

# NEURAL NETWORK-BASED VECTOR REPRESENTATION OF DOCUMENTS FOR READER- EMOTION CATEGORIZATION

Yu-Lun Hsieh, Shih-Hung Liu,  
Yung-Chun Chang, Wen-Lian Hsu

Institute of Information Science, Academia Sinica, Taiwan

# OUTLINE

- Introduction
- Previous Work
- Proposed Method
- Experiments
- Results & Discussion
- Conclusion

# INTRODUCTION

- Sharing online has become increasingly easy, and people often post on social media websites their experiences and emotions regarding virtually anything
- By means of modern computational technologies, we can quickly collect and classify data about human emotions for further research
- More and more businesses have realized the potential of analyzing online opinions toward their products and services

# INTRODUCTION (II)

- Emotion classification: to predict the emotion (e.g., *happy* or *angry*) of the given text. There are two aspects:
  - Writer's emotion: the emotion expressed by the author of an article. A writer may directly express her feelings through some emotional words or emoticons 😊
  - Reader's emotion: the reader's response after reading it. It can be invoked by not only the content but also personal experiences or knowledge
- A news title “Dozens killed after a plane crashes” is likely to trigger *angry* or *worried* emotions in its readers, despite the fact that it is a description of an event which contains no emotional words

# INTRODUCTION (III)

- Recently there is an increasing interest in vector space representations for words and documents through neural network or deep learning models
- It inspired us to use vectors learned from neural networks, and a classifier to categorize reader-emotions in news articles
- We hope to utilize the power of deep learning to capture hidden connections between the words and the potential invocation of human emotions

# OUTLINE

- Introduction
- Previous Work
- Proposed Method
- Experiments
- Results & Discussion
- Conclusion

# PREVIOUS WORK

- Previous work about emotion detection were mainly focused on the writer's emotion
- **Emoticons:** an important feature. They were taken as emotion tags, and keywords were taken as features. (Read, 2005) Others use emoticons as tags to train SVMs at the document or sentence level. (Mishne, 2005; Yang & Chen, 2006)
- **Pure text:** movie reviews (Pang et al., 2002), student's daily expressions (Wu et al., 2006)

# PREVIOUS WORK (II)

- However, as found in Yang et al (2009), writers and readers do not always share the same emotions regarding the same text
- Classifying emotion from the reader's perspective is a challenging task, and research on this topic is relatively sparse as compared to those considering the writer's point of view



# OUTLINE

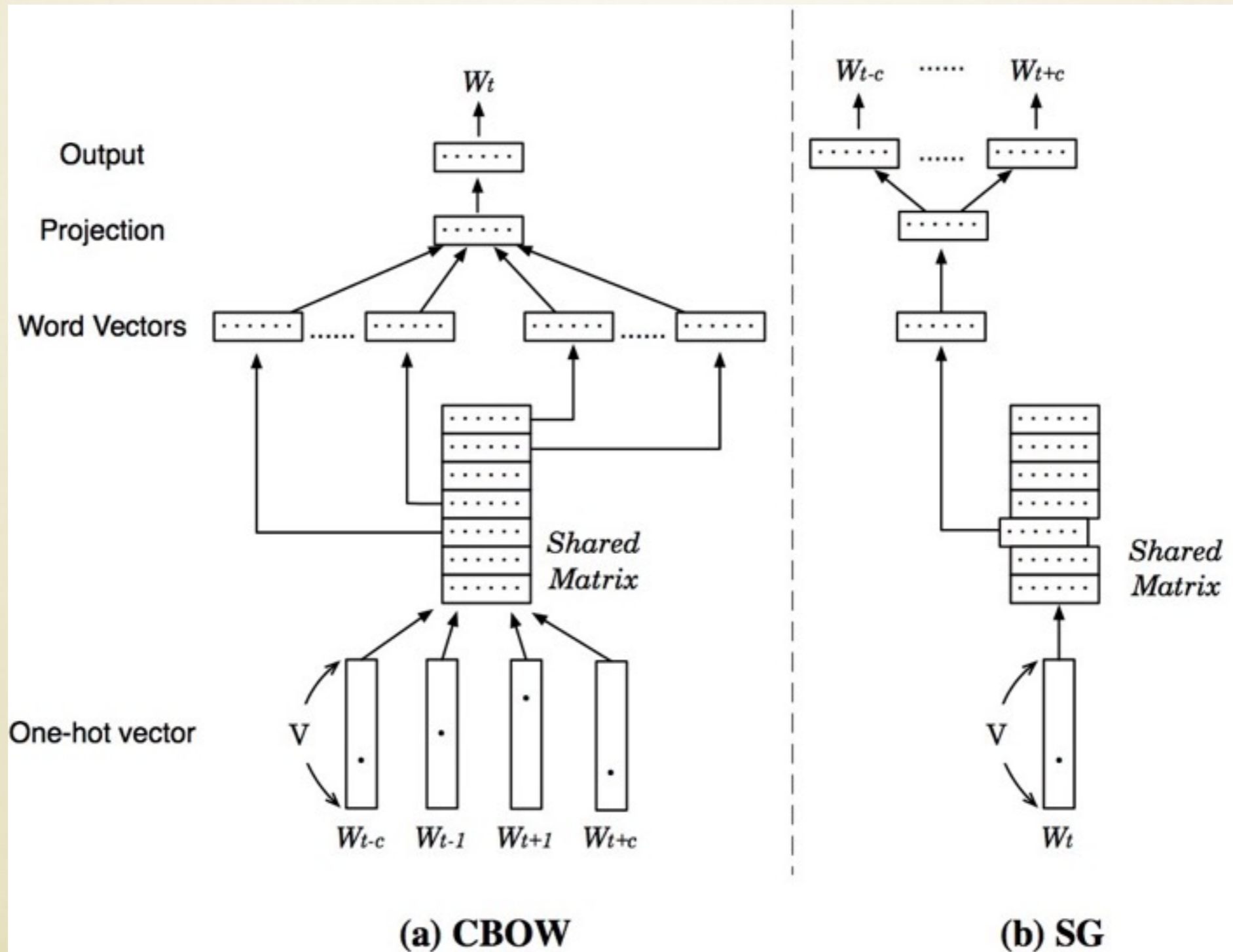
- Introduction
- Previous Work
- Proposed Method
- Experiments
- Results & Discussion
- Conclusion

# METHOD

- We propose a novel usage of document embedding for emotion classification
- Document embedding (or representations): a by-product of neural network language model
- It can learn latent semantic and syntactic regularities in various NLP applications
- Representative methods for the word level include the continuous bag-of-word (CBOW) model and the skip-gram (SG) model (Mikolov et al., 2013)

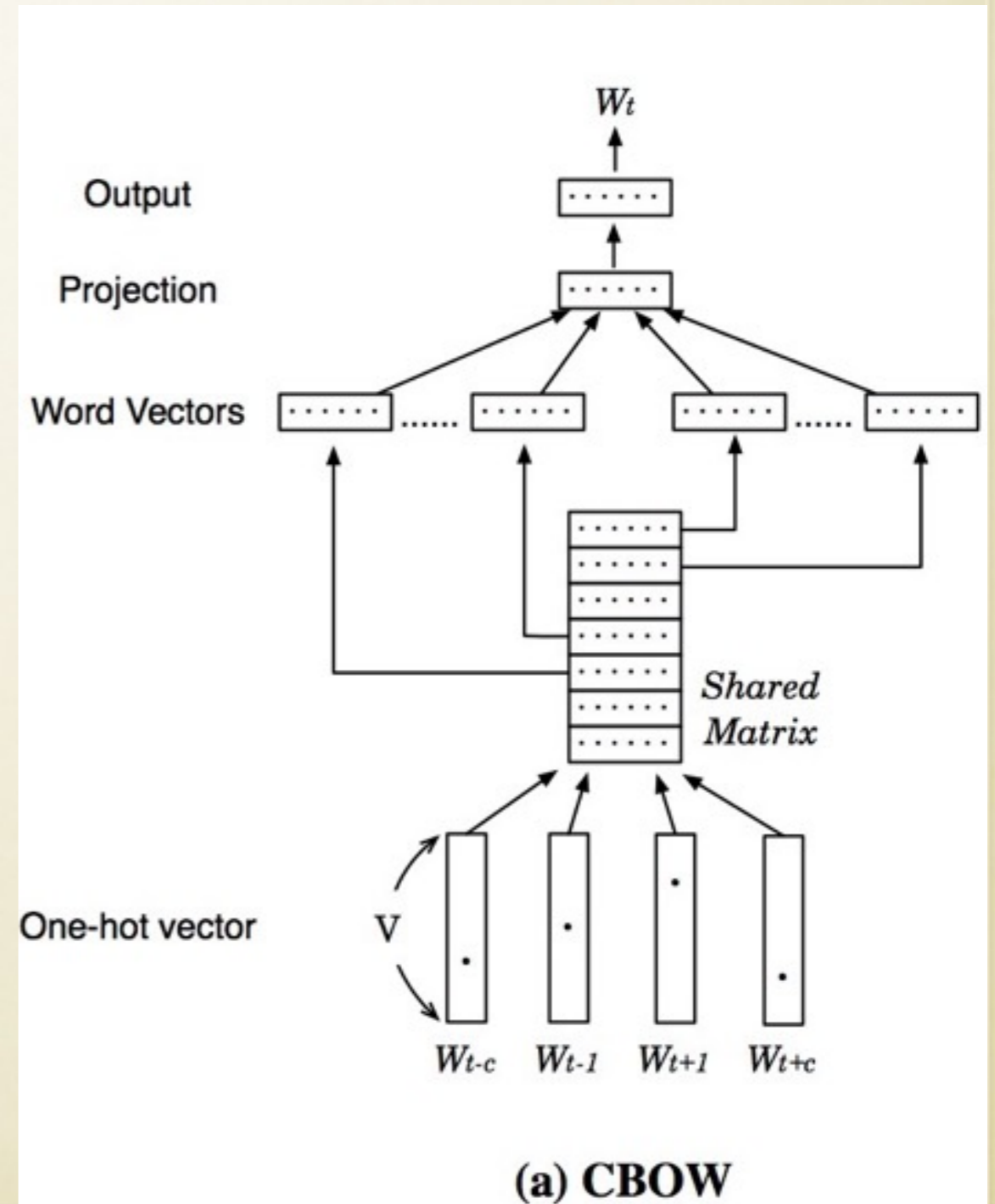
# VECTOR REPRESENTATION OF A WORD

• Mikolov et al. (2013)



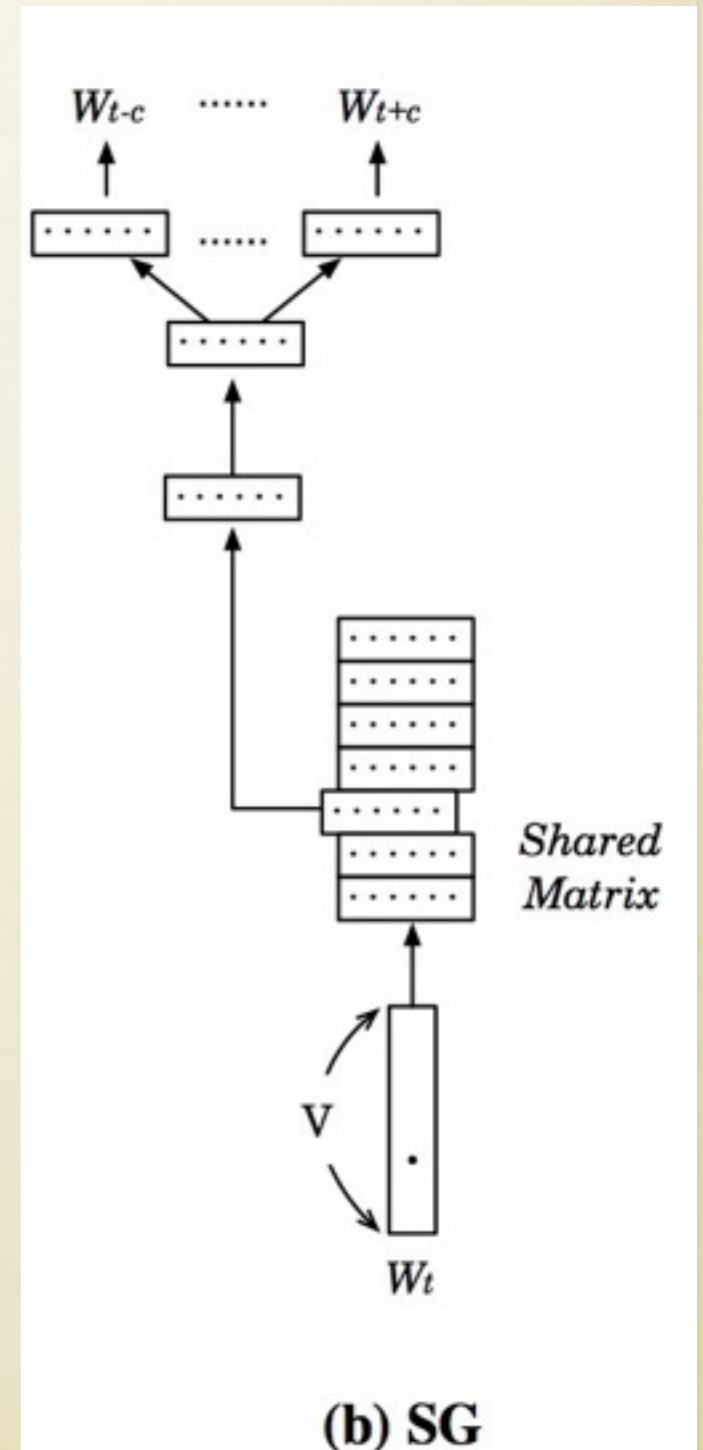
# CBOW

- Predict this word based on its neighbors
- Sum vectors of context words
- Linear activation function in hidden layer
- Output a vector
- Back-propagation to adjust the input vectors



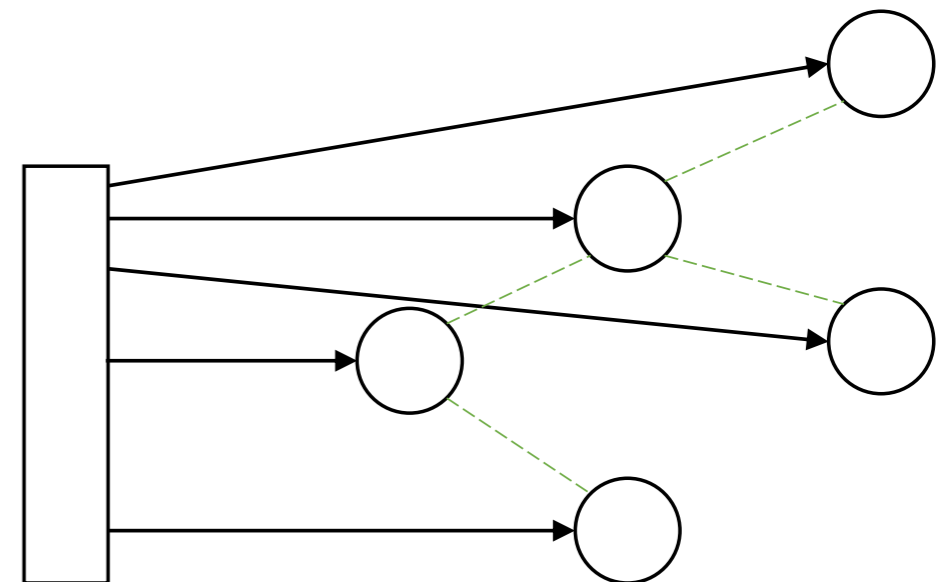
# SKIP-GRAM (SG)

- Predict neighbors word based on this word
- Input vector of this word
- Linear activation function in hidden layer
- Output n other words
- Back-propagation to adjust the input vector



# HIERARCHICAL SOFTMAX

- Improvements to the training procedure have been proposed to increase speed and effectiveness
- Hierarchical softmax involves constructing a Huffman tree to increase speed (frequent words have short binary codes)
- Only update those on the path



HIDDEN LAYER

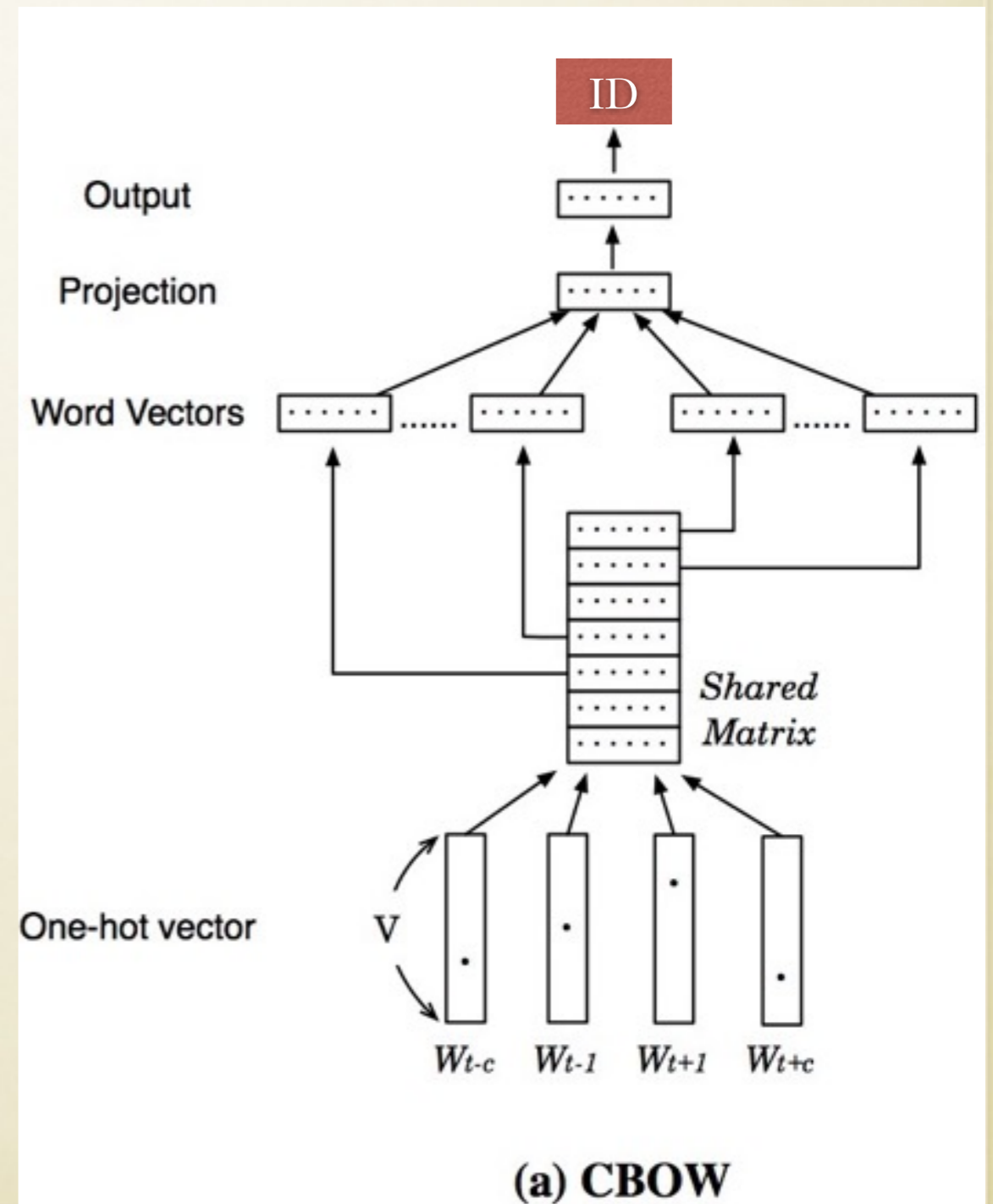
HIERARCHICAL SOFTMAX

# FROM WORD TO DOCUMENT

- By the same line of thought, we can represent a sentence/paragraph/document using a vector.  
(Le and Mikolov, 2014)
- A sentence or document ID is put into the vocabulary as a special word.
- Train the ID with the whole sentence/document as the context.

# CBOW FOR DOCUMENT

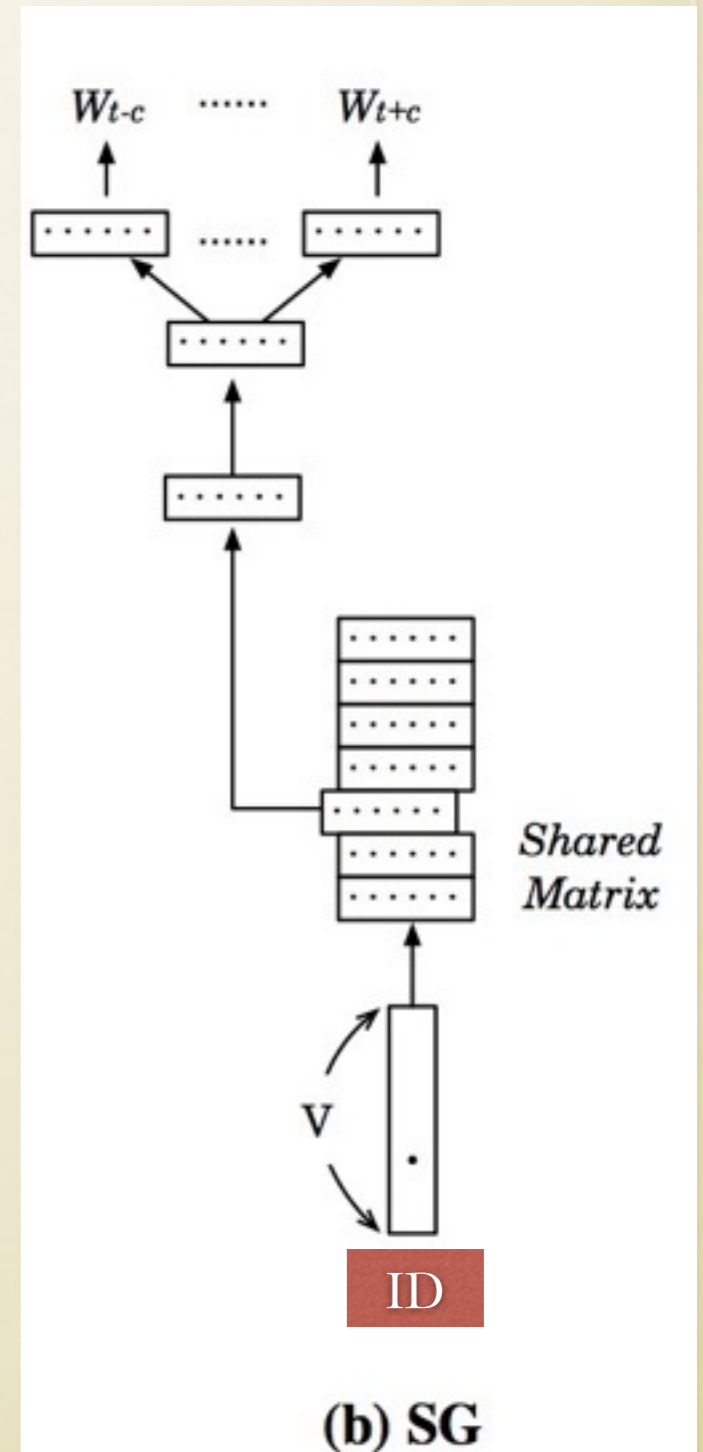
- The ID is used as a special word in the document and slide throughout the whole document
- During learning, both ID and word vectors can be updated





# SG FOR DOCUMENT

- The ID is used to predict every word in the document



# OUTLINE

- Introduction
- Previous Work
- Proposed Method
- Experiments
- Results & Discussion
- Conclusion

# CORPUS

- We collected a corpus of Chinese news articles from Yahoo! online news
- In which, each article is voted from readers with emotion tags in eight categories: *angry*, *worried*, *boring*, *happy*, *odd*, *depressing*, *warm*, and *informative*
- We consider the voted emotions as the reader's emotion toward the news
- Following previous studies that used a similar source, we exclude *informative* as it is not considered as an emotion category

# EXPERIMENTAL SETTING

- We only consider coarse-grained emotion categories. Thus, fine-grained emotions *happy*, *warm*, and *odd* are merged into ‘*positive*’, while *angry*, *boring*, *depressing*, and *worried* are merged into ‘*negative*’
- Only articles with a clear statistical distinction between the highest vote of emotion and others determined by *t*-test with a 95% confidence level are retained
- 27,000 articles are kept and divided into the training set and the test set, each containing 10,000 and 17,000 articles, respectively
- Evaluation: we adopt the convention of using accuracy

# PROPOSED METHOD

- DV-SVM: the proposed method first train CBOW and SG word vectors, and then document vectors
- Afterwards, they are used as representations for the documents for SVM in order to classify the reader-emotion of the document
- We first experiment with various settings for the dimensionality of the vector, and the best settings are compared with other methods described next

# COMPARISONS

- Naive Bayes (denoted as NB)
- Probabilistic graphical model LDA + SVM (denoted as LDA)
- Keyword-based model + SVM (denoted as KW)
- CF: state-of-the-art reader-emotion recognition method in Lin et al. (2007) that combines various features including bi-grams, words, meta-data, and emotion category words

# OUTLINE

- Introduction
- Previous Work
- Proposed Method
- Experiments
- Results & Discussion
- Conclusion

# RESULTS I

<i>Dimensionality</i>	<i>Model</i>	
	CBOW	SG
10	76.69	75.98
50	83.94	80.48
100	85.97	81.81
150	86.67	82.63
300	<b>87.37</b>	<b>85.47</b>
400	84.62	83.38

- Using only 10 dimensions can achieve a substantial accuracy of over 75%.
- Performance is generally positively related to dimensionality.
- Difference between the two models is not very obvious.
- Increasing the dimensionality does not promise to improve the performance.
- Best performance for both models: 300.
- CBOW model reaches a slightly better accuracy of **87.37%** than SG's **85.47%**.



# RESULTS II

<i>Methods</i>	<i>Accuracy(%)</i>
<b>NB</b>	52.78
<b>LDA</b>	74.16
<b>KW</b>	80.81
<b>CF</b>	85.70
<b>DV-SVM<sub>CBOW300</sub></b>	<b>87.37</b>

- Using only surface word weightings is not enough
- LDA's ability of including both local and long-distance word relations may be the reason for its success
- Reader-emotion can largely be recognized by using only keywords
- High performance of CF suggests that, in order to capture more profound emotions hidden in the text, we have to consider not only surface words, but also the relations and semantics
- Our method can successfully encode the relations between words and emotion into a dense vector, leading to the best performance

# OUTLINE

- Introduction
- Previous Work
- Proposed Method
- Experiments
- Results & Discussion
- Conclusion

# CONCLUSIONS

- We present a novel approach for reader- emotion classification using document embedding as features for SVM
- Higher dimension does not always guarantee better performance, but maybe related to the characteristics of the corpus
- We demonstrate that using document embedding for reader-emotion classification can yield substantial success

# THANK YOU

Any questions or comments can be sent to:

[morphe@iis.sinica.edu.tw](mailto:morphe@iis.sinica.edu.tw)