Relevant Emotion Ranking from Text Constrained with Emotion Relationships

Proceedings of NAACL-HLT 2018

作者

Deyu Zhou, Yang Yang, Yulan He

School of Computer Science and Engineering, Key Laboratory of Computer Network and Information Integration, Ministry of Education, Southeast University, China School of Engineering and Applied Science, Aston University, UK

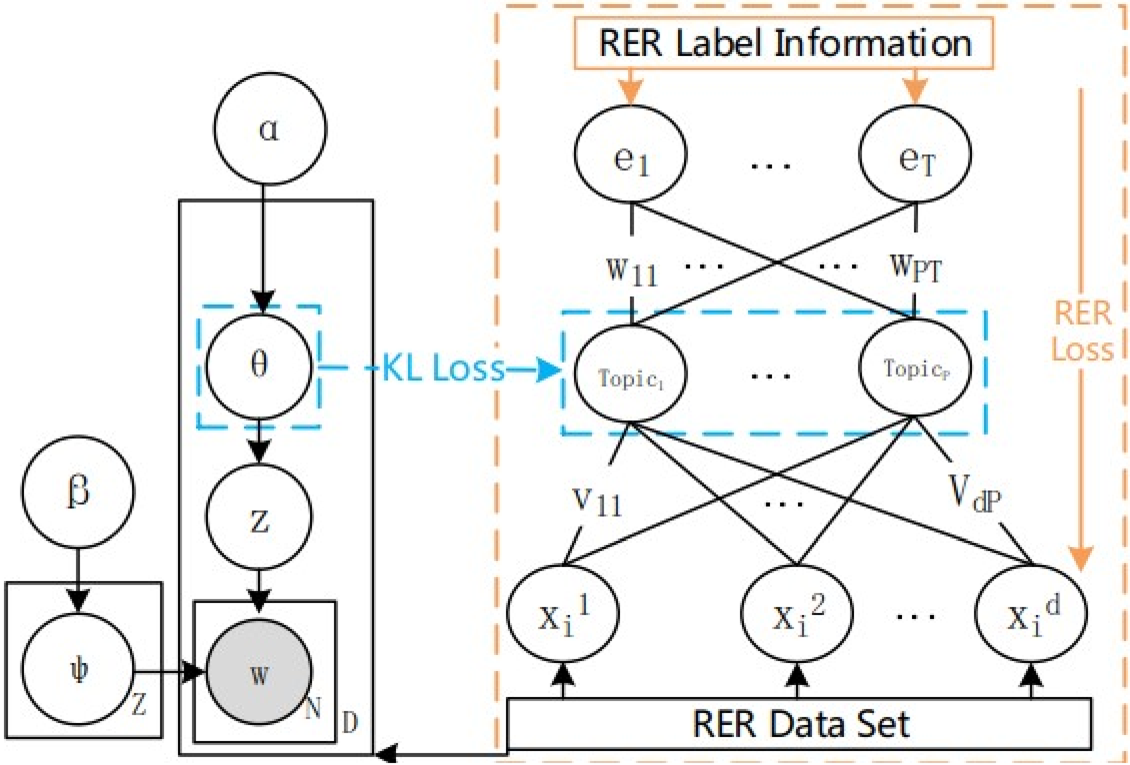
Outline

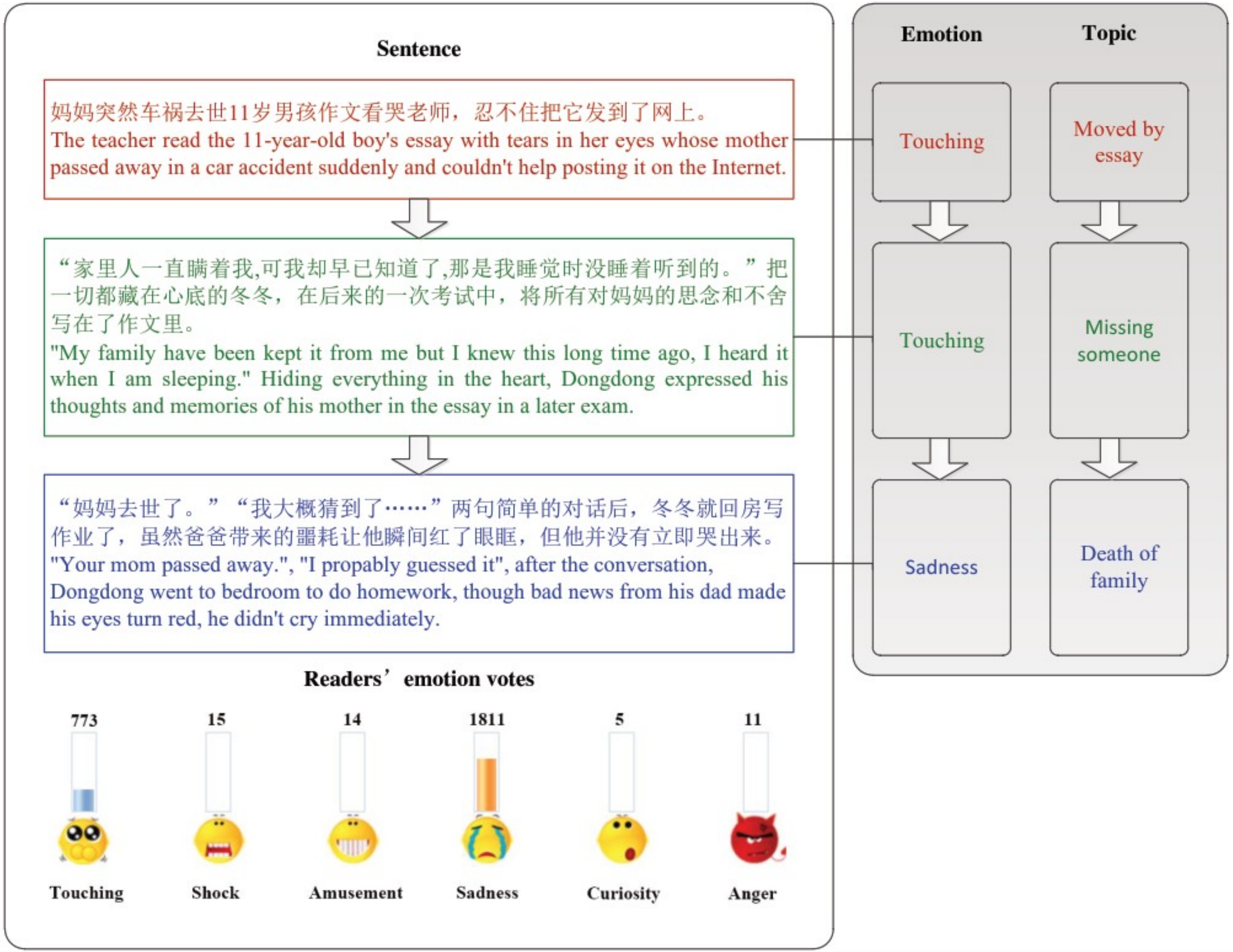
- Emotion Representations
- Multi-Emotion Detection
- Interpretable Neural Network for Emotion Ranking
- Hidden Topic-Emotion Transition Model
- Emotion Cause Extraction

1~38~77

Interpretable Neural Network for Relevant Emotion Ranking (INN-RER)

(Yang, Zhou and He, EMNLP 2018)





Sentence
妈妈突然车祸去世11岁男孩作文看哭老师，忍不住把它发到了网上。
The teacher read the 11-year-old boy's essay with tears in her eyes whose mother passed away in a car accident suddenly and couldn't help posting it on the Internet.
“家里人一直瞒着我,可我却早已知道了,那是我睡觉时没睡着听到的。”把一切都藏在心底的冬冬，在后来的一次考试中，将所有对妈妈的思念和不舍写在了作文里。
"My family have been kept it from me but I knew this long time ago, I heard it when I am sleeping." Hiding everything in the heart, Dongdong expressed his thoughts and memories of his mother in the essay in a later exam.
“妈妈去世了。”“我大概猜到了……”两句简单的对话后，冬冬就回房写作业了，虽然爸爸带来的噩耗让他瞬间红了眼眶，但他并没有立即哭出来。
"Your mom passed away.", "I propably guessed it", after the conversation, Dongdong went to bedroom to do homework, though bad news from his dad made his eyes turn red, he didn't cry immediately.
Readers’ emotion votes
773
15
14
1811
5
11
Touching
Shock
Amusement
Sadness
Curiosity
Anger
Emotion
Topic
Touching
Moved by essay
Touching
Missing someone
Sadness
Death of family

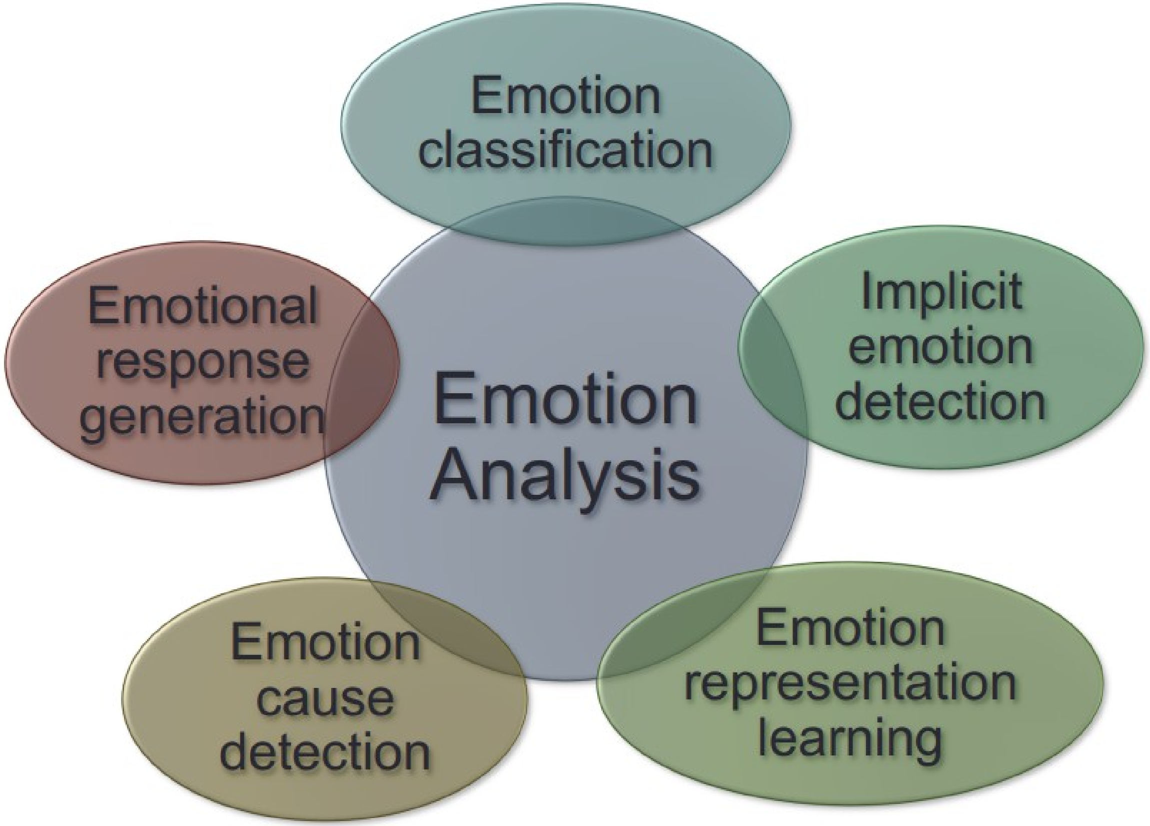
1 Introduction

- https://drive.google.com/file/d/1wS3P042lyHOCYzwbsMlAGaNADtXbGCdT/view?usp=sharing
- https://drive.google.com/file/d/17MVO_-YgJnWfVlzwz1ygaNv9FKKi43wd/view?usp=sharing

Emotion Representations

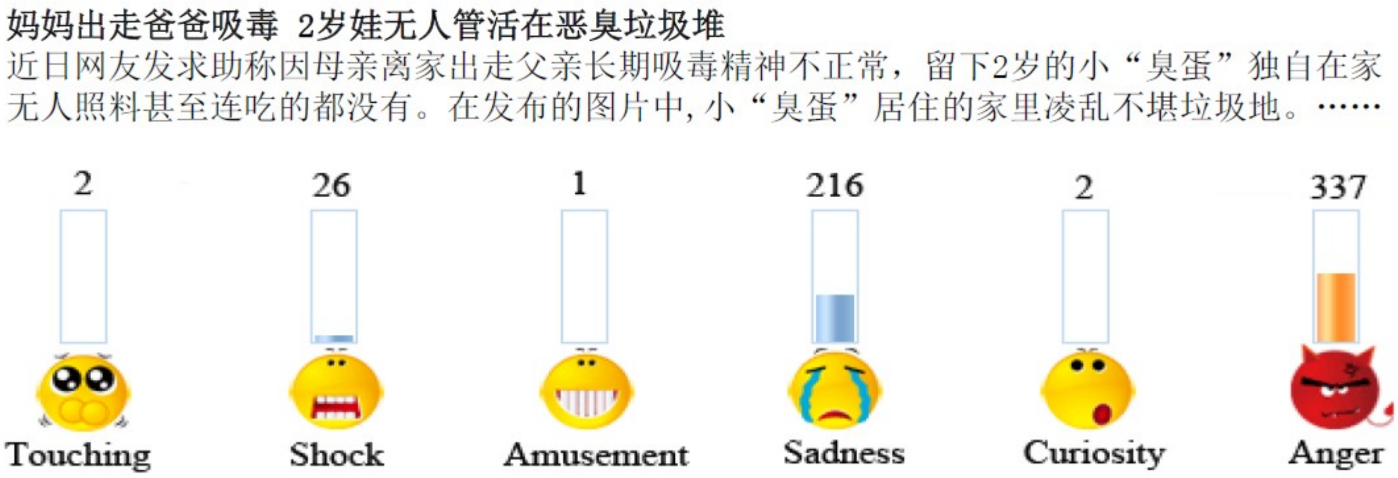
- Sentiment
  - Positive, negative, neutral
- Multi-category emotion representations
- Multi-dimensional emotion representations

Emotion Analysis Tasks



Multi-Emotion Detection

Sina News



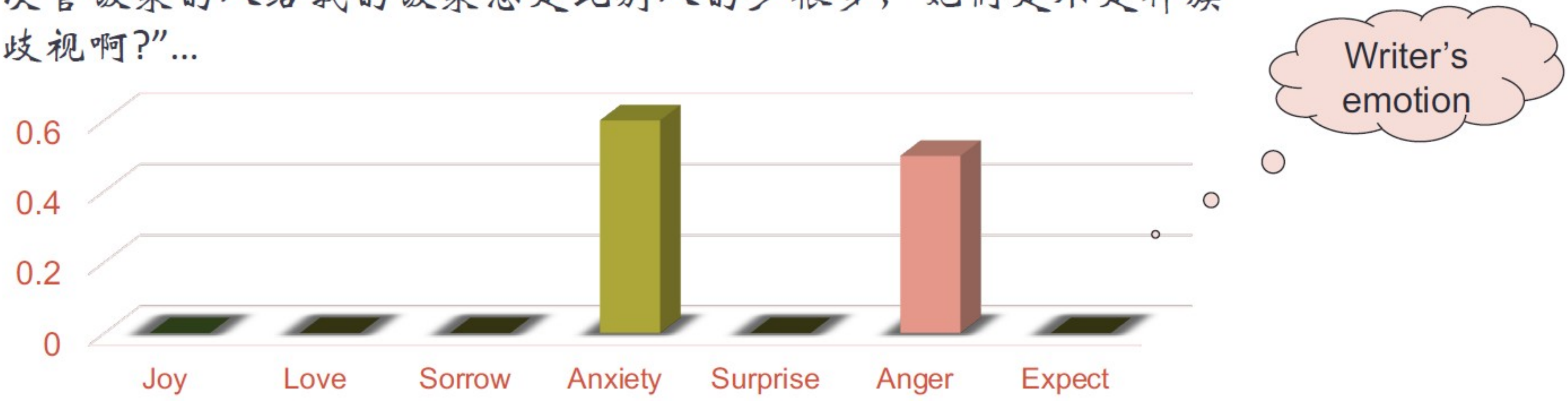
Readers' reactions

- Readers expressed different emotions with the majority showed “Sadness” and “Anger”
- Emotions receiving very few votes could be considered as *irrelevant ones*

Chinese Blogs

女儿在学校遭遇不平

昨天下午，女儿一进门就对我说，妈妈，我有重要的事跟你说:“每次管饭菜的人给我的饭菜总是比别人的少很多，她们是不是种族歧视啊?”...



2 Related work-Lexicon-based

- Aman and Szpakowicz (2007) classified emotional and non-emotional sentences with a predefined emotion lexicon.
    - Emotional dictionaries constructed from training corpora of news articles were used to predict the readers’ emotion of a new articles (Lei et al., 2014; Rao et al., 2012)
    - Use linguistic templates to predict reader’s emotions (Chang et al., 2015)
- emotion vector
- (Wang and Pal, 2015) can also output multiple emotions with intensities using non-negative matrix factorization with constraints derived based on an emotion lexicon

2 Related work-Learning-based

- Unsupervised
  - emotion-topic models (Bao et al., 2012, 2009)
  - Latent Dirichlet Allocation (LDA) (Blei et al., 2003)
- Supervised
  - classification problem
    - Single-label
      - knowledge-based and corpus-based
    - Multi-label
      - Quan et al. (2015) proposed a logistic regression model
  - A joint model to co-train a polarity classifier and an emotion classifier (Gao et al., 2013)
  - A multi-task Gaussian-process based classification (Beck et al., 2014)
  - Emotion distribution learning (Zhou et al., 2016)

3 Methodology

- Assuming a set of T emotions $E = \{e_1, e_2, ...e_T\}$
- a set of n instances $X = \{x_1, x_2, x_3, ..., x_n\}$

$$x_i \in \mathbb{R}^d$$

$$R_i \subseteq E$$

$$\overline{R_i} = E - R_i$$

$$\mathbf{g}(x_i) = [g_1(x_i), ..., g_T(x_i)]$$

  - assigning a score $g_t(x_i)$ to each emotion $e_t, (t \in \{1, ..., T\})$
  - multi-label learning
  - Relevant emotion ranking

- Relevant emotion ranking

	emotions				
instances	e1	e2	…	et	e threshold
X1	g1(x1)	g2(x1)	…	gt(x1)	eth(x1)
X2	g1(x2)	g2(x2)	…	gt(x2)	eth(x2)
…	…	…	…	…	…
Xn	g1(xn)	g2(xn)	…	gt(xn)	eth(xn)

3 Methodology

- define a threshold $g_\Theta(x)$ which could be simply set to 0 or learned from data (Furnkranz et al., 2008)
- the predicted relevant emotions of instance x_i are denoted as
$\hat{R}_i = \{e_t \in E | g_t(x_i) > g_\Theta(x_i)\}$
    - Those emotions with scores lower than the threshold will be considered as irrelevant and hence discarded.

3 Methodology-Emotion Loss Function

- The goal of relevant emotion ranking is to learn the parameter of the ranking function g.
- Without loss of generality, we assume that g are linear models

$$g_t(x_i) = w_t^\intercal \cdot x_i,\ t \in \{1, 2, 3, ..., T\} \cup \{\Theta\}$$

3 Methodology-Emotion Loss Function

- Several evaluation criteria typically used in multi-label learning can also be used to measure the ranking function's ability of distinguishing relevant emotions from irrelevant ones, such as hamming loss, one error, coverage, ranking loss, and average precision as suggested in (Zhang and Zhou, 2014).
- However, these multi-label criteria cannot meet our requirement exactly as none of them considers the ranking among emotions which are considered relevant.

3 Methodology-Emotion Loss Function

- Therefore, by incorporating PRO loss (Xu et al., 2013), the loss function for the instance x_i is defined as follows:

$$L(x_i, R_i, \prec, \mathbf{g}) = \sum_{e_t \in R_i \cup \{\Theta\}} \sum_{e_s \in \prec(e_t)} \frac{1}{norm_{t,s}} l_{t,s} \tag{1}$$

$$l_{t,s} = \begin{cases} 1, & g_t(x_i) < g_s(x_i) \\ \frac{1}{2}, & g_t(x_i) = g_s(x_i) \\ 0, & \text{otherwise} \end{cases}$$

>=

3 Methodology-Emotion Loss Function

(e_s, e_t) represents 4 types of emotion pairs:

- (relevant, relevant)
- (relevant, irrelevant)
- (relevant, threshold)
- (irrelevant, threshold)

3 Emotion Loss Function

where e_t refers to the emotion belonging to relevant emotion set R_i or the threshold Θ of instance x_i while e_s refers to the emotion which is less relevant than e_t denoted as $\prec$. Thus, (e_t, e_s) represents four types of emotion pairs: i.e., *(relevant, relevant)*, *(relevant, irrelevant)*, *(relevant, threshold)*, and *(threshold, irrelevant)*. The normalization term $norm_{t,s}$ is used to balance those four types of emotion pairs to avoid dominated terms by their respective set sizes. The set sizes of the four different types of emotion pairs mentioned

3 Methodology-Emotion Loss Function

- Note that $l_{t,s}$ is non-convex and difficult to optimize. Thus, a large margin surrogate convex loss (Vapnik and Vapnik, 1998) implemented in hinge form is used instead as follows:

$$\hat{L}(x_i, R_i, \prec, \mathbf{g}) = \sum_{e_t \in R_i \cup \{\Theta\}} \sum_{e_s \in \prec(e_t)} \frac{1}{norm_{t,s}} (1 + g_s(x_i) - g_t(x_i))_+$$

where $(u)_+ = \max\{0, u\}$.

情緒有差異 > 0 + 整數

無情緒差異 <= 0 - 負數

3 Methodology-Emotion Loss Function

However, Eq. 2 ignores the relationships between different emotions. As mentioned in Introduction section, some emotions often co-occur such as “joy” and “love” while some rarely co-exist such as “joy” and “anger”. Such relationship information among emotions can provide important clues for emotion ranking. Therefore, we incorporate this information into the emotion loss function as constraints.

3 Methodology-Emotion Loss Function

- $\hat{L}(x_i, R_i, \prec, \mathbf{g})$ can be redefined as:

$$\hat{L_\omega}(x_i, R_i, \prec, \mathbf{g}) = \sum_{e_t \in R_i \cup \{\Theta\}} \sum_{e_s \in \prec(e_t)} \frac{1}{norm_{t,s}} \times (1 + g_s(x_i) - g_t(x_i) + \omega_{ts}(w_t - w_s))_+$$

3 Methodology-Emotion Loss Function

- where the weight ω_{ts} models the relationship between the t-th emotion and the s-th emotion in the emotion set and can be calculated in multiple ways. Since the Pearson correlation coefficient (Nicewander, 1988) is the most familiar measure of relationship between two variables, we use it to measure the relationship of two emotions using their original emotion scores across each corpus.

3.2 Relevant Emotion Ranking

- minimize the empirical error measured by the appropriate loss function
- control the complexity of the resulting model

3.2 Relevant Emotion Ranking

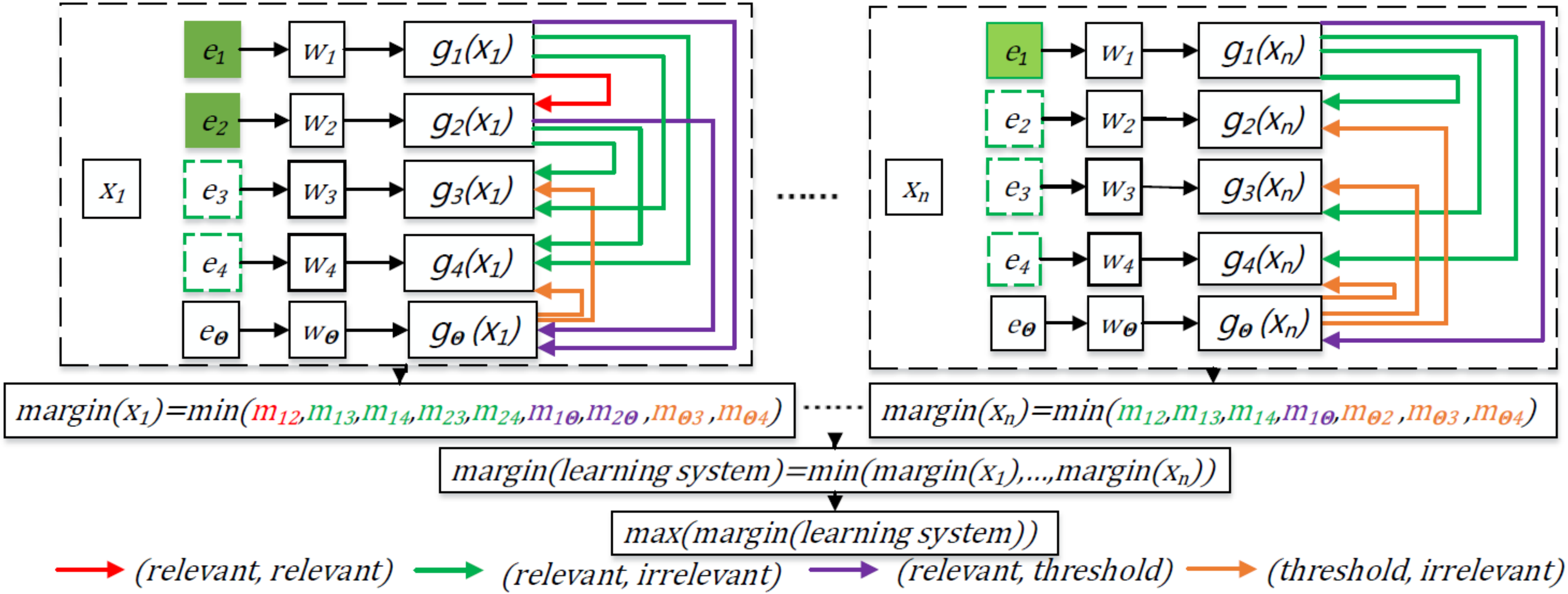
- a label ranking method (Elisseeff andWeston, 2001)
    - The margin of the (relevant, relevant) label pair needs to be dealt with carefully

3.2 Relevant Emotion Ranking

- relevant emotion ranking (RER)

$$\min_{e_t \in R_i \cup \{\Theta\}, e_s \in \prec(e_t)} \frac{\langle w_t - w_s, x_i \rangle}{||w_t - w_s||} \quad (4)$$

Various Margins-(RER)



幅度 v.s. 相關
w↑ m↓ r↑

3.2 Relevant Emotion Ranking

- the hyperplane $\langle w_t - w_s, x_i \rangle = 0$.

$$\min_{x_i \in G} \min_{e_t \in R_i \cup \{\Theta\}, e_s \in \prec(e_t)} \frac{\langle w_t - w_s, x_i \rangle}{||w_t - w_s||}$$

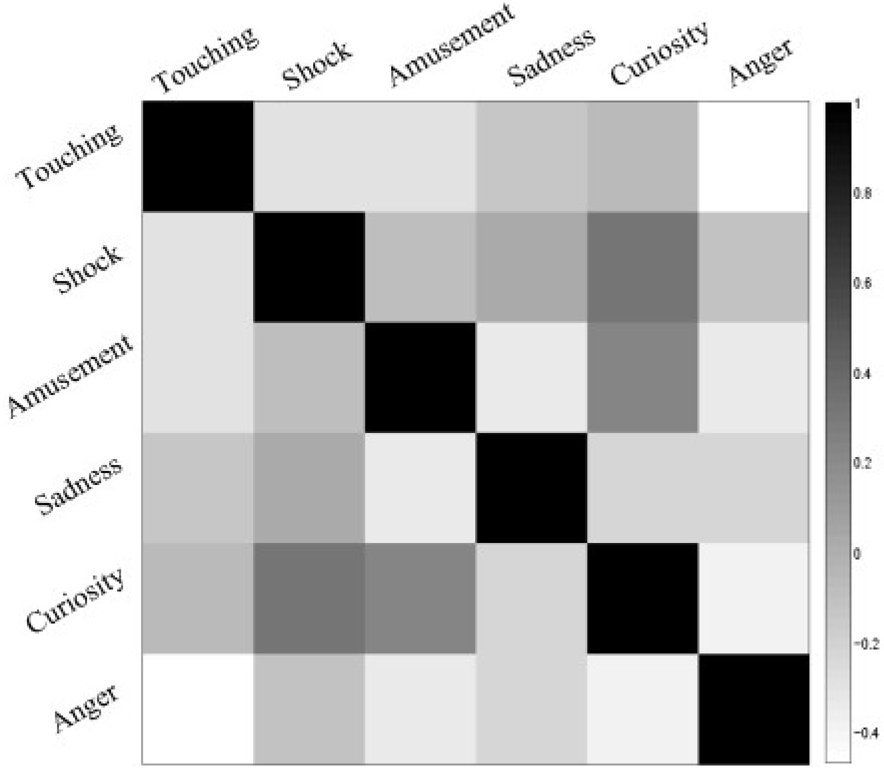
3.2 Relevant Emotion Ranking

- a positive margin

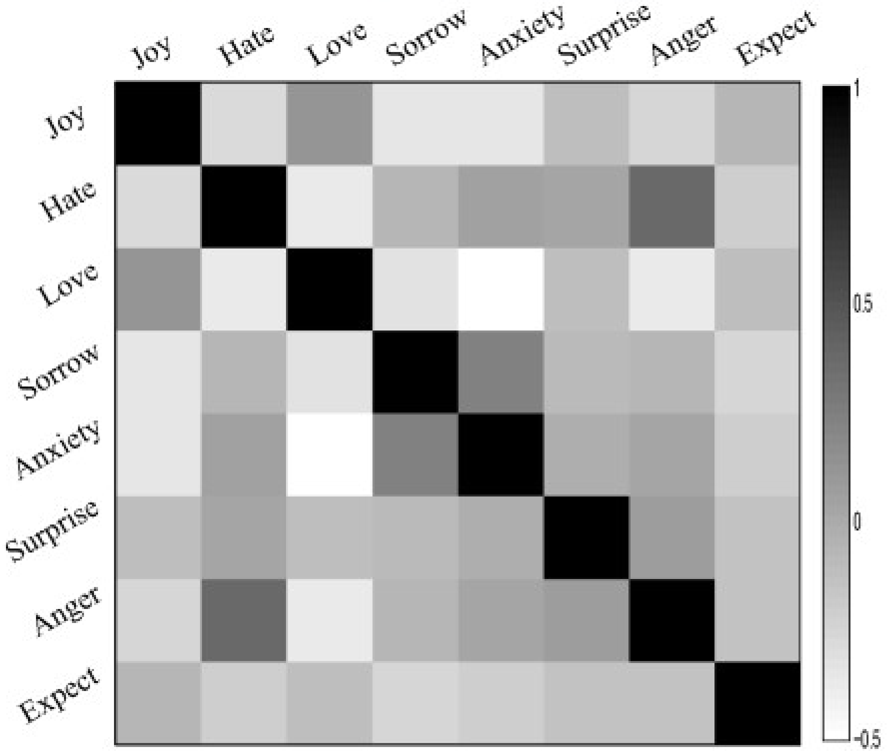
$$\max_{w_j} \min_{x_i \in G} \min_{e_t \in R_i \cup \{\Theta\}, e_s \in \prec(e_t)} \frac{1}{||w_t - w_s||}$$

$$s.t. \langle w_t - w_s, x_i \rangle \geq 1, 1 \leq i \leq n, 1 \leq j \leq T + 1$$

Emotion Relationships - Sina News



Emotion Relationships – Blogs



3.2 Relevant Emotion Ranking

- to overcome the complexity brought in by the max operator, the objective of the optimization problem can be re-written by approximating the max operator with the sum operator.

$$
\begin{aligned}
&\min_{w_j} \sum_{t=1}^{T+1} ||w_t||^2 \\
&s.t.\ \langle w_t - w_s, x_i \rangle \geq 1, 1 \leq i \leq n, \qquad (7) \\
&1 \leq j \leq T+1, e_t \in R_i \cup \{\Theta\}, e_s \in \prec (e_t)
\end{aligned}
$$

3.2 Relevant Emotion Ranking

- To accommodate real-world scenarios
- $$\min_{w_j,\xi_{its}} \sum_{t=1}^{T+1} ||w_t||^2 + \lambda \sum_{i=1}^{n} \sum_{e_t \in R_i \cup \{\Theta\}} \sum_{e_s \in \prec(e_t)} \frac{1}{norm_{t,s}} \xi_{its}$$
  $$s.t. \langle w_t - w_s, x_i \rangle \geq 1 - \xi_{its}, 1 \leq j \leq T+1, \xi_{its} \geq 0$$

3.2 Relevant Emotion Ranking

- Since ξ_{its} does not need to be optimized since it can be easily determined by wt, ws.
- two parts balanced by the trade-off parameter λ.

$$\min_{w_t, \hat{L}} \sum_{t=1}^{T+1} ||w_t||^2 + \lambda \sum_{i=1}^{n} \hat{L}(x_i, R_i, \prec, \mathbf{g})$$

3.3 Parameter Estimation

- SVM-type
- $$\min_{\mathbf{w},\boldsymbol{\xi}} \frac{1}{2}||\mathbf{w}||^2 + \lambda C^\top \boldsymbol{\xi}$$
  $$s.t.\ \ A\mathbf{w} \geq \mathbf{1}_p - \boldsymbol{\xi}, \boldsymbol{\xi} \geq \mathbf{0}_p$$
- ξ does not need to be optimized since it can be easily determined by w.

3.3 Parameter Estimation

- Alternating Direction Method of Multipliers (ADMM)
- (Bertsekas and Tsitsiklis, 1989)

$$\min_{\mathbf{w}} F(\mathbf{w}, G) = \frac{1}{2}||\mathbf{w}||^2 + \lambda C^\top (1_p - A\mathbf{w})_+$$

3.3 Parameter Estimation

- ADMM

$$\min_{\mathbf{w}^0, \mathbf{w}^1, \mathbf{w}^m} \sum_{m=1}^{M} F(\mathbf{w}^m, G^m),$$

$$s.t.\ \ \mathbf{w}^m = \mathbf{w}^0, \forall m = 1, ..., M$$

3.3 Parameter Estimation

- Lagrangian Function (LF)
- 數學最優化問題中，拉格朗日乘數法（以數學家約瑟夫•拉格朗日命名）是一種尋找變數受一個或多個條件所限制的多元函數的極值的方法。

$$LF(\{\mathbf{w}^0, \mathbf{w}^1, ..., \mathbf{w}^m\}, \{\alpha^m\}_{m=1}^{M}, \beta) =$$

$$\sum_{m=1}^{M} F(\mathbf{w}^m, G^m) \quad + \sum_{m=1}^{M} (\alpha^m)^\top (\mathbf{w}^m - \mathbf{w}^0) + \frac{\beta}{2} \sum_{m=1}^{M} ||\mathbf{w}^m - \mathbf{w}^0||^2$$

3.3 Parameter Estimation

Algorithm 1 Parameter updating process.

1: Decompose data set G into M disjoint subsets i.e., $\{G_1, G_2, ..., G_M\}$. Set iteration $i = 0$.
2: Initialize $\{\mathbf{w}_0^0, \mathbf{w}_0^1, ..., \mathbf{w}_0^M, \alpha_0^1, ..., \alpha_0^M\}$ as zeros.
3: **while** not converged **do**
4: Set $i = i + 1$
5: Update $\mathbf{w}_i^0, \{\mathbf{w}_i^m, \alpha_i^m\}_{m=1}^M$ as:
$\{\mathbf{w}_i^m\}_{m=1}^M = \underset{\mathbf{w}^1..\mathbf{w}^m}{\operatorname{argmin}} LF(\mathbf{w}_{i-1}^0, \{\mathbf{w}_{i-1}^m, \alpha_{i-1}^m\}_{m=1}^M, \beta)$
$\mathbf{w}_i^0 = \underset{\mathbf{w}^0}{\operatorname{argmin}} LF(\mathbf{w}^0, \{\mathbf{w}_{i-1}^m, \alpha_{i-1}^m\}_{m=1}^M, \beta$
$\alpha_i^m = \alpha_{i-1}^m + \beta(\mathbf{w}_i^m - \mathbf{w}_i^0)^\top, \forall m = 1, 2, ..., M$
6: **end while**
Output: Final $\mathbf{w}^0$

4 Experiments

4.1 Setup

Sina Social News		Ren-CECps Corpus	
Category	#Votes	Category	#Scores
Touching	694,006	Joy	1,349.6
Shock	572,651	Hate	6,103.9
Amusement	869,464	Love	2,911.1
Sadness	837,431	Sorrow	2,042.5
Curiosity	212,559	Surprise	3,873.9
Anger	1,109,315	Anger	7,832.1
		Anxiety	5,006.4
		Expect	610.4
All	4,295,426	All	29,729.9

Table 1: Statistics for the two corpora used in our experiments.

4.1 Setup

- the single layer long short-term memory (LSTM) networks (Hochreiter and Schmidhuber, 1997)
- the dimension of each text representation to 100
- a learning rate of 0.001
- a dropout rate of 0.3
- categorical cross-entropy as the loss function
- The mini batch (Cotter et al., 2011) size is set to 32.

4.2 Comparison with Baselines

- previously achieved the state-of-the-art performances on multi-emotion detection
- Emotion Distribution Learning (EDL) (Zhou et al., 2016)
- EmoDetect (Wang and Pal, 2015)

4.2 Comparison with Baselines

- Θ which is initialized as 0.15 after normalization
- RERc
- ML-KNN (Zhang and Zhou, 2007),
- LIFT (Zhang, 2011)
- CLR (F¨urnkranz et al., 2008)
- Rank-SVM (Zhang and Zhou, 2014) ,
- MLLOC(Huang and Zhou, 2012),
- BP-MLL (Zhang and Zhou, 2006),
- ECC (Read et al., 2009)
- ML-RBF(Zhang, 2009).

Name	**Definition**
PRO Loss	$\frac{1}{n}\sum_{i=1}^{n}\sum\limits_{e_t\in R_i\cup\{\Theta\}}\sum\limits_{e_s\in\prec(e_t)}\frac{1}{norm_{t,s}}l_{t,s}$ $l_{t,s}$ is a modified 0-1 error; $norm_{t,s}$ is the set size of label pair (t, s)
Hamming Loss	$\frac{1}{nT}\sum_{i=1}^{n}\lvert\hat{R}_i\triangle R_i\rvert$
Ranking Loss	$\frac{1}{n}\sum_{i=1}^{n}(\sum_{(e_t,e_s)\in R_i\times\overline{R_i}}\delta[g_t(x_i)<g_s(x_i)])/(\lvert R_i\rvert\times\lvert\overline{R_i}\rvert)$ where δ is the indicator function.
One Error	$\frac{1}{n}\sum_{i=1}^{n}\delta[\underset{\mathbf{e_t}}{\operatorname{argmax}}\, g_t(x_i)\notin R_i]$
Average Precision	$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lvert R_i\rvert}\times$ $(\sum\limits_{t:e_t\in R_i}\lvert\{e_s\in R_i\lvert g_s(x_i)>g_t(x_i)\}\rvert)/(\lvert\{e_s\lvert g_s(x_i)>g_t(x_i)\}\rvert)$
Coverage	$\frac{1}{n}\sum_{i=1}^{n}\max_{t:e_t\in R_i}\lvert\{e_s\lvert g_s(x_i)>g_t(x_i)\}\rvert$
Subset Accuracy	$\frac{1}{n}\sum_{i=1}^{n}\delta[\hat{R}_i=R_i]$
$F1_{exam}$	$\frac{1}{n}\sum_{i=1}^{n}2\lvert R_i\cap\hat{R}_i\rvert/(\lvert R_i\rvert+\lvert\hat{R}_i\rvert)$
MicroF1	$F1(\sum\limits_{t=1}^{T}TP_t,\sum\limits_{t=1}^{T}FP_t,\sum\limits_{t=1}^{T}TN_t,\sum\limits_{t=1}^{T}FN_t)$
MacroF1	$\frac{1}{T}\sum\limits_{t=1}^{T}F1(TP_t,FP_t,TN_t,FN_t)$

Datasets	Evaluation Criterion	Methods			
		RER	RERc	EDL	EmoDetect
News	PRO loss(↓)	0.1992	**0.1913**	0.2596	0.2465
	Hamming Loss(↓)	0.2318	**0.2277**	0.2671	0.2696
	Ranking Loss(↓)	0.1477	**0.1405**	0.1689	0.1769
	One-error(↓)	0.1579	**0.1562**	0.2115	0.1903
	Average Precision(↑)	0.8775	**0.8789**	0.8028	0.7865
	Coverage(↓)	2.1398	**2.1316**	2.1595	2.2348
	Subset Accuracy(↓)	0.1899	**0.1822**	0.2026	0.2243
	$F1_{exam}$(↑)	0.7062	**0.7143**	0.6503	0.6469
	MicroF1(↑)	0.7086	**0.7171**	0.6346	0.6375
	MacroF1(↑)	0.6244	**0.6291**	0.5641	0.5767
Blogs	PRO loss(↓)	0.2354	**0.2321**	0.2739	0.2912
	Hamming Loss(↓)	0.2054	**0.2014**	0.2102	0.2202
	Ranking Loss(↓)	0.2137	**0.2102**	0.2589	0.2781
	One-error(↓)	0.4556	**0.4550**	0.5227	0.5352
	Average Precision(↑)	0.6749	**0.6803**	0.6411	0.5663
	Coverage(↓)	2.1269	**2.1268**	2.1699	2.8956
	Subset Accuracy(↓)	**0.1663**	**0.1663**	0.2116	0.2321
	$F1_{exam}$(↑)	0.5080	**0.5114**	0.4606	0.4650
	MicroF1(↑)	0.5093	**0.5116**	0.4620	0.4552
	MacroF1(↑)	0.4102	**0.4161**	0.3923	0.3622

Anger	Anxiety	Expect	Hate
生气(angry)	害怕(fear)	祝福(blessing)	讨厌(hate)
愤怒(rage)	失去(lose)	幸福(happy)	虚伪(hypocrisy)
抱怨(complain)	孤独(lonely)	美好(fine)	炒作(hype)
批评(criticize)	压力(pressure)	梦想(dream)	无耻(shameless)
利益(interest)	现实(reality)	自由(freedom)	手段(means)
歧视(discriminate)	陌生(strange)	渴望(long for)	愚蠢(silly)
制止(stop)	心灵(heart)	希望(hope)	浪费(waste)
指责(accuse)	痛苦(pain)	学习(learn）	背后(behind)
懊恼(annoy)	想象(imagine)	信念(faith)	肮脏(dirty)
无耻(shameless)	伤害(hurt)	家里(home)	欺骗(lie)
Joy	**Love**	**Sorrow**	**Surprised**
快乐(happy)	美丽(beautiful)	孤独(lonely)	好奇(curious)
高兴(joyful)	爱情(love)	眼泪(tears)	惊讶(surprise)
朋友(friend)	朋友(friend)	爱情(love)	震惊(shock)
感动(touching)	幸福(happiness)	寂寞(solitude)	惊奇(wonder)
心情(mood)	孩子(child)	痛苦(pain)	惊人(amazing)
温暖(warm)	生命(life)	感情(feeling)	意外(accident)
享受(enjoy)	阳光(sunshine)	伤害(hurt)	惊吓(fright)
兴奋(excited)	温暖(warmth)	失去(lose)	惊呼(scream)
收获(harvest)	思念(miss)	思念(miss)	不经意(accidently)
微笑(smile)	可爱(lovely)	生活(life)	诧异(amazed)

Datasets	**Evaluation Criterion**	**Methods**								
		RERc	ML-KNN	LIFT	CLR	Rank-SVM	MLLOC	BP-MLL	ECC	ML-RBF
News	PRO loss(↓)	**0.1913**	0.2551	0.2426	0.2487	0.2670	0.3429	0.2603	0.2823	0.2658
	Hamming Loss(↓)	**0.2277**	0.2876	0.3118	0.3023	0.3127	0.3241	0.3040	0.3079	0.3599
	Ranking Loss(↓)	**0.1405**	0.1898	0.1987	0.2142	0.2271	0.3234	0.1897	0.2563	0.1949
	One-error(↓)	**0.1562**	0.2366	0.1881	0.2242	0.2258	0.2025	0.2043	0.2151	0.2240
	Average Precision(↑)	**0.8789**	0.8095	0.7945	0.7916	0.8001	0.7545	0.8044	0.6245	0.8106
	Coverage(↓)	**2.1316**	2.3602	2.4641	2.3453	2.6093	3.1272	2.4032	2.4122	2.4390
	Subset Accuracy(↓)	**0.1822**	0.1916	0.1857	0.2386	0.1839	0.2107	0.2765	0.2222	0.2609
	$F1_{exam}$(↑)	**0.7143**	0.6215	0.6262	0.6032	0.6244	0.5193	0.5879	0.5108	0.6147
	MicroF1(↑)	**0.7171**	0.6280	0.6131	0.6177	0.6268	0.5389	0.6231	0.5699	0.6160
	MacroF1(↑)	**0.6291**	0.5587	0.5593	0.5658	0.5613	0.4913	0.5563	0.4573	0.5543
Blogs	PRO loss(↓)	**0.2321**	0.3036	0.2912	0.3041	0.2869	0.3523	0.3429	0.2867	0.2922
	Hamming Loss(↓)	**0.2014**	0.2409	0.2242	0.2162	0.2585	0.2156	0.2241	0.2301	0.2204
	Ranking Loss(↓)	**0.2102**	0.2928	0.2881	0.2947	0.3024	0.4532	0.3234	0.3345	0.2364
	One-error(↓)	**0.4550**	0.5543	0.5152	0.5229	0.5606	0.6143	0.4625	0.6635	0.4679
	Average Precision(↑)	**0.6803**	0.5897	0.5963	0.6370	0.5832	0.4532	0.5545	0.5256	0.6412
	Coverage(↓)	**2.1268**	2.4448	2.4356	2.2671	2.5962	3.5634	3.1272	2.7756	2.5067
	Subset Accuracy(↓)	**0.1663**	0.1978	0.2116	0.1938	0.2321	0.2251	0.2107	0.2236	0.1803
	$F1_{exam}$(↑)	**0.5114**	0.4616	0.4620	0.4509	0.4832	0.4931	0.5093	0.4986	0.4997
	MicroF1(↑)	**0.5116**	0.4720	0.4552	0.4859	0.4962	0.4902	0.4889	0.5003	0.5051
	MacroF1(↑)	**0.4161**	0.3632	0.3656	0.4056	0.3965	0.3853	0.3813	0.3957	0.4086

4.3 Result Analysis

- the word “happy” delivers the emotion of Joy
- the word “tears” tells Sorrow
- “friend” appears in both Joy and Love

5 Conclusions

- a novel framework based on relevant emotion ranking to identify multiple emotions
  - the rankings of relevant emotions from text
- emotion relationship can provide important cues for emotion detection
  - we observe that some emotions co-occur more often while other emotions rarely coexist.
- the proposed framework can effectively deal with the emotion detection problem
  - better than the state-of-the-art emotion detection methods and multi-label learning methods.

Goal

- Emotions might be evoked by hidden topics
  - Unveil the topical information to understand how the emotions are evoked.
- A novel interpretable neural network approach for relevant emotion ranking
  - The neural network is initialised to make the hidden layer approximate the behavior of topic models.
  - A novel error function is defined to optimize the whole neural network for relevant emotion ranking.