

Sentence-level Emotion Classification with Label and Context Dependence

Shoushan Li^{†‡}, Lei Huang[†], Rong Wang[†], Guodong Zhou^{†*}

[†]Natural Language Processing Lab, Soochow University, China

[‡] Collaborative Innovation Center of Novel Software Technology and Industrialization
{shoushan.li, lei.huang2013, wangrong2022}@gmail.com,
gdzhou@suda.edu.cn

Abstract

Predicting emotion categories, such as *anger*, *joy*, and *anxiety*, expressed by a sentence is challenging due to its inherent multi-label classification difficulty and data sparseness. In this paper, we address above two challenges by incorporating the label dependence among the emotion labels and the context dependence among the contextual instances into a factor graph model. Specifically, we recast sentence-level emotion classification as a factor graph inferring problem in which the label and context dependence are modeled as various factor functions. Empirical evaluation demonstrates the great potential and effectiveness of our proposed approach to sentence-level emotion classification.

1 Introduction

Predicting emotion categories, such as *anger*, *joy*, and *anxiety*, expressed by a piece of text encompasses a variety of applications, such as online chatting (Galik et al., 2012), news classification (Liu et al., 2013) and stock marketing (Bollen et al., 2011). Over the past decade, there has been a substantial body of research on emotion classification, where a considerable amount of work has focused on document-level emotion classification.

Recently, the research community has become increasingly aware of the need on sentence-level emotion classification due to its wide potential applications, e.g. the massively growing importance of analyzing short text in social media (Kiritchenko et al., 2014; Wen and Wan, 2014). In general, sentence-level emotion classification exhibits two challenges.

.....

<S1>她们都睡了，我蹑手蹑脚摸黑上了床，凑上去想亲她一下，她突然一个转身，小手“啪”地搭在了我的脸颊上。</S1> <S2>现在我终于如愿以偿。</S2> <S3>感受着小手的温度，享受着这份她对我的依恋，生怕动一下，会让她的小手离我而去。</S3>.....

(English:

<S1> The girls fall to sleep, so I make my way noiselessly onto the bed, wishing I could get a chance to give a kiss to Yan, suddenly she turn over to me and her little soft hand fall onto my face.</S1> <S2>Praise the Lord, that is all I want.</S2> <S3>Feeling the warm of her hand and the attachment she hold to me, I couldn't afford to move even a little, fearing I may lost her hand.</S3>.....)

Sentence-level Emotion Classification

► **Input:** S1, S2, S3

► **Output:** S1 : joy, love

S2: joy

S3: joy, love, anxiety

Figure 1: An example of a paragraph and the sentences therein with their emotion categories from the corpus collected by Quan and Ren (2009)

On one hand, like document-level emotion classification, sentence-level emotion classification is naturally a multi-label classification problem. That is, each sentence might involve more than one emotion category. For example, as shown in Figure 1, in one paragraph, two sentences, i.e., S1 and S3, have two and three emotion categories respectively. Automatically classifying instances with multiple possible categories is

* Corresponding author

sometimes much more difficult than classifying instances with a single label.

On the other hand, unlike document-level emotion classification, sentence-level emotion classification is prone to the data sparseness problem because a sentence normally contains much less content. Given the short text of a sentence, it is often difficult to predict its emotion due to the limited information therein. For example, in **S2**, only one phrase “*如願以償*(*that is all I want*)” expresses the *joy* emotion. Once this phrase fails to appear in the training data, it will be hard for the classifier to give a correct prediction according to the limited content in this sentence.

In this paper, we address above two challenges in sentence-level emotion classification by modeling both the label and context dependence. Here, the label dependence indicates that multiple emotion labels of an instance are highly correlated to each other. For instance, the two positive emotions, *joy* and *love*, are more likely to appear at the same time than the two counterpart emotions, *joy* and *hate*. The context dependence indicates that two neighboring sentences or two sentences in the same paragraph (or document) might share the same emotion categories. For instance, in Figure 1, **S1**, **S2**, and **S3**, from the same paragraph, all share the emotion category *joy*.

Specifically, we propose a factor graph, namely Dependence Factor Graph (DFG), to model the label and context dependence in sentence-level emotion classification. In our DFG approach, both the label and context dependence are modeled as various factor functions and the learning task aims to maximize the joint probability of all these factor functions. Empirical evaluation demonstrates the effectiveness of our DFG approach to capturing the inherent label and context dependence. To the best of our knowledge, this work is the first attempt to incorporate both the label and context dependence of sentence-level emotion classification into a unified framework.

The remainder of this paper is organized as follows. Section 2 overviews related work on emotion analysis. Section 3 presents our observations on label and context dependence in the corpus. Section 4 proposes our DFG approach to sentence-level emotion classification. Section 5 evaluates the proposed approach. Finally, Section 6 gives the conclusion and future work.

2 Related Work

Over the last decade, there has been an explosion of work exploring various aspects of emotion analysis, such as emotion resource creation (Wiebe et al., 2005; Quan and Ren, 2009; Xu et al., 2010), writer’s emotion vs. reader’s emotion analysis (Lin et al., 2008; Liu et al., 2013), emotion cause event analysis (Chen et al., 2010), document-level emotion classification (Alm et al., 2005; Li et al., 2014) and sentence-level or short text-level emotion classification (Tokushisa et al., 2008; Bhowmick et al., 2009; Xu et al., 2012). This work focuses on sentence-level emotion classification.

Among the studies on sentence-level emotion classification, Tokushisa et al. (2008) propose a data-oriented method for inferring the emotion of an utterance sentence in a dialog system. They leverage a huge collection of emotion-provoking event instances from the Web to deal with the data sparseness problem in sentence-level emotion classification. Bhowmick et al. (2009) and Bhowmick et al. (2010) apply KNN-based classification algorithms to classify news sentences into multiple reader emotion categories. Although the multi-label classification difficulty has been noticed in their study, the label dependence is not exploited. More recently, Xu et al. (2012) proposes a coarse-to-fine strategy for sentence-level emotion classification. They deal with the data sparseness problem by incorporating the transfer probabilities from the neighboring sentences to refine the emotion categories. To some extent, this can be seen a specific kind of context information. However, they ignore the label dependence by directly applying Binary Relevance to overcome the multi-label classification difficulty.

Unlike all above studies, this paper emphasizes the importance of the label dependence and exploits it in sentence-level emotion classification via a factor graph model. Moreover, besides the label dependence, our factor graph-based approach incorporates the context dependence in a unified framework to further improve the performance of sentence-level emotion classification.

3 Observations

To better illustrate our motivation of modeling the label and context dependence, we systematically investigate both dependence phenomena in our evaluation corpus.

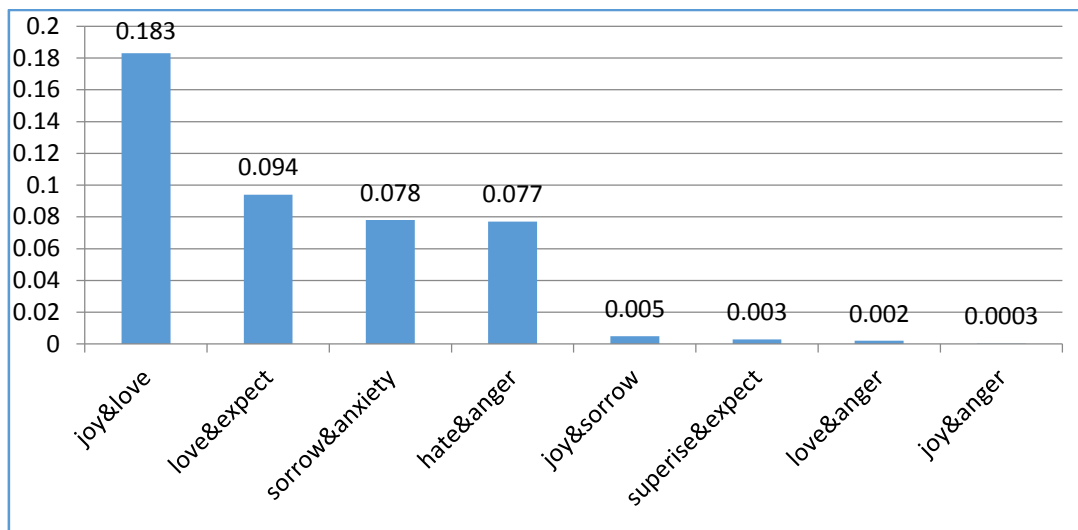


Figure 2: Probability distribution of most and least frequently-occurred pairs of emotion categories, with left four most frequently-occurred and right four least frequently-occurred, among all 28 pairs

The corpus contains 100 documents, randomly selected from Quan and Ren (2009). There are totally 2751 sentences and each of them is manually annotated with one or more emotion labels.

Table 1: The numbers of the sentences in each emotion category

Emotion	#Sentence	Emotion	#Sentence
<i>joy</i>	691	<i>anxiety</i>	567
<i>hate</i>	532	<i>surprise</i>	180
<i>love</i>	1025	<i>anger</i>	287
<i>sorrow</i>	611	<i>expect</i>	603

Table 2: The numbers of the sentences grouped by the emotion labels they contain

	#Sentence
No Label	180
One Label	1096
Two Labels	1081
Three Labels	346
Four or more labels	48
ALL	2751

Table 1 shows the sentence distribution of the eight emotion categories. Obviously, the distribution is a bit imbalanced. While about to one quarter of sentences express the emotion category *love*, only ~6% and ~10% express surprise and anger respectively, with the remaining 5 emotion categories distributed rather evenly from ~20% to

~25%. Table 2 shows the numbers of the sentences grouped by the emotion labels they contain. From this table, we can see that more than half sentences have two or more emotion labels. This indicates the popularity of the multi-label issue in sentence-level emotion classification.

To investigate the phenomenon of label dependence, we first assume that $X \subset R^d$ denotes an input domain of instances and $Y = \{l_1, l_2, \dots, l_m\}$ be a finite domain of possible emotion labels. Each instance is associated with a subset of Y and this subset is described as an m -dimensional vector $y = \{y^1, y^2, \dots, y^m\}$ where $y^i = 1$ only if instance x has label l_i , and $y^i = 0$ otherwise. Then, we can calculate the probability that an instance takes both emotion labels l_i and l_j , denoted as

$p(l_i, l_j)$. Figure 2 shows the probability distribution of most and least frequently-occurred pairs of emotion categories, with left four most frequently-occurred and right four least frequently-occurred, among all 28 pairs. From this figure, we can see that some pairs, e.g., *joy* and *love*, are much more likely to be taken by one sentence than some other pairs, e.g. *joy* and *anger*.

Finally, we investigate the phenomenon of the context dependence by calculating the probabilities that two instances x_k and x_l have at least one identical emotion label, i.e., $p(y_k \cap y_l)$ in different settings.

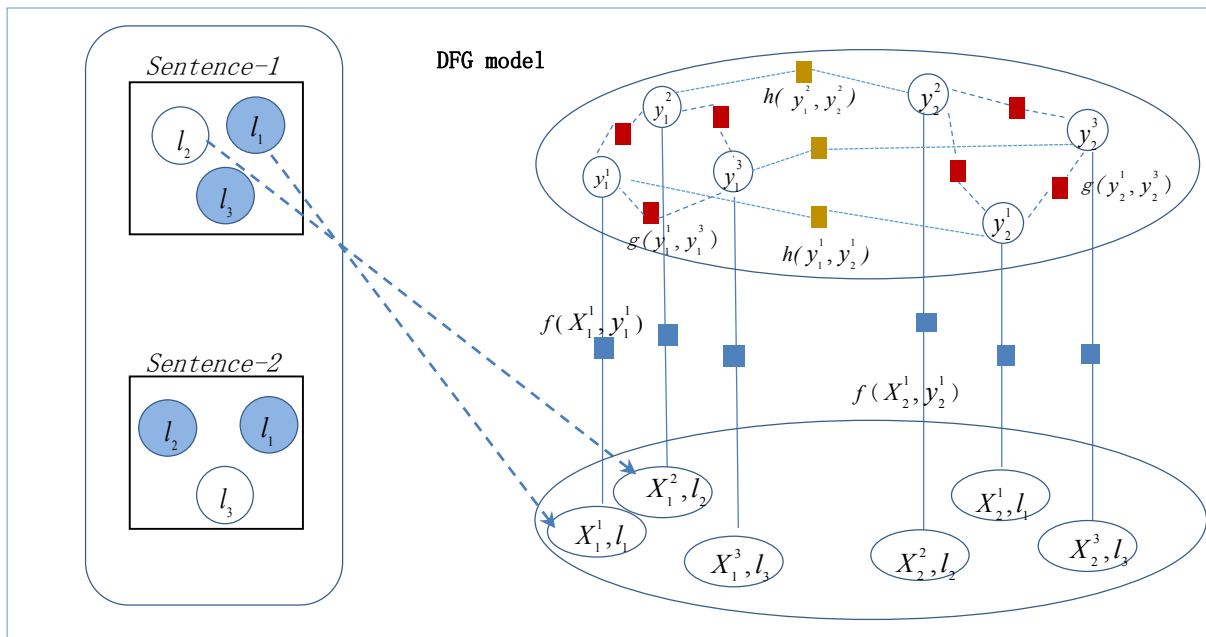


Figure 4: An example of DFG when two instances are involved: *sentence-1* with the label vector $[1, 0, 1]$ and *sentence-2* with the label vector $[1, 1, 0]$

Note: each multi-label instance is transformed into three pseudo samples, represented as X_i^k ($k=1,2,3$). $f(\cdot)$ represents a factor function for modeling textual features. $g(\cdot)$ represents a factor function for modeling the label dependence between two pseudo samples. $h(\cdot)$ represents a factor function for modeling the context dependence between two instances in the same context.

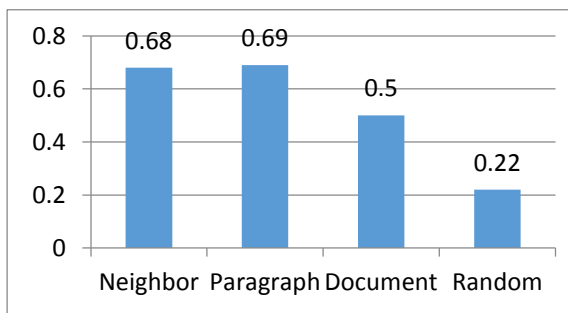


Figure 3: Probabilities that two instances have an identical emotion label in different settings

Figure 3 shows the probabilities that two instances have at least one identical emotion label in different settings, where *neighbor*, *paragraph*, *document* and *random* mean two neighboring instances, two instances from the same paragraph, two instances from the same document, and two instances from a random selection, respectively. From this figure, we can see that two instances from the same context are much more likely to take an identical emotion label than two random instances.

From above statistics, we come to two basic observations:

- 1) **Label dependency:** One sentence is more likely to take some pair of emotion labels, e.g.,

hate and *angry* than some other pair of emotion labels, e.g., *hate* and *happy*.

- 2) **Context dependency:** Two instances from the same context are more likely to share the same emotion label than those from a random selection.

4 Dependence Factor Graph Model

In this section, we propose a dependence factor graph (DFG) model for learning emotion labels of sentences with both label and context dependence.

4.1 Preliminary

Factor Graph

A factor graph consists of two layers of nodes, i.e., variable nodes and factor nodes, with links between them. The joint distribution over the whole set of variables can be factorized as a product of all factors. Figure 4 gives an example of our dependence factor graph (DFG) when two instances, i.e., *sentence-1* and *sentence-2* are involved.

Binary Relevance

A popular solution to multi-label classification is called binary relevance which constructs a binary classifier for each label, resulting a set of inde-

pendent binary classification problems (Tsoumakas and Katakis, 2007; Tsoumakas et al., 2009). In our approach, binary relevance is utilized as a preliminary step so that each original instance is transformed into K pseudo samples, where K is the number of categories. For example, in Figure 4, X_1^1 , X_1^2 , and X_1^3 represent the three pseudo samples, generated from the same original instance *sentence-1*.

4.2 Model Definition

Formally, let $G = (V, E, X)$ represent an instance network, where V denotes a set of sentence instances. $E \subset V \times V$ is a set of relationships between sentences. Two kinds of relationship exist in our instance network: One represents the label dependence between each two pseudo instances generated from the same original instance, while the other represents the context dependence when the two instances are from the same context, e.g., the same paragraph. X is the textual feature vector associated with a sentence.

We model the above network with a factor graph and our objective is to infer the emotion categories of instances by learning the following joint distribution:

$$P(Y|G) = \prod_k \prod_i f(X_i^k, y_i^k) g(y_i^k, G(y_i^k)) h(y_i^k, H(y_i^k)) \quad (1)$$

where three kinds of factor functions are used.

1) Textual feature factor function: $f(X_i^k, y_i^k)$

denotes the traditional textual feature factor functions associated with each text X_i^k . The textual feature factor function is instantiated as follows:

$$f(X_i^k, y_i^k) = \frac{1}{Z_1} \exp\left(\sum_j \alpha_{ij} \Phi(x_{ij}^k, y_i^k)\right) \quad (2)$$

Where $\Phi(x_{ij}^k, y_i^k)$ is a feature function and x_{ij}^k represents a textual feature, i.e., a word feature in this study.

2) Label dependence factor function:

$g(y_i^k, G(y_i^k))$ denotes the additional label dependence relationship among the pseudo instances, where $G(y_i^k)$ is the label set of the instances connected to y_i^k . $G(y_i^k)$ and y_i^k are labels of the pseudo instances generated from the same original instance. The label dependence factor function is instantiated as follows:

$$g(y_i^k, G(y_i^k)) = \frac{1}{Z_2} \exp\left\{\sum_{y_i^l \in G(y_i^k)} \beta_{ikl} (y_i^k - y_i^l)^2\right\} \quad (3)$$

Where β_{ikl} is the weight of the function, representing the influence degree of the two instances y_i^k and y_i^l .

3) Context dependence factor function:

$h(y_i^k, H(y_i^k))$ denotes the additional context dependence relationship among the instances, where $H(y_i^k)$ is the set of the instances connected to y_i^k . $H(y_i^k)$ and y_i^k are the labels of the pseudo instances from the same context but generated from different original instances. The context dependence factor function is instantiated as follows:

$$h(y_i^k, H(y_i^k)) = \frac{1}{Z_3} \exp\left\{\sum_{y_j^k \in H(y_i^k)} \delta_{ijk} (y_i^k - y_j^k)^2\right\} \quad (4)$$

Where δ_{ijk} is the weight of the function, representing the influence degree of the two instances y_i^k and y_j^k .

4.3 Model Learning

Learning the DFG model is to estimate the best parameter configuration $\theta = (\{\alpha\}, \{\beta\}, \{\delta\})$ to maximize the log-likelihood objective function $L(\theta) = \log P_\theta(Y|G)$, i.e.,

$$\theta^* = \arg \max L(\theta) \quad (5)$$

In this study, we employ the gradient decent method to optimize the objective function. For example, we can write the gradient of each α_{kj} with regard to the objective function:

$$\frac{\partial L(\theta)}{\partial \alpha_{kj}} = E[\Phi(x_{ij}, y_i^k)] - E_{P_{\alpha_{ij}}(Y|G)}[\Phi(x_{ij}, y_i^k)] \quad (6)$$

Where $E[\Phi(x_{ij}, y_i^k)]$ is the expectation of feature function $\Phi(x_{ij}, y_i^k)$ given the data distribution. $E_{P_{\alpha_{ij}}(Y|G)}[\Phi(x_{ij}, y_i^k)]$ is the expectation of feature function $\Phi(x_{ij}, y_i^k)$ under the distribution $P_{\alpha_{ij}}(Y|G)$ given by the estimated model. Figure 5 illustrates the detailed algorithm for learning the parameter α . Note that LBP denotes the Loopy

Belief Propagation (LBP) algorithm which is applied to approximately infer the marginal distribution in a factor graph (Frey and MacKay, 1998). A similar gradient can be derived for the other parameters.

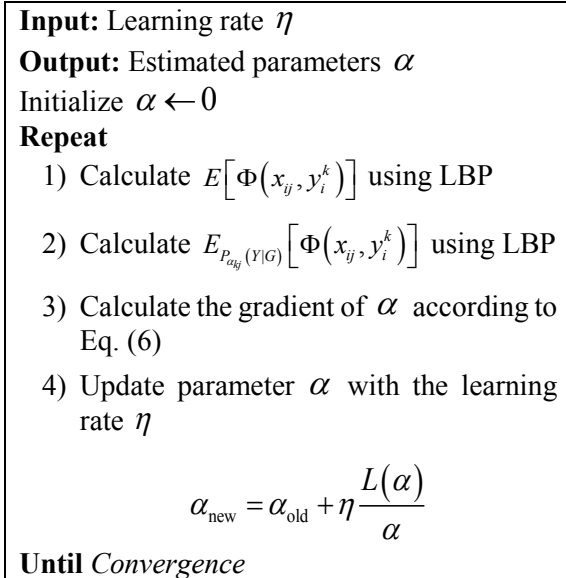


Figure 5: The learning algorithm for DGP model

4.4 Model Prediction

With the learned parameter configuration θ , the prediction task is to find a Y^{U*} which optimizes the objective function, i.e.,

$$Y^{U*} = \arg \max P(Y^U | Y^L, G, \theta) \quad (7)$$

Where Y^{U*} are the labels of the instances in the testing data.

Again, we utilize LBP to calculate the marginal probability of each instance $P(y_i^k | Y^L, G, \theta)$ and predict the label with the largest marginal probability. As all instances in the test data are concerned, above prediction is performed in an iteration process until the results converge.

5 Experimentation

We have systematically evaluated our DFG approach to sentence-level emotion classification.

5.1 Experimental Setting

Corpus

The corpus contains 100 documents (2751 sentences) from the Ren-CECps corpus (Quan and

Ren, 2009). In our experiments, we use 80 documents as the training data and the remaining 20 documents as the test data.

Features

Each instance is treated as a bag-of-words and transformed into a binary vector encoding the presence or absence of word unigrams.

Evaluation Metrics

In our study, we employ three evaluation metrics to measure the performances of different approaches to sentence-level emotion classification. These metrics have been popularly used in some multi-label classification problems (Godbole and Sarawagi, 2004; Schapire and Singer, 2000).

- 1) *Hamming loss*: It evaluates how many times an instance-label pair is misclassified considering the predicted set of labels and the ground truth set of labels, i.e.,

$$hloss = 1 - \frac{1}{mq} \sum_{i=1}^q \sum_{j=1}^m \mathbf{1}_{y_i' = y_i^j} \quad (8)$$

where q is the number of all test instances and m is the number of all emotion labels. y_i^j is the estimated label while y_i^j is the true label.

- 2) *Accuracy*: It gives an average degree of the similarity between the predicted and the ground truth label sets of all test examples, i.e.,

$$Accuracy = \frac{1}{q} \sum_{i=1}^q \left| \frac{y_i \cap y_i'}{y_i \cup y_i'} \right| \quad (9)$$

- 3) *F1-measure*: It is the harmonic mean between precision and recall. It can be calculated from true positives, true negatives, false positive and false negatives based on the predictions and the corresponding actual values, i.e.,

$$F1 = \frac{1}{q} \sum_{i=1}^q \frac{|y_i \cap y_i'|}{|y_i| + |y_i'|} \quad (10)$$

Note that smaller *Hamming loss* corresponds to better classification quality, while larger *accuracy* and *F-measure* corresponds to better classification quality.

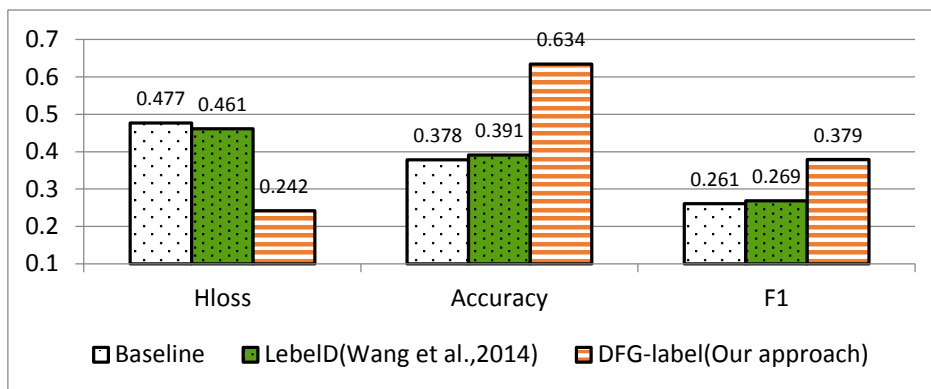


Figure 6: Performance comparison of different approaches to sentence-level emotion classification with the label dependence only

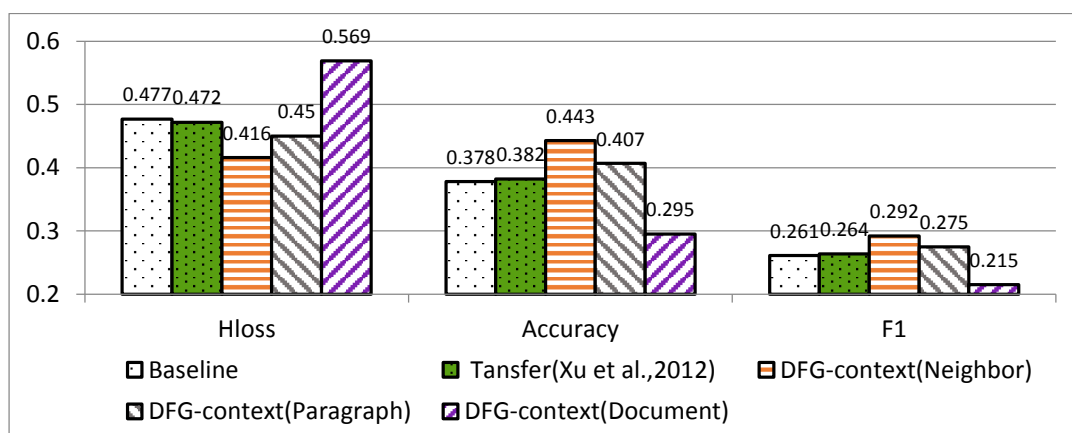


Figure 7: Performance comparison of different approaches to sentence-level emotion classification with the context dependence only

5.2 Experimental Results with Label Dependence

In this section, we compare following approaches which only consider the label dependence among pseudo instances:

- **Baseline:** As a baseline, this approach applies a maximum entropy (ME) classifier with only textual features, ignoring both the label and context dependence.
- **LabelD:** As the state-of-the-art approach to handling multi-label classification, this approach incorporates label dependence, as described in (Wang et al., 2014). Specifically, this approach first utilizes a Bayesian network to infer the relationship among the labels and then employ them in the classifier.
- **DFG-label:** Our DFG approach with the label dependence.

Figure 6 compares the performance of different approaches to sentence-level emotion classification with the label dependence. From this figure, we can see that our DFG approach improves the

baseline approach with an impressive improvement in all three kinds of evaluation metrics, i.e., 23.5% reduction in *Hloss*, 25.6% increase in *Accuracy*, and 11.8% increase in *F1*. This result verifies the effectiveness of incorporating the label dependence in sentence-level emotion classification. Compared to the state-of-the-art LabelD approach, our DFG approach is much superior. Significant test show that our DFG approach significantly outperforms both the baseline approach and LabelD (p -value<0.01). One reason that LabelD performs worse than our approach is possibly due to their separating learning on textual features and label relationships. Also, different from ours, their approach could not capture the information between two conflict emotion labels, such as “happy” and “sad” (they are not possibly appearing together).

5.3 Experimental Results with Context Dependence

In this section, we compare following approaches which only consider the context dependence among pseudo instances:

- **Baseline**: same as the one in Section 5.2, which applies a maximum entropy (ME) classifier with only textual features, ignoring both the label and context dependence.
- **Transfer**: As the state-of-the-art approach to incorporating contextual information in sentence-level emotion classification (Xu et al., 2012), this approach utilizes the label transformation probability to refine the classification results.
- **DFG-label (Neighbor)**: Our DFG approach with the context dependence only. Specifically, the neighboring instances are considered as context.
- **DFG-label (Paragraph)**: Our DFG approach with the context dependence only. Specifically, the instances in the same paragraph are considered as context.
- **DFG-label (Document)**: Our DFG approach with the context dependence only. Specifically, the instances in the same document are considered as context.

Figure 7 compares the performance of different approaches to sentence-level emotion classification with the context dependence only. From this figure, we can see that our DFG approach consistently improves the state-of-the-art in all three kinds of evaluation metrics, i.e., 6.1% reduction in *Hloss*, 6.5% increase in *Accuracy*, and 3.1% increase in *F1* when the neighboring instances are considered as context. Among the three kinds of context, the neighboring setting performs best. We also find that using the whole document as the context is not helpful and it performs even worse than the baseline approach. Compared to the state-of-the-art Transfer approach, our DFG approach with the neighboring context dependence is much superior. Significant test show that our DFG approach with the neighboring context dependence significantly outperforms the baseline approach and the state-of-the-art LabelD approach (p -value<0.01).

5.4 Experimental Results with Both Label and Context Dependence

Table 3 shows the performance of our DFG approach with both label and context dependence, denoted as DGF-both. From this table, we can see that using both label and context dependence further improves the performance.

Figure 8 shows the performance of our DGF-both approach when different sizes of training data are used to train the model. From this figure, we can see that incorporating both the label and context dependence consistently improves the

performance with a large margin, irrespective of the amount of training data available.

Table 3: Performance of our DFG approach with both label and context dependence

	<i>Hloss</i>	<i>Accuracy</i>	<i>F1</i>
Baseline	0.447	0.378	0.261
DFG-label	0.254	0.621	0.372
DFG-context	0.416	0.443	0.292
DFG-both	0.242	0.634	0.379

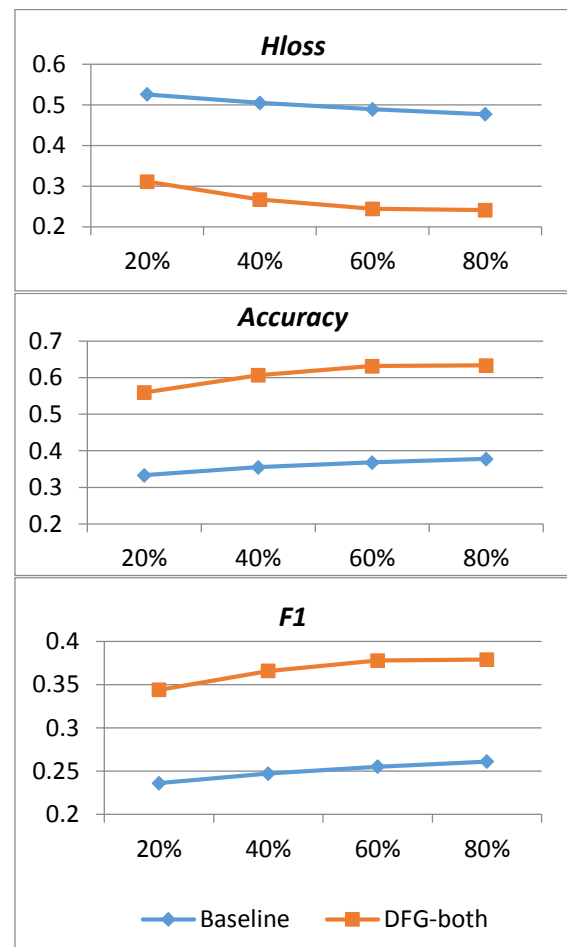


Figure 8: Performance of our DGF-both approach when different sizes of training data are used

6 Conclusion

In this paper, we propose a novel approach to sentence-level emotion classification by incorporating both the label dependence among the emotion labels and the context dependence among the contextual instances into a factor graph, where the label and context dependence is modeled as various factor functions. Empirical evaluation shows that

our DFG approach performs significantly better than the state-of-the-art.

In the future work, we would like to explore better ways of modeling the label and context dependence and apply our DFG approach in more applications, e.g. micro-blogging emotion classification.

Acknowledgments

This research work has been partially supported by three NSFC grants, No.61273320, No.61375073, No.61331011, and Collaborative Innovation Center of Novel Software Technology and Industrialization.

References

- Alm C., D. Roth and R. Sproat. 2005. Emotions from Text: Machine Learning for Text-based Emotion Prediction. In *Proceedings of EMNLP-05*, pp.579-586.
- Bhowmick P., A. Basu, P. Mitra, and A. Prasad. 2009. Multi-label Text Classification Approach for Sentence Level News Emotion Analysis. *Pattern Recognition and Machine Intelligence. Lecture Notes in Computer Science, Volume 5909*, pp 261-266.
- Bhowmick P., A. Basu, P. Mitra, and A. Prasad. 2010. Sentence Level News Emotion Analysis in Fuzzy Multi-label Classification Framework. *Research in Computing Science. Special Issue: Natural Language Processing and its Applications*, pp.143-154.
- Bollen J., H. Mao, and X.-J. Zeng. 2011. Twitter Mood Predicts the Stock Market. *Journal of Computational Science*, 2(1):1-8, 2011.
- Chen Y., S. Lee, S. Li and C. Huang. 2010. Emotion Cause Detection with Linguistic Constructions. In *Proceedings of COLING-10*, pp.179-187.
- Frey B. and D. MacKay. 1998. A Revolution: Belief Propagation in Graphs with Cycles. In *Proceedings of NIPS-98*, pp.479-485.
- Galik M. and S. Rank. 2012. Modelling Emotional Trajectories of Individuals in an Online Chat. In *Proceedings of Springer-Verlag Berlin Heidelberg-12*, pp.96-105.
- Godbole S. and S. Sarawagi. 2004. Discriminative Methods for Multi-labeled Classification. In *Advances in knowledge discovery and data mining*. pp. 22-30.
- Kiritchenko S., X. Zhu, and S. Mohammad. 2014. Sentiment Analysis of Short Informal Texts. *Journal of Artificial Intelligence Research*, 50(2014), pp.723-762.
- Li C., H. Wu, and Q. Jin. 2014. Emotion Classification of Chinese Microblog Text via Fusion of BoW and eVector Feature Representations. In *Proceedings of NLP&CC-14*, pp.217-228.
- Lin K., C. Yang, and H. Chen. 2008. Emotion Classification of Online News Articles from the Reader's Perspective. In *Proceedings of the International Conference on Web Intelligence and Intelligent Agent Technology-08*, pp.220-226.
- Liu H., S. Li, G. Zhou, C. Huang, and P. Li. 2013. Joint Modeling of News Reader's and Comment Writer's Emotions. In *Proceedings of ACL-13*, short paper, pp.511-515.
- Quan C. and F. Ren. 2009. Construction of a Blog Emotion Corpus for Chinese Emotional Expression Analysis. In *Proceedings of EMNLP-09*, pp.1446-1454.
- Schapire R. E and Y. Singer. 2000. A Boosting-based System for Text Categorization. *Machine learning*, pp. 135-168
- TOKUHISA R., K. Inui, and Y. Matsumoto. 2008. Emotion Classification Using Massive Examples Extracted from the Web. In *Proceedings of COLING-2008*, pp.881-888.
- Tsoumakas G. and I. Katakis. 2007. Multi-label Classification: An Overview. In *Proceedings of International Journal of Data Warehousing and Mining*, 3(3), pp.1-13.
- Tsoumakas G., I. Katakis, and I. Vlahavas. 2009. Mining Multi-label Data. *Data Mining and Knowledge Discovery Handbook*, pages 1-19.
- Wen S. and X. Wan. 2014. Emotion Classification in Microblog Texts Using Class Sequential Rules. In *Proceedings of AAAI-14*, 187-193.
- Wang S., J. Wang, Z. Wang, and Q. Ji. 2014. Enhancing Multi-label Classification by Modeling Dependencies among Labels. *Pattern Recognition*. Vol. 47. Issue 10: 3405-3413, 2014.
- Wiebe J., T. Wilson, and C. Cardie. 2005. Annotating Expressions of Opinions and Emotions in Language. *Language Resources and Evaluation*, 39, 65-210.
- Xu G., X. Meng and H. Wang. 2010. Build Chinese Emotion Lexicons Using A Graph-based Algorithm and Multiple Resources. In *Proceedings of COLING-10*, pp.1209-1217.
- Xu J., R. Xu, Q. Lu, and X. Wang. 2012. Coarse-to-fine Sentence-level Emotion Classification based on the Intra-sentence Features and Sentential Context. In *Proceedings of CIKM-12*, poster, pp.2455-2458.