Word Vector Modeling for Sentiment Analysis of Product Reviews

Yuan Wang, Zhaohui Li*, Jie Liu, Zhicheng He, Yaluo Huang, and Dong Li

Intelligent Information Processing Laboratory

Nankai University, Tianjin, China

2014/12/9

Product Reviews

39/44 人认为此评论有用

坏了, 2014年8月8

评论者 武建伟 - 查看此用户发表

已确认购买 (这是什么?)

评论的商品: HUAWEI 华为 荣耀
移动版 3G手机(白色)真8核处理

用了一个月，手机死机现象严重。
三分之一的位置有很多竖的线。怎

您的投票很重要

这条评论对您有用吗? 是

2/2 人认为此评论有用

千元备用机首选.,

评论者 聂之景 - 查看此用户发表

已确认购买 (这是什么?)

评论的商品: HUAWEI 华为 荣耀
移动版 3G手机(白色)真8核处理

外观漂亮..像素也行..[illegible] 中国网
没想象中那么大..男女皆宜..网上
不清晰什么的..别人的手机我不知
米note间纠结..建议你选购荣耀..
乐时外放音量偏小了点..第一次在

您的投票很重要

这条评论对您有用吗? 是

Poor Battery Life!!!!!, November 21, 2014

By **Inkmast3r** - See all my reviews

Verified Purchase (What's this?)

This review is from: TDK Life on Record TREK Max A34 Wireless Weatherproof Speaker (Electronics)

Do not waste your money the advertising of of plays for 8 hours is not true I purchased it and charged it like the instructions say and only could get 4 hours of play out of it before it went dead and I was playing it at medium volume so two days of the same problem is not worth the headache and a $150!!

Help other customers find the most helpful reviews

Was this review helpful to you? Yes No

Report abuse | Permalink

Comment

0 of 1 people found the following review helpful

Loved it at first, November 21, 2014

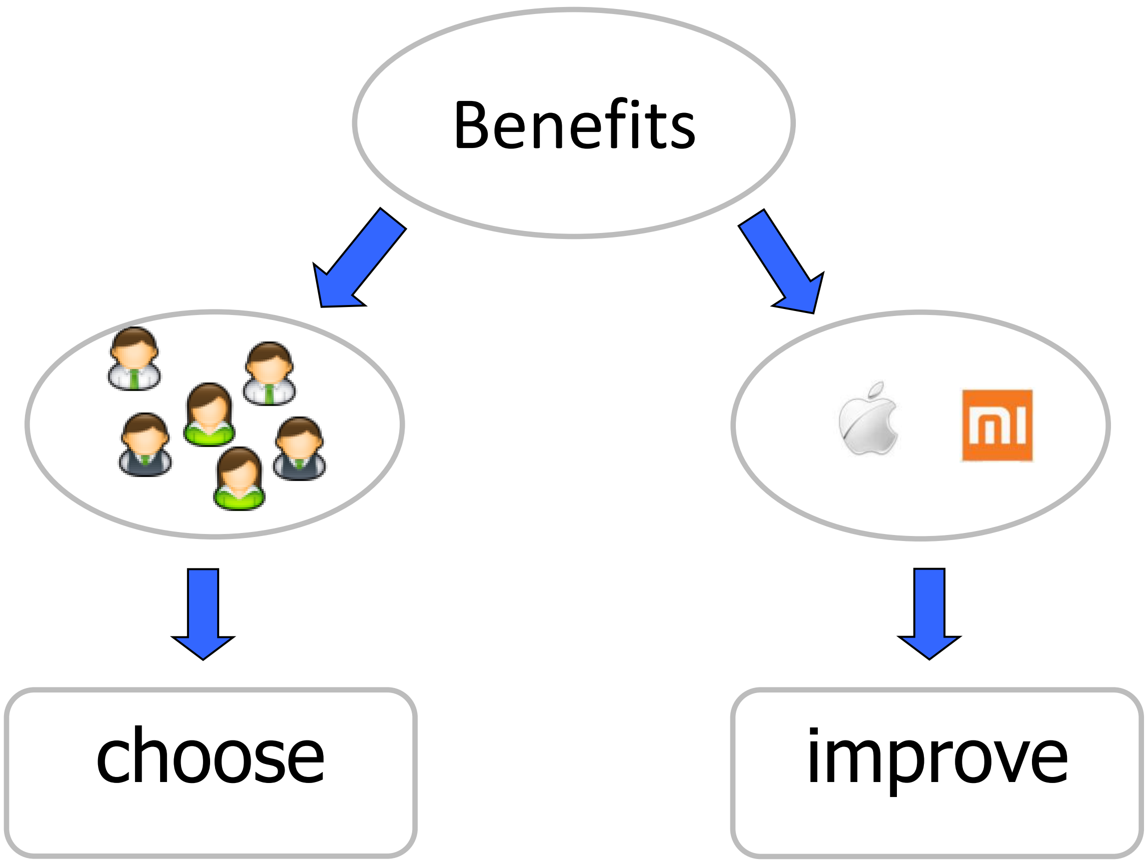
By **Anthony V. Martin** - See all my reviews

Verified Purchase (What's this?)

This review is from: TDK Life on Record TREK Max A34 Wireless Weatherproof Speaker (Electronics)

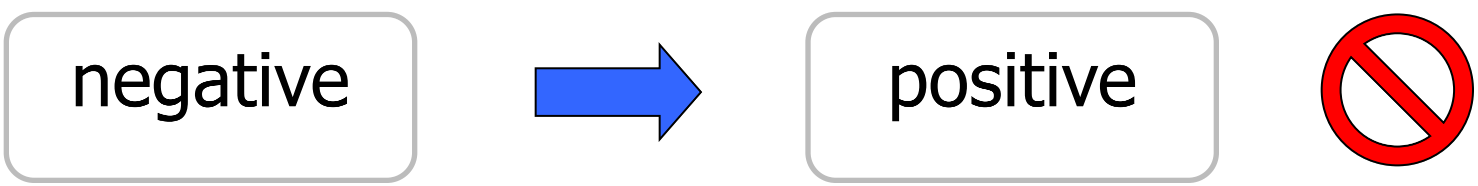
Loved it at first. Great sound. Have had the speaker for 3 and a half months, and it will no longer come on. It doesn't seem to accept a charge. There is no customer

Motivation



Motivation

- Example：
  - < review id=“9214”> The Rice cooker works great. Missing the Steaming Basket. Please send it asap. That’s the reason I purchased because it said “Rice Cooker/Steamer” Been waiting too long< /review>



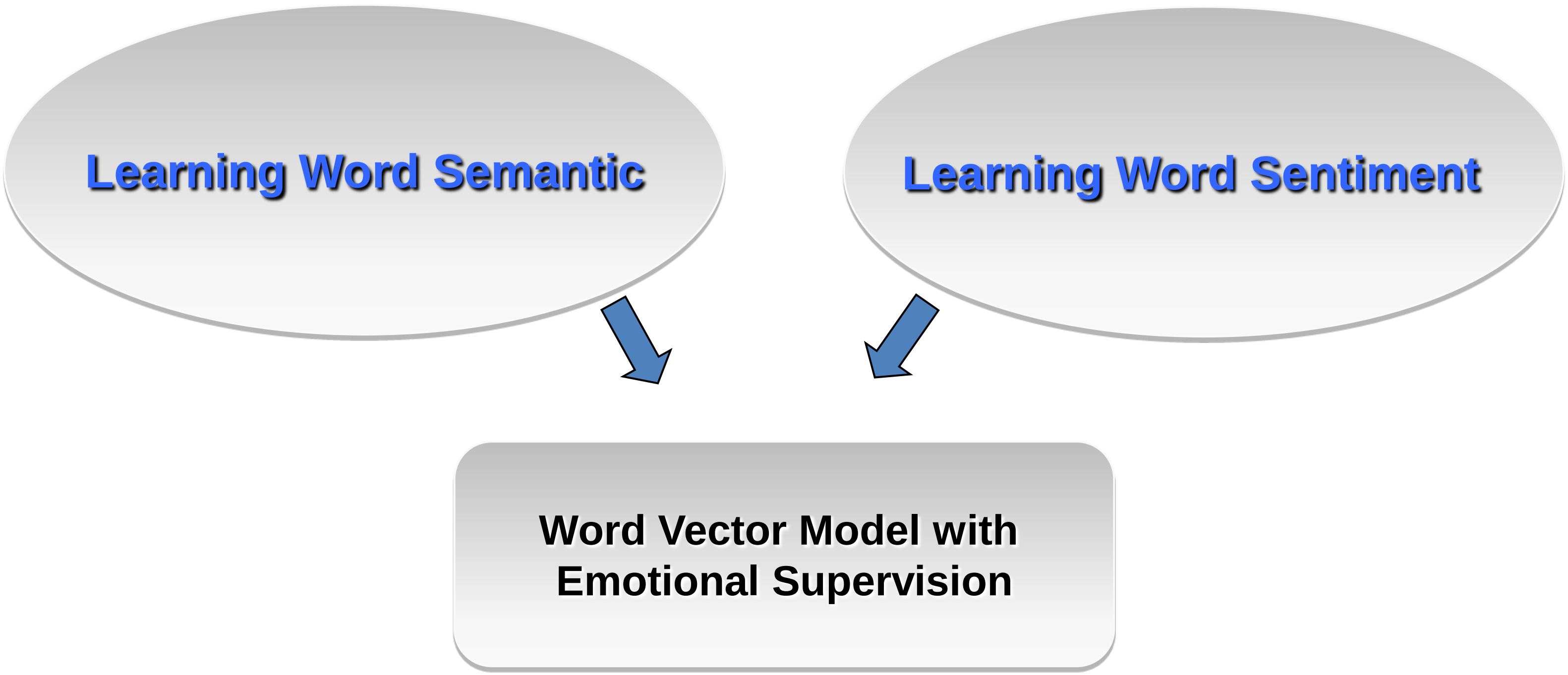
Related Works

- Research directions
  - Sentiment polarity classification
    - **binary classification**(positive and negative)
    - multivariate classification
  - Subjectivity classification
- Classification method
  - The method based on sentiment knowledge
  - The method based on **machine learning techniques**

Related Works

- Word Vector Model for Sentiment Analysis
  - Word representation based on continuous vectors
    - Language model of three-layer neural networks based on "n-gram" assumption, proposed by Bengio in 2003
    - Probabilistic topic models, such as PLSA[12] and LDA[13]
      - Pros: Model semantic information effectively
    - Add supervised sentiment information, such as Maas in 2001[14]
      - Pros: Learns the sentiment information

Framework of our Model



Our work：Considers supervised sentiment information in the learning process of neural network language model, builds a semi-supervised model, and learns the word vectors with both the abilities of sentiment expression and semantics expression.

①Neural Network Language Model

- Every word ***w*** represents a vector C(w), and the current word is predicted by words around it.
  - The generation probability of sentence with $\boldsymbol{N_d}$ words is:

$$P(w_1^{N_d}) = \prod_{t=1}^{N_d} P(w_t \mid w_1^{t-1}) \propto \prod_{t=1}^{N_d} P(w_t \mid w_{t-n+1}^{t-1})$$

  - The likelihood optimization of the model is:

$$\max \sum_{d} \sum_{t=1}^{N_d} \log P(w_t \mid w_{t-n+1}^{t-1})$$

①Neural Network Language Model

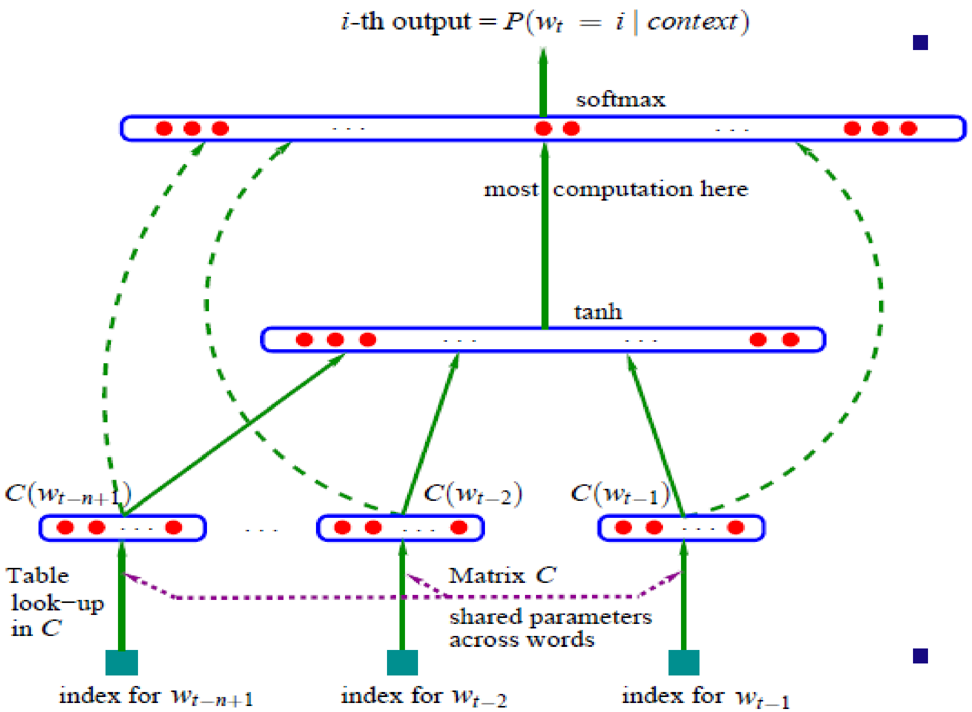


Fig.Neural Network Language Model[5]

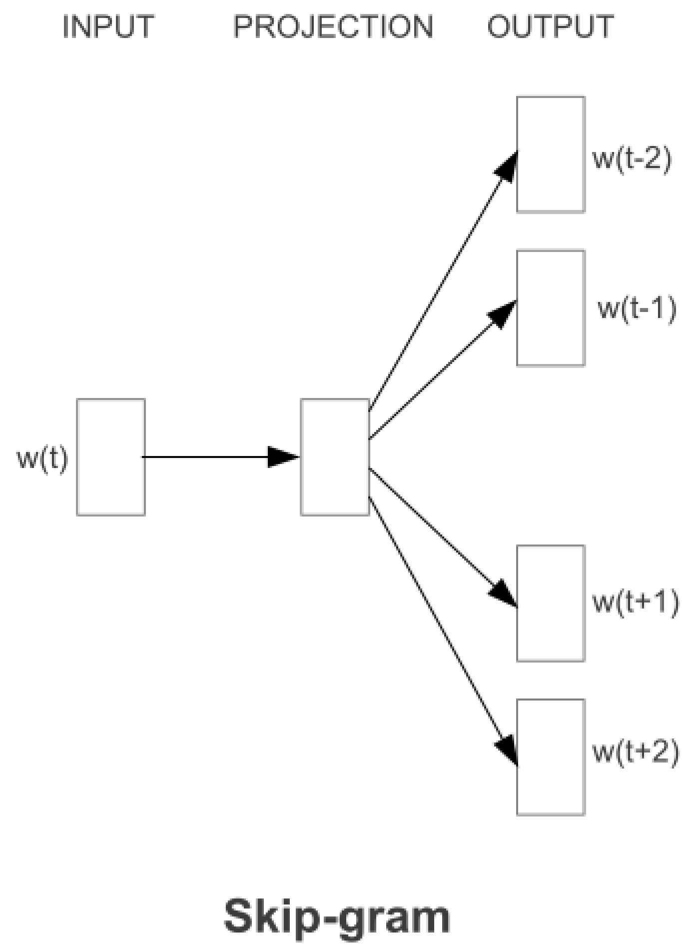
- The hidden laye：
  - the input is ***d + Hx***
  - the output is a **| V |** dimension vector ***y*** ， the calculation of ***y*** is：

$$y = b + Wx + U * \tanh(d + Hx)$$

- The probability of the next word is:

$$P(w_t \mid w_{t-n+1}^{t-1}) = \frac{e^{y_{w_t}}}{\sum_i e^{y_{w_i}}}$$

Skip-gram model



- NNLM structure in semantic learning
- Skip-gram model[15]
  - Range window
  - Simplify the Hidden layer
  - Goal：

$$\sum_{d}\sum_{t=1}^{N_d}\sum_{-c\le j\le c,\, j\ne 0} P(w_{t+j}\mid w_t)$$

  - y = b + W x

②Learning Word Sentiment

- Learning words'sentiment polarity
  - Assumption:
    - Every word has the same capability to express sentiment independently
  - Strategy:
    - Add the Sentiment information of reviews
    - Use Logistic Regression Linear model to model sentiment information of each word

② Learning Word Sentiment

- Map sentiment labels into {0, 1}, denoted as ***s***.
  - The generation probability of a sentence sentiment label is:

$$P(s \mid d; C, a, d_s) = \frac{\mathbf{1}}{N_d} \sum_{t=\mathbf{1}}^{N_d} P(s \mid w_t; C, a, b_s)$$

  - For every word:

$$P(s \mid w_t; C, t, a) = \sigma(aC(w_t) + b_s)$$

This model makes words with similar sentiment information closer in the semantic space, and can be separated by a hyper plane.

$$\max \sum_{d} \frac{1}{N_d} \sum_{t=\mathbf{1}} P(s \mid w_t; C, a, b_s)$$

Word Vector Model with Emotional Supervision

- The combination of the two part：
  - Purpose：
    - Learn semantic information and sentiment polarity characteristics of words simultaneously
  - Combined objective function：

$$L = \sum_{d} \sum_{t=1}^{N_d} \sum_{-c \le j \le c, j \ne 0} P(w_{t+j} \mid w_t; C, W, b) + \beta \sum_{d} \frac{1}{N_d} \sum_{t=1}^{N_d} \log P(s \mid w_t; C, a, b_s)$$

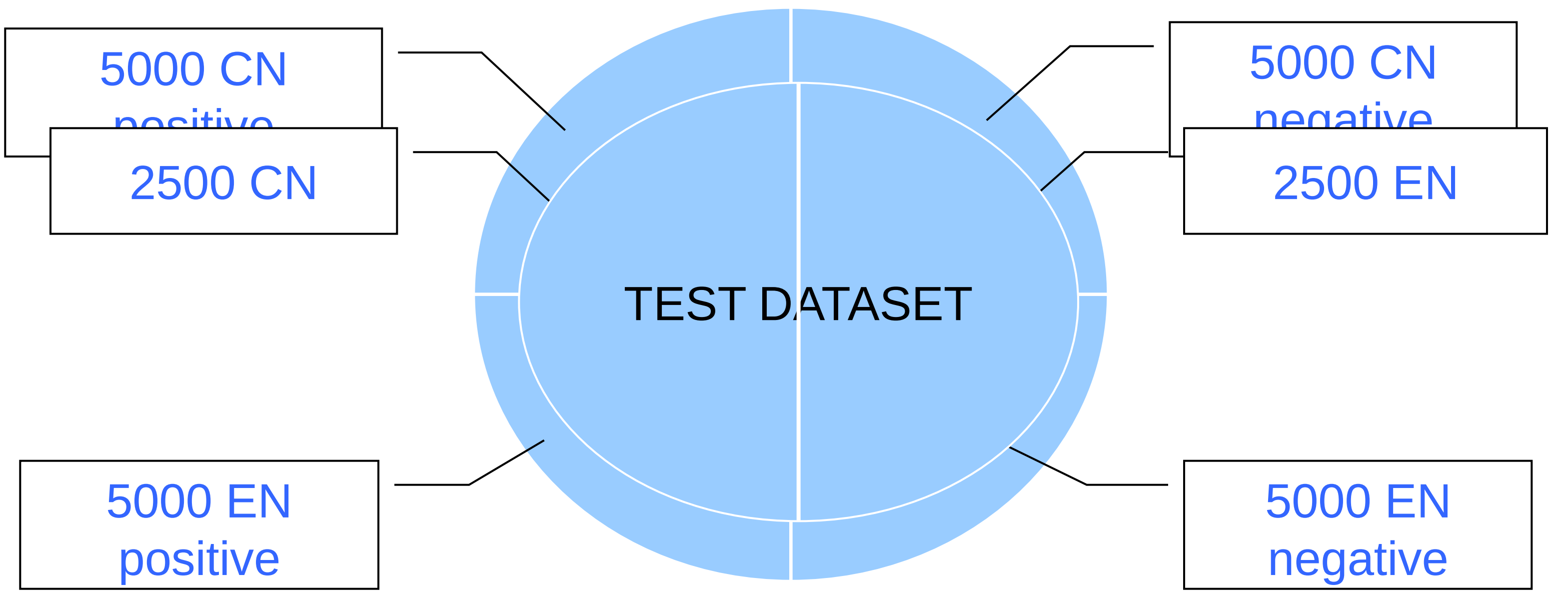
    - β is a parameter to balance two parts. The parameters need to learn is:

$$\theta = (C, W, b, a, b_s)$$

    - We learn parameters with stochastic gradient ascent method.

Experiment Dataset

- Resources
  - Product review dataset in NLPCC2014 “Sentiment Classification with Deep Learning Technology”
  - The dataset is divided into training and testing sets.

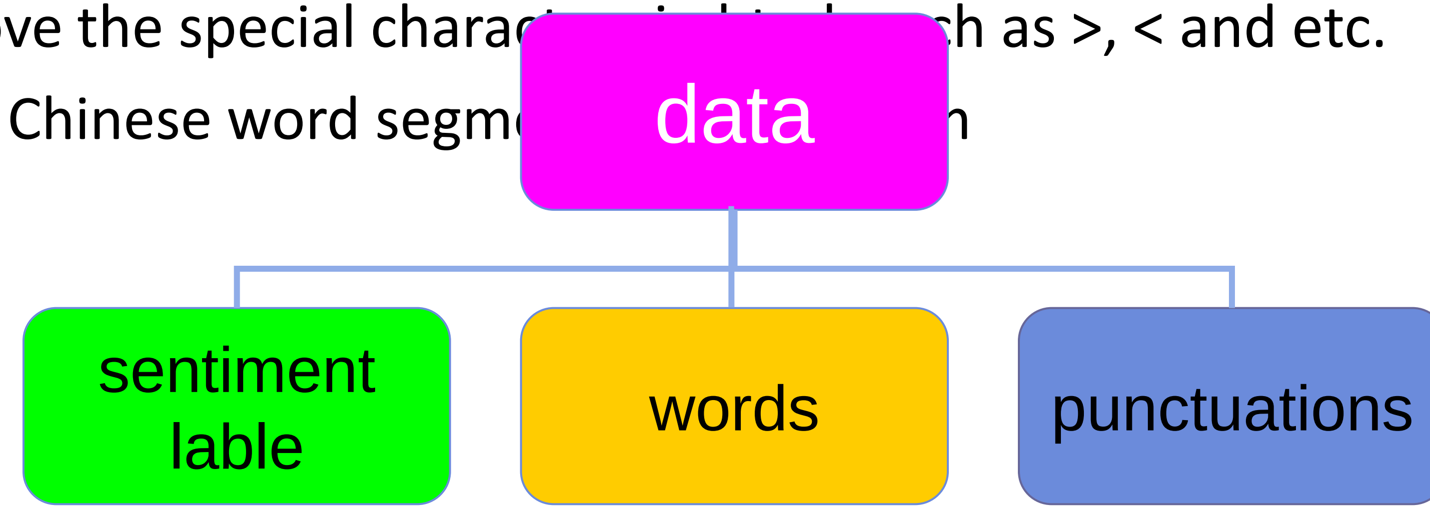


Dataset Pretreatment

- Example：

<review id="8882">
最喜欢的就是田连元了！！！ 有一个网址上有各种评书的 MP3， 可以下载的！！
http://music.jschina.com.cn/adsu.asp?id=385&userid=62039
</review>

  - 1)Remove all hyperlinks, such as http://music.jschina.com.cn/adsu.asp?id=385&userid=62039.
  - 2)Remove the special charac ch as >, < and etc.
  - 3)Yebol Chinese word segme n



Quantitative Evaluation

- In training phase
  - Acc for guiding parameter selection
- In testing phase
  - precision (P), recall (R) and F1

$$P_{pos} = \frac{TP}{TP+FP}, \qquad R_{pos} = \frac{TP}{TP+FN}, \qquad F\mathbf{1}_{pos} = \frac{\mathbf{2}P_{pos}R_{pos}}{P_{pos}+R_{pos}}$$

$$P_{neg} = \frac{TN}{TN+FN}, \qquad R_{neg} = \frac{TN}{TN+FP}, \qquad F\mathbf{1}_{neg} = \frac{\mathbf{2}P_{neg}R_{neg}}{P_{pos}+R_{neg}}$$

$$Acc = \frac{TP+TN}{TP+FP+TN+FN}$$

Experimental Process

- Linear support vector machine with 5-fold cross validation
- Experiment feature
  - Sentiment knowledge features (Senti)
  - Term frequency features[1] (TF)
  - our model feature (our model)
- Training
  - 1)Compare our model with TF and Senti
  - 2)Combine the best two sets of features in experiment 1) as final features and do parameter selection
- Testing
  - Use the final feature and the best parameters to predict the sentiment polarity of the test dataset

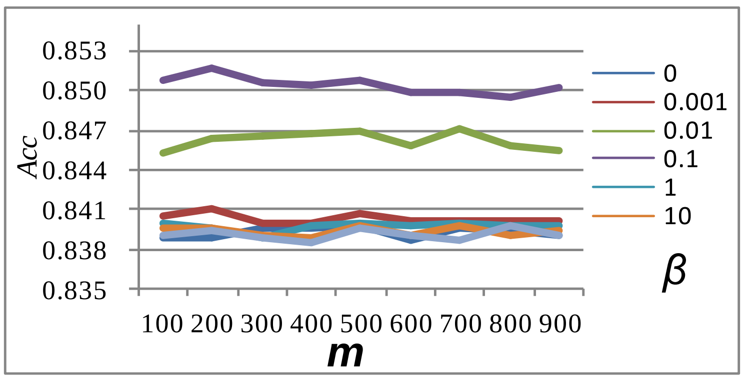
Sentiment knowledge features

Index	Characteristic
1-2	the number of positive/negative words
3	the number of verbs
4	the number of nouns
5	the number of adjectives
6	the number of exclamations
7	the number of question marks
8	the number of ellipses
9	the number of tildes
10-11	the number of positive/negative phrases
12	emotional index
13-14	emotion index of the first/last sentence
15-16	the number of positive/negative sentences
17	the number of neutral sentences

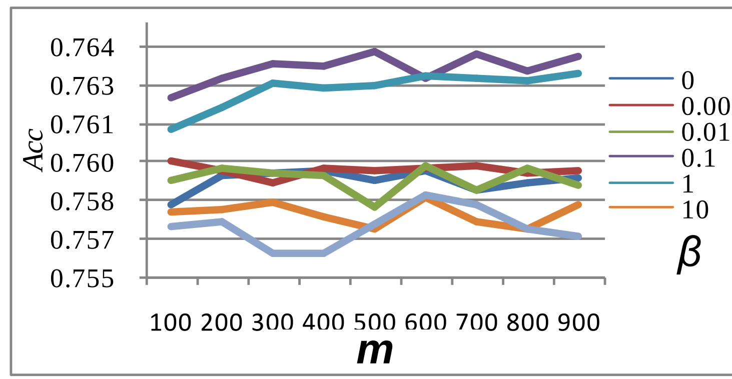
Experimental Results

- **In Training and Validation phase**
  - Classification accuracy with different features

Dataset	Senti	TF	Our Model
English	75.68%	**81.48%**	79.22%
Chinese	66.37%	70.06%	**71.45%**



EN



CN

  - **m**: the dimension of word vector ***C(w)***
  - ***β*** : the balance factor

Experimental Results

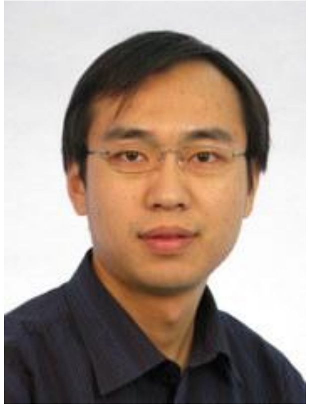
- Classification results on test data

Dataset	Parameters	Negative			Positive		
		P	R	F1	P	R	F1
English	$m = 200, \beta = 0.1$	0.864	0.855	0.860	0.856	0.866	0.861
	$m = 100, \beta = 0.1$	0.865	0.851	0.858	0.853	0.867	0.860
Chinese	$m = 500, \beta = 0.1$	0.780	0.748	0.764	0.758	0.789	0.773
	$m = 700, \beta = 0.1$	0.678	0.452	0.545	0.591	0.794	0.678

Conclusion

- **Proposes a word vector learning method aiming at sentiment analysis**
- **Explores the performance of deep learning models on the task of sentiment classification for product reviews**

Thank you！



Jie Liu

Yuan Wang

Thank you !

contact: zhaohuilee@yeah.net

Reference

- 1. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up?: Sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10. EMNLP '02, Stroudsburg, PA, USA, Association for Computational Linguistics (2002) 79–86
- 2. Linhong, X., Hongfei, L., Jing, Z.: Construction and analysis of emotional corpus. Journal of Chinese Information Processing 22 (2008) 116–122
- 3. Pang, B., Lee, L.: A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In: Proceedings of the 42Nd Annual Meeting on Association for Computational Linguistics. ACL '04, Stroudsburg, PA, USA, Association for Computational Linguistics (2004) 271–278
- 4. Pang, B., Lee, L.: Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics. ACL '05, Stroudsburg, PA, USA, Association for Computational Linguistics (2005) 115–124
- 5. Bengio, Y., Ducharme, R., Vincent, P., Jauvin, C.: A neural probabilistic language model. Journal of Machine Learning Research 3 (2003) 1137–1155
- 6. Collobert, R., Weston, J.: A unified architecture for natural language processing: Deep neural networks with multitask learning. In: Proceedings of the 25th International Conference on Machine Learning. ICML '08, New York, NY, USA, ACM (2008) 160–167
- 7. Toms, M., Martin, K., Luks, B., Jan, C., Sanjeev, K.: Recurrent neural network based language model. In: Proceedings of the 11th Annual Conference of the International Speech Communication Association. INTERSPEECH '10, Makuhari, Chiba, Japan (2010) 1045–1048
- 8. Kim, S.M., Hovy, E.: Automatic detection of opinion bearing words and sentences. In: Companion Volume to the Proceedings of IJCNLP-05, the Second International Joint Conference on Natural Language Processing, Jeju Island, KR (2005) –66
- 9. Tong, R.M.: An operational system for detecting and tracking opinions in on-line discussion (September 2001)
- 10. Turney, P.D.: Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. ACL '02, Stroudsburg, PA, USA, Association for Computational Linguistics (2002) 417–424
- 11. Kim, S.M., Hovy, E.: Automatic identification of pro and con reasons in online reviews. In: Proceedings of the COLING/ACL on Main Conference Poster Sessions. COLING-ACL '06, Stroudsburg, PA, USA, Association for Computational Linguistics (2006) 483–490
- 12. Hofmann, T.: Probabilistic latent semantic indexing. In: Proceedings of the 22Nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '99, New York, NY, USA, ACM (1999) 50–5713. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. 3 (March 2003) 993–1022
- 14. Maas, A.L., Daly, R.E., Pham, P.T., Huang, D., Ng, A.Y., Potts, C.: Learning word vectors for sentiment analysis. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies- Volume 1. HLT '11, Stroudsburg, PA, USA, Association for Computational Linguistics (2011) 142–150
- 15. Mikolov, T., tau Yih, W., Zweig, G.: Linguistic regularities in continuous space word representations. In: Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies(NAACL-HLT-2013), Association for Computational Linguistics (May 2013) 746–751
- 16. Morin, F., Bengio, Y.: Hierarchical probabilistic neural network language model. In: Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics, Citeseer (2005) 246–252
- 17. Jinggang, W., Xiao, Y., Dandan, S., Lejian, L., Wei, Z.: A chinese bag-of-opinions method for weibo sentiment analysis. In: Proceedings of the First CCF Conference on Natural Language Processing and Chinese Computing. (2012) 1–6
- 18. Lixing, X., Ming, Z., Maosong, S.: Hierarchical structure based hybrid approach to sentiment analysis of chinese micro blog and its feature extraction. Journal of Chinese Information Processing 26 (2012) 73–83

①Neural Network Language Model

- The hidden laye：
  - the input is ***d + Hx***
  - the output is a | V | dimension vector ***y*** ， the calculation of ***y*** is：

$$y = b + Wx + U * \tanh(d + Hx)$$

- The probability of the next word is:

$$P(w_t \mid w_{t-n+1}^{t-1}) = \frac{e^{y_{w_t}}}{\sum_i e^{y_{w_i}}}$$

- We estimate parameters with the gradient descent method. ***C*** is randomly initialized, ***(b, W, U, d, H)*** are all initialized with 0

The Optimization of the Model

- WHY?
  - Do an iterative calculation of all output nodes because of the softmax method
  - Sentiment Analysis complexity of our model increases exp onentially with | V | increases.
- HOW?
  - Hierarchical softmax method
    - using Huffman coding to code output layer as a node tree
    - Details are in [16]
  - Tree encoding can adopt any hierarchical agglomerative clustering method
    - ignores normalizing term in softmax without normalized operation