are usually short. Adopting the idea that there is only a single topic for each such short document in social media [9], the document-word relationship R_w is modeled as below.

For each document d, a K-dimensional one-hot variable z_d is associated to d, representing the unique topic of d. z_d is generated by a corpus-wise topic distribution, which is a K-dimensional multinomial distribution parameterized by θ, that $\forall k, \theta^k \geq 0, \sum_k \theta^k = 1$. Without ambiguity, the corpus-wise topic distribution is referred to as its parameter θ, similarly for other distributions introduced

[1] www.facebook.com/cnn

in this work. θ is generated according to a symmetric Dirichlet distribution parameterized by α.

Consistently with conventional topic model, the topic z_d generates the n_d words of the document d, i.i.d., according to the topic-word distributions parameterized by ϕ. For each topic $k \in [K]$, the topic-word distribution is a multinomial distribution parameterized by ϕ_k, that $\forall v \in [V]\, \phi_k^v \geq 0, \sum_v \phi_k^v = 1$ and ϕ_k is generated according to a symmetric Dirichlet distribution parameterized by β.

4.2 A Module for Ternary Relationships

Extended from the module for documents, each document d is summarized by its topic z_d.

Inspired by the social norm and latent user group theories [12], we assume users form G groups. For each emotional reaction ϵ_{dm}, indexed as $m \in [M_d]$ in reactions to the document d, the user $u_{dm} \in [U]$ [2] belongs to one group, represented by G-dimensional one-hot variable x_{dm}. x_{dm} is generated by u_{dm}, according to the user-group distributions parameterized by ψ. For each user $\mu \in [U]$, the user-group distribution is a multinomial distribution parameterized by ψ_u, that $\forall g \in [G]\, \psi_\mu^g \geq 0, \sum_g \psi_\mu^g = 1$. In other words, the group x_{dm} to which a user u_{dm} belongs is generated i.i.d according to $\psi_{u_{dm}}$, when all reactions from a user are considered. For each user μ, ψ_μ is generated according to a symmetric Dirichlet distribution parameterized by ζ.

To model the emotional reactions from different users toward different documents, we assume that the users from the same group react the same to the documents of the same topic. Formally, each emotional reaction ϵ_{dm} is generated by the combination of the topic of the document d, z_d and the group of the user u_{dm}, according to the group-topic-emotion distributions parameterized by η. For each topic k and group g, the group-topic-emotion distribution is a multinomial distribution parameterized by η_{gk}, that $\forall e \in [E]\, \eta_{gk}^e \geq 0, \sum_e \eta_{gk}^e = 1$ and η_{gk} is generated according to a symmetric Dirichlet distribution parameterized by γ.

4.3 PROMO: PRobabilistic generative cO-factorization MOdel

Our final PROMO model is made by combining two aforementioned modules. The annotations are summarized in Table.1. The graphic model representation and the generative process of PROMO is summarized below.

4.4 Interpretation

The discrete latent variable structure of PROMO gives clear translation of PSEM problem. *Which emotion who feels about what* is translated to *which emotion the user from which group feels about document about which topic*, which can be

[2] The U-dimensional one-hot variable u_{dm} is also used as its index of the non-zero entry interchangeably, which is applied to all one-hot variables in this work.

Require: K, G, α, ζ, β, γ
  $\theta^k \sim Dirichlet(\alpha)$
  $\psi_u^g \sim Dirichlet(\zeta)$
  $\phi_k^w \sim Dirichlet(\beta)$
  $\eta_{gk}^e \sim Dirichlet(\gamma)$
  for all $d \in [D]$ **do**
    $z_d^k \sim Multinomial(\theta)$
    $w_n^v \sim Multinomial(\phi_{z_d})$
    $x_{dm}^g \sim Multinomial(\psi_{u_{dm}})$
    $\epsilon_{dm}^e \sim Multinomial(\eta_{x_{dm} z_d})$
  end for

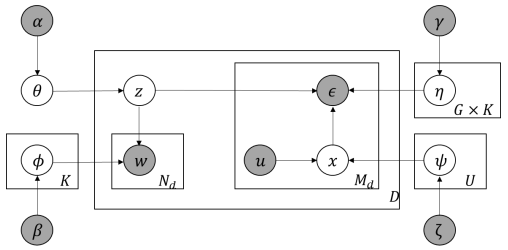


Fig. 2: Generative process and Graphical model representation of PROMO

interpreted from PROMO model variables. "Users from which group" is the user-group distribution ψ; "document about which topic" is the document topic z; and η can be interpreted as which emotion a group of users will feel about which topic of documents, which carries the core hidden structure of the emotional reaction data, that is, the answer toward the PSEM problem.

4.5 Inference

For the inference of PROMO with data, the stochastic variational inference [15] method is used. In the PSEM problem, the number of emotional reactions per document M can be very large (e.g., several thousands), which makes the collapsed Gibbs sampling [13], a more easily derivable inference method, too slow, due to the sequential sampling of each hidden variables that are not collapsed. With stochastic variational inference, within one document, the same type of latent variables can be inferred in parallel; and also the inference of each document within a batch can be trivially parallerized.

To approximate the intractable posterior distribution, we use a fully decomposable variational distribution q,

$$p(\theta, \psi, \phi, \eta, z, x | w, \epsilon; \Theta) \approx q(\theta, \psi, \phi, \eta, z, x | \overline{\theta}, \overline{\psi}, \overline{\eta}, \overline{z}, \overline{x}), \tag{1}$$

where Θ represents the set of hyperparameters $\{\alpha, \beta, \zeta, \gamma\}$, $\bar{\cdot}$ (e.g., $\overline{\theta}$, $\overline{\psi}$) are the parameters of q , the approximation is in terms of KL-divergence $D(q||p)$ and q can be decomposed as $q = q(\theta|\overline{\theta})q(\psi|\overline{\psi})q(\phi|\overline{\phi})q(\eta|\overline{\eta})q(z|\overline{z})q(x|\overline{x})$. For the variables as the parameters of multinomial distributions in PROMO, that is, θ, ψ, ϕ and η, the variational distributions are Dirichlet distributions; for the one-hot variables, z and x, the variational distributions are multinomial distributions.

The inference task, calculating the posterior distribution p is then reduced to finding the best variational distributions,

$$(\overline{\theta}^*, \overline{\psi}^*, \overline{\phi}^*, \overline{\eta}^*, \overline{z}^*, \overline{x}^*) = argmin_{\overline{\theta}, \overline{\psi}, \overline{\phi}, \overline{\eta}, \overline{z}, \overline{x}}(D(q||p)). \tag{2}$$

The optimization in eq.2 is done by iteratively optimizing each parameter. Readers who are interested in derivation detail can refer to the previous stochastic

Algorithm 1 Inference algorithm for PROMO

```
Initialize θ̄, ψ̄, φ̄, η̄ randomly.
set learning rate lr(t) function and batch size bs.
for t in 1 to MAXITERATION do
    Sample bs documents D_batch uniformly from data.
    for all d ∈ D_batch do
        Initialize z̄_d randomly.
        repeat
            For all g and m, update x̄^g_dm according to eq.3
            For all k, update z̄^k_d according to eq.4
        until converge
    end for
    for all par in {θ̄, ψ̄, φ̄, η̄} do
        Update par* according to eq.5
        Update par ← (1 − lr(t))par + lr(t)par*
    end for
end for
```

variational inference works [15, 6]. The update rules for those parameters are followed. Within each document d, for each emotional reaction indexed by m, the group distribution of the user u_{dm} is updated as

$$\overline{x}^g_{dm} \propto exp(\sum_{\mu} u^{\mu}_{dm} F(\overline{\psi}_{\mu})^g + \sum_k \sum_e \overline{z}^k_d \epsilon^e_{dm} F(\overline{\eta}_{gk})^e); \tag{3}$$

and the topic distribution of the document d is updated as

$$\overline{z}^k_d \propto exp(F(\overline{\theta})^k + \sum_n \sum_v w^v_{dn} F(\overline{\phi}_k)^v + \sum_m \sum_g \sum_e \overline{x}^g_{dm} \epsilon^e_{dm} F(\overline{\eta}_{gk})^e), \tag{4}$$

where $F(y)^l = \Psi(y^l) - \Psi(\sum_l y^l)$, with $\Psi()$ the digamma function. For corpus-level parameters, the updating rules are

$$\begin{aligned} \overline{\theta}^{k*} &= \alpha + \frac{D}{bs} \sum_{d \in \mathcal{D}_{batch}} \overline{z}^k_d, & \overline{\psi}^{g*}_{\mu} &= \zeta + \frac{D}{bs} \sum_{d \in \mathcal{D}_{batch}} \sum_m^{M_d} u^{\mu}_{dm} \overline{x}^g_{dm}, \\ \overline{\phi}^{v*}_k &= \beta + \frac{D}{bs} \sum_{d \in \mathcal{D}_{batch}} \sum_n^{N_d} w^v_{dn} \overline{z}^k_d, & \overline{\eta}^{e*}_{gk} &= \gamma + \frac{D}{bs} \sum_{d \in \mathcal{D}_{batch}} \sum_m^{M_d} \epsilon^e_{dm} \overline{x}^g_{dm} \overline{z}^k_d, \end{aligned} \tag{5}$$

where $\mathcal{D}_{batch}$ is the set of documents in a mini-batch and $bs = |\mathcal{D}_{batch}|$.

The complete stochastic variational inference algorithm for PROMO is shown in Algorithm. 1.

4.6 Supervised Learning Tasks

In order to validate the ability of the PROMO model to reveal the hidden structures of the PSEM data, we propose two supervised learning tasks which the PROMO can be applied to. The two tasks test whether the hidden structures PROMO reveals can be generalizable to unseen data.

1. Warm-start emotion prediction. This task is to predict the emotional reaction given a user and a document. It is called warm-start, because both the user and the document exist in R_e for training. For each user $\mu \in [U]$, document $d \in [D]$ and without loss of generality, indexing the emotional reaction as m, the posterior probability of $p(\epsilon_{dm}|\mu, d, R_w, R_e; \Theta)$ can be calculated in PROMO as

$$\int dz_d dx_{dm} d\eta d\psi \, p(\epsilon_{dm}|z_d, x_{dm}, \eta) p(z_d, \eta|\mu, R_w, R_e; \Theta) p(x_{dm}, \psi|\mu, R_w, R_e; \Theta) \tag{6}$$

where the two posterior distributions of z_d and x_{dm} can be replaced with the fitted variational distributions as $p(z_d, \eta|\mu, R_w, R_e; \Theta) \approx q(z_d|\overline{z}_d)q(\eta|\overline{\eta})$ and $p(x_{dm}, \psi|\mu, R_w, R_e; \Theta) \approx p(x_{dm}|\mu, \psi)q(\psi|\overline{\psi})$. Finally, because of the one-hot property of z_d, ϵ_{dm} and x_{dm}, the posterior distribution $p(\epsilon_{dm}|\mu, d, R_w, R_e; \Theta)$ can be derived as

$$\prod_e (\sum_k \sum_g \overline{z}_d^k \langle \eta \rangle_{gk}^e \langle \psi \rangle_\mu^g)^{\epsilon_{dm}^e}, \tag{7}$$

where for any $f \in \{\theta, \psi, \phi, \eta\}$, $\langle f \rangle$ is the variational mean of f, which is $\langle f \rangle^l = \overline{f}^l / \sum_{l'} \overline{f}^{l'}$ for $f \sim Dirichlet(\overline{f})$.

2. Cold-start emotion prediction. With the module for documents, PROMO can be applied to predicting the emotional reaction given a user and a new document. For a new document d with words $w_d = \{w_{dn}|n \in [N_d]\}$, the posterior probability of $< \mu, \epsilon_{dm}, d >$ can be calculated followed the similar derivations for eq.6 and eq.7, as

$$p(\epsilon_{dm}|\mu, w_d, R_w, R_e; \Theta) \approx \prod_e (\sum_k \sum_g \widehat{z}_d^k \langle \eta \rangle_{gk}^e \langle \psi \rangle_\mu^g)^{\epsilon_{dm}^e}, \tag{8}$$

where $\widehat{z}_d$ is the estimated topic distribution of d, which can be calculated as $\widehat{z}_d^k \propto \prod_v \langle \phi \rangle_k^{v(\sum_n^{N_d} w_{dn}^v)} \langle \theta \rangle^k$. Compared with the warm-start prediction, eq.7, where $\overline{z}_d$ is inferred from both the words and also the emotional reactions of the document d, the cold-start prediction, eq.8, uses $\widehat{z}_d$, which is estimated only from the words w_d of the document d.

4.7 The Relation to Existing Models

In PROMO, we introduce a group-topic-emotion distribution η to address the challenge of the ternary relationship user-emotion-document. The reconstruction of R_e in the warm-start emotion prediction (Sec.4.6), eq.7, can be translated as the factorization of the $R_e \in \{0, 1\}^{U \times E \times D}$ into the document latent vectors $\overline{z} \in [0, 1]^{D \times K}$, the user latent vectors $\langle \psi \rangle \in [0, 1]^{U \times G}$ and the emotion interaction core $\langle \eta \rangle \in [0, 1]^{G \times K \times E}$. It is equivalent in terms of expressive power to the RESCAL [26] model, a variant of Tucker decomposition, proposed for multi-relational learning. It can be proved by observing that any R_e constructed

with the RESCAL model can also be constructed with PROMO (i.e., eq.7) by applying rescale factors to the corresponding vectors in RESCAL.

One of the differences between PROMO and RESCAL is that the document latent vector in PROMO, $\overline{z}$ are inferred from both R_w and R_e, while the counterpart in RESCAL only from R_e. Therefore, in warm-start prediction task, R_w serves as the regularization. On the other hand, the generative architecture grants PROMO better interpretability.

5 Experimental Validation

In this section, we apply PROMO to real-world data to answer two questions: (1) *Interpretability*: what can be revealed from emotional reactions data by PROMO? (2) *Validity*: Is PROMO a valid model for emotional reactions data? All code[3] and dataset[4] are publicly available.

5.1 Data Description

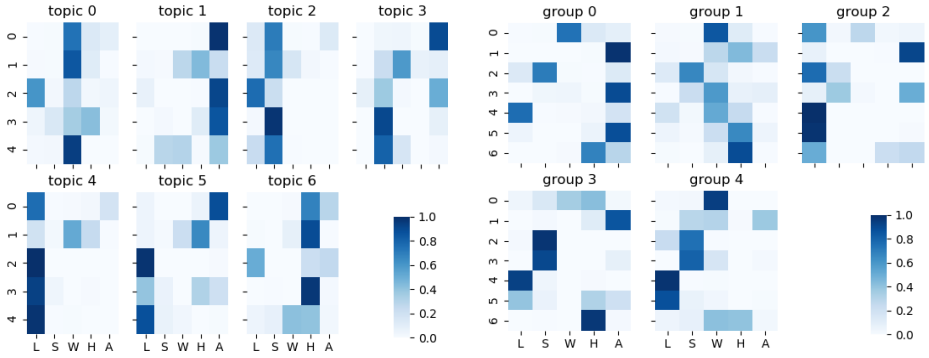
We crawled the emotional reactions data from the public posts published in news pages of Facebook. More specifically, we crawled posts and corresponding user emotional reactions from Fox News [5] page from May 13th to October 17th, 2016, and CNN [6] page from March 1, 2016 to July 14, 2017. As for documents, we use the post message, which is short and headline-like text, appearing in most posts of news pages; as for emotional reactions, we use the emoticon labels that users click for the posts. Besides the two data sets, we combine them and keep posts within the same publication period into a new data set.

We excludes posts without any words. For each document, we remove URL's and stop words; and all numbers are replaced with a unique indicator "NUMBER", as number in news title is often informative. For emotional reactions, clicks of "like" are excluded due to its ambiguous meaning [21]; after that, to eliminate noise, only documents and users with more than 10 emotional reactions are used, which is 10-core decomposition. The resulting data statistics is shown in Table.2.

5.2 Interpretability: What can PROMO reveal?

We apply PROMO to COMBINE data with $K = 7$ and $G = 5$. After the inference, we visualize and analyze the topic-word distribution ϕ and the group-topic-emotion distribution η to describe the hidden structures of the emotional reaction data, which is the answer to the PSEM problem.

[3] http://github.com/JasonLC506/PSEM
[4] http://tiny.cc/ecml20
[5] http://www.facebook.com/FoxNews/
[6] http://www.facebook.com/cnn/



(a) Group-emotion distribution for each topic, where rows are groups, and columns are emotion labels

(b) Topic-emotion distribution for each group, where rows are topics, and columns are emotion labels

Fig. 3: The group-topic-emotion distributions are visualized as slices on the topic dimension and the group dimension, respectively. The marks of emotion labels are L for LOVE, S for SAD, W for WOW (surprise), H for HAHA, and A for ANGRY.

Topic words. The posterior topic word distribution, $\langle \phi \rangle$, is visualized as top words for each topic in Table.3. We found four clusters of topics in terms of their top words. Topic 0 is about surprising news, topic 2 is about disaster news, topic 3 is about violence news and topic 1,4,5,6 are all about political news. This phenomenon is due to the fact that the topic in PROMO is inferred from both R_w and R_e. Therefore, topics with similar topic-word distributions appear when their emotional reactions are highly different.

Emotional reactions. First, the slices of the posterior group-topic-emotion distribution $\langle \eta \rangle$ on the topic dimension are visualized in Fig.3a to show the user difference in emotional reactions within each single topic. For example, within topic 2, known as some disaster news from Table.3, cause most groups of users to react with SAD, while the users from the group 2 tend to react with LOVE. Similar example can be observed in topic 3, where two kinds of groups of users tend to use SAD and ANGRY, respectively. In addition, as an example of controversial topics, topic 5, related to political news, attracts LOVE from users in group 2, 3 and 4, but ANGRY from those in group 0.

Next, the slices of $\langle \eta \rangle$ on the group dimension are visualized in Fig.3b to show the emotional reaction difference toward different topics from each single group of users. For example, users from group 0 tend to react with LOVE to the topic 4 but ANGRY to the topic 1, 3 and 5, though topics 1, 4 and 5 are similar in their topic-word distributions. Moreover, comparing group 1 and group 2 in general, users from group 1 generally prefer SAD, WOW and HAHA rather than emotions with more tension, such as LOVE and ANGRY, and vice versa

	CNN	Fox News	COMBINE
# users	869,714	355,244	524,888
# documents	23,236	3,934	10,981
# reactions	38,931,052	12,257,816	17,327,148
# emotion labels	5	5	5
# words (total)	300,135	66,241	158,104
% LOVE	23.4	26.8	25.9
% SAD	18.4	12.2	15.4
% WOW	13.1	8.3	10.2
% HAHA	21.4	14.2	15.7
% ANGRY	23.7	38.5	32.8

Table 2: data statistics

Topic	Top words
0	NUMBER, world, one, new, years, life, found
1	clinton, hillary, trump, donald, said, j, NUMBER
2	NUMBER, one, police, people, died, said, us
3	NUMBER, police, people, said, one, new, killed
4	trump, j, donald, NUMBER, said, people, clinton
5	trump, clinton, hillary, j, donald, president, obama
6	trump, donald, j, clinton, hillary, NUMBER, said

Table 3: Topics found on COMBINE data

for those from group 2. It shows that such results can also show the general difference between two groups of users.

5.3 Validity: Is PROMO good for emotional reaction data?

We answer this question by comparing the performance of PROMO and that of competing methods in two supervised learning tasks in real-world emotional reaction datasets.

Experimental setting. For each data set, among all posts, 10% are randomly selected as the cold-start prediction test set, 5% as the validation set; Within the remaining posts, among all emotional reactions, 10% are randomly selected as the warm-start prediction test set, 5% as the validation set, and remaining as the training set for both tasks. The validation set is used for hyperparameter tuning and the convergence check for each task.

Evaluation measures. We formulate both tasks as multi-class classification problems, given that there can be only one emotional reaction given a single <document, user> pair. As a result, macro AUC-precision-recall, $AUCPR$ is used as evaluation measure, which is known more robust in imbalanced case.

Warm-start emotion prediction. We use tensor factorization based multi-relational learning methods as baselines to show that PROMO learns from text data. Besides, we use an adapted personalized emotion perception model, the factorization machine and a multi-task sentiment classification model to assert the necessity to explicitly consider the users-emotions-posts ternary relationship in PSEM.

- PROMO: the model proposed in this work.
- PROMO_NT: PROMO trained without text data of posts.
- RESCAL: model based on Tucker decomposition with relation dimension summed up with core tensor, which is similar to PROMO [26].
- CP: CANDECOMP/PARAFAC (CP) decomposition [8].
- MultiMF: separate matrix factorization for each relation type [21].
- NTN (Neural Tensor Network) [29]: a method combining tensor factorization and neural network to catch complex interactions among different dimensions.

The former five baselines (PROMO_NT, RESCAL, CP, MultiMF and NTN) are tensor factorization based multi-relational learning methods; while the followings make use of the textual content of the post.

- BPEP (Bayesian Personalized Emotion Perception): We adopt the Bayesian text domain module of the personalized Emotion perception work [28]. It builds naive Bayesian estimate of emotion distribution of each word for each user.
- FM (Factorization Machine): We use the rank-2 factorization machine [27] to model the interaction between textual content, posts and users. The feature is the concatenation of bag-of-word textual feature, post identifier and user identifier, both as one-hot vectors.
- MT-LinAdapt (Multi-Task Linear Model Adaptation) [12]: a state-of-the-art multi-task sentiment classification method that is shown to outperform its two-stage counterpart LinAdapt [1]. In our case, we adopt the setting that each user is a task. The bag-of-word textual feature is used as input for minimum comparison.
- CUE-CNN [2]: a state-of-the-art model for personalized sentiment classification, which employs convolutional neural network to capture textual content.
- CADEN [36]: a state-of-the-art model for personalized sentiment classification, which employs recurrent neural network to capture textual content.

All methods are adapted to use negative log likelihood as loss function for training. We implement RESCAL, CP, MultiMF, NTN, FM, MT-LinAdapt and CADEN using stochastic gradient descent with Adam [20]. We used validation set to search for the best hyperparameters for each model. For PROMO, $K = 50$ and $G = 64$, $\alpha = 1.0/K$, $\beta = 1.0$, $\gamma = 1.0$, $\zeta = 0.1$ and batch size is set to full batch; same setting for PROMO_NT, except $K = 30$ and $G = 16$.

As results shown in Table.4, PROMO consistently outperforms baseline methods in all data sets. From the comparison between PROMO and the tensor factorization based methods, we assert that PROMO learns from text data. On the other hand, the comparison between PROMO and BPEP, FM, MT-LinAdapt, CUE-CNN and CADEN supports that only text information, without explicitly considering the users-emotions-posts ternary relationship, is not enough in PSEM problem, which also distinguish PSEM problem from personalized sentiment analysis [12, 1, 36].

A detail check over the performance in different datasets provides some clue for the superiority of PROMO. From Table.4, in CNN dataset, the best of former five tensor factorization based methods, NTN, outperforms the latter three methods that focus more on textual information; while in Fox News and COMBINE datasets, the best of the latter three methods, FM, is comparable or better than the former five. This observation reveals the different contribution of users-emotions-posts interactions and textual information in different datasets (i.e., more contribution of the interactions for CNN while less in Fox News and COMBINE). However, the Bayesian architecture of PROMO provides an automatic balance between the two aspects, so that consistently outperforms others in all datasets.

	CNN	Fox News	COMBINE
PROMO	**0.809**	**0.779**	**0.791**
PROMO_NT	0.791	0.766	0.773
RESCAL	0.793	0.768	0.759
CP	0.776	0.753	0.760
MultiMF	0.748	0.697	0.698
NTN	0.805	0.685	0.733
BPEP	0.445	0.431	0.445
FM	0.785	0.774	0.773
MT-LinAdapt	0.741	0.750	0.733
CUE-CNN	0.787	0.706	0.736
CADEN	0.740	0.665	0.695

Table 4: results on warm-start emotion prediction task

	CNN	Fox News	COMBINE
PROMO	0.601	**0.462**	**0.525**
PROMO_NG	0.520	0.404	0.424
BPEP	0.431	0.429	0.450
eToT	0.396	0.364	0.391
CUE-CNN	**0.625**	0.421	0.501
CADEN	0.565	0.388	0.453

Table 5: results on cold-start emotion prediction task

Cold-start emotion prediction. There are less existing works that can be applied to cold-start emotion prediction task. We test following competing models, besides PROMO, BPEP, CUE-CNN and CADEN described in the previous task.

- PROMO_NG: the PROMO with number of group set to 1, that is $G = 1$. This variant shows the situation when user difference is not considered.
- eToT: a probabilistic generative model for social emotion mining with temporal data, [38], excluding the temporal components, which is implemented using collapsed Gibbs sampling [13].

We used validation set to search for the best hyperparameters for each model. For PROMO, we set $K = 50$ and $G = 16$, $\alpha = 1.0/K$, $\beta = 0.01$, $\gamma = 100.0$, $\zeta = 0.1$ and batch size is set to full batch. For eToT, $K = 50$ is used.

As results shown in Table.5, PROMO outperforms other models in all but CNN dataset. Compared with that in warm-start task, models with more advanced textual feature extraction methods, i.e., CUE-CNN and CADEN perform much better. For example, CUE-CNN obtains a result even better than PROMO. However, those deep learning based model lost the interpretability of PROMO. In more detail, the improvement from PROMO_NG to PROMO supports the basic assumption of PSEM that users are different in emotional reaction. Besides, the comparable results between PROMO_NG and BPEP support the conclusion in the previous experiment, that only text information, as used in BPEP, may not be enough to describe user difference.Finally, the comparison between results of PROMO and eToT show that PROMO takes a superior probabilistic generative architecture for PSEM problem.

6 Conclusion

In this work, we formulate the novel Personalized Social Emotion Mining (PSEM) problem, to find the hidden structures of emotional reaction data. As a solution, then, we develop the PROMO (PRobabilistic generative cO-factorization

MOdel), which is both well interpretable and expressive to address the PSEM problem. We showcase its interpretability by the meaningful hidden structures found by PROMO on a real-world data set. We also demonstrate that PROMO is a valid and effective model for emotional reactions data by showing its superiority against competing methods in two supervised learning tasks.

7 Acknowledgement

The authors would like to thank the anonymous referees and Noseong Park at GMU for their valuable comments. This work was supported in part by NSF awards #1422215, #1525601, #1742702, and Samsung GRO 2015 awards.

References

1. Al Boni, M., Zhou, K., Wang, H., Gerber, M.S.: Model adaptation for personalized opinion analysis. In: ACL. vol. 2, pp. 769–774 (2015)
2. Amir, S., Wallace, B.C., Lyu, H., Carvalho, P., Silva, M.J.: Modelling context with user embeddings for sarcasm detection in social media. In: SIGNLL (2016)
3. Arrow, K.J.: A difficulty in the concept of social welfare. Journal of political economy **58**(4), 328–346 (1950)
4. Bai, S., Ning, Y., Yuan, S., Zhu, T.: Predicting readers emotion on chinese web news articles. In: ICPCA-SWS. pp. 16–27. Springer (2012)
5. Bao, S., Xu, S., Zhang, L., Yan, R., Su, Z., Han, D., Yu, Y.: Joint emotion-topic modeling for social affective text mining. In: 2009 Ninth IEEE International Conference on Data Mining. pp. 699–704. IEEE (2009)
6. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. JMLR **3**(Jan), 993–1022 (2003)
7. Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., Yakhnenko, O.: Translating embeddings for modeling multi-relational data. In: NIPS. pp. 2787–2795 (2013)
8. Carroll, J.D., Chang, J.J.: Analysis of individual differences in multidimensional scaling via an n-way generalization of eckart-young decomposition. Psychometrika **35**(3), 283–319 (1970)
9. Ding, Z., Qiu, X., Zhang, Q., Huang, X.: Learning topical translation model for microblog hashtag suggestion. In: IJCAI. pp. 2078–2084 (2013)
10. Feng, W., Wang, J.: Incorporating heterogeneous information for personalized tag recommendation in social tagging systems. In: SIGKDD. pp. 1276–1284 (2012)
11. Gerrish, S., Blei, D.M.: How they vote: Issue-adjusted models of legislative behavior. In: NIPS. pp. 2753–2761 (2012)
12. Gong, L., Al Boni, M., Wang, H.: Modeling social norms evolution for personalized sentiment classification. In: ACL. vol. 1, pp. 855–865 (2016)
13. Griffiths, T.L., Steyvers, M.: Finding scientific topics. PNAS **101**(suppl 1), 5228–5235 (2004)
14. Gu, Y., Sun, Y., Jiang, N., Wang, B., Chen, T.: Topic-factorized ideal point estimation model for legislative voting network. In: SIGKDD. pp. 183–192 (2014)
15. Hoffman, M.D., Blei, D.M., Wang, C., Paisley, J.: Stochastic variational inference. The Journal of Machine Learning Research **14**(1), 1303–1347 (2013)
16. Hofmann, T.: Probabilistic latent semantic analysis. In: UAI. pp. 289–296 (1999)

17. Jia, Y., Chen, Z., Yu, S.: Reader emotion classification of news headlines. In: NLP-KE 2009. pp. 1–6. IEEE (2009)
18. Jin, B., Yang, H., Xiao, C., Zhang, P., Wei, X., Wang, F.: Multitask dyadic prediction and its application in prediction of adverse drug-drug interaction. In: AAAI. pp. 1367–1373 (2017)
19. Karatzoglou, A., Amatriain, X., Baltrunas, L., Oliver, N.: Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering. In: RecSys. pp. 79–86. ACM (2010)
20. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
21. Lee, S.Y., Hansen, S.S., Lee, J.K.: What makes us click like on facebook? examining psychological, technological, and motivational factors on virtual endorsement. Computer Communications **73**, 332–341 (2016)
22. Lei, J., Rao, Y., Li, Q., Quan, X., Wenyin, L.: Towards building a social emotion detection system for online news. Future Generation Computer Systems **37**, 438–448 (2014)
23. Lin, K.H.Y., Chen, H.H.: Ranking reader emotions using pairwise loss minimization and emotional distribution regression. In: EMNLP. pp. 136–144 (2008)
24. Lin, K.H.Y., Yang, C., Chen, H.H.: What emotions do news articles trigger in their readers? In: SIGIR. pp. 733–734. ACM (2007)
25. Lin, K.H.Y., Yang, C., Chen, H.H.: Emotion classification of online news articles from the reader's perspective. In: WI-IAT'08. vol. 1, pp. 220–226. IEEE (2008)
26. Nickel, M., Tresp, V., Kriegel, H.P.: A three-way model for collective learning on multi-relational data. In: ICML. vol. 11, pp. 809–816 (2011)
27. Rendle, S.: Factorization machines. In: ICDM. pp. 995–1000. IEEE (2010)
28. Rui, T., Cui, P., Zhu, W.: Joint user-interest and social-influence emotion prediction for individuals. Neurocomputing **230** (2017)
29. Socher, R., Chen, D., Manning, C.D., Ng, A.: Reasoning with neural tensor networks for knowledge base completion. In: NIPS. pp. 926–934 (2013)
30. Tang, Y.j., Chen, H.H.: Emotion modeling from writer/reader perspectives using a microblog dataset. In: Proceedings of IJCNLP Workshop on sentiment analysis where AI meets psychology. pp. 11–19 (2011)
31. Tucker, L.R.: Some mathematical notes on three-mode factor analysis. Psychometrika **31**(3), 279–311 (1966)
32. Wang, E., Liu, D., Silva, J., Carin, L., Dunson, D.B.: Joint analysis of time-evolving binary matrices and associated documents. In: NIPS. pp. 2370–2378 (2010)
33. Yang, B., Yih, W.t., He, X., Gao, J., Deng, L.: Embedding entities and relations for learning and inference in knowledge bases. arXiv preprint arXiv:1412.6575 (2014)
34. Yang, Y., Cui, P., Zhu, W., Yang, S.: User interest and social influence based emotion prediction for individuals. Multimedia - MM '13 pp. 785–788 (2013)
35. Zhang, J.J., Lee, D.: Roar: Robust label ranking for social emotion mining. In: AAAI (2018)
36. Zhang, L., Xiao, K., Zhu, H., Liu, C., Yang, J., Jin, B.: CADEN : A Context-Aware Deep Embedding Network for Financial Opinions Mining. ICDM pp. 757–766 (2018)
37. Zhao, S., Yao, H., Gao, Y., Ding, G., Chua, T.S.: Predicting Personalized Image Emotion Perceptions in Social Networks. TAC **X**(X), 1–1 (2016)
38. Zhu, C., Zhu, H., Ge, Y., Chen, E., Liu, Q.: Tracking the evolution of social emotions: A time-aware topic modeling perspective. In: ICDM. pp. 697–706. IEEE (2014)