Chinese Sentiment Analysis Using Bidirectional LSTM with Word Embedding

Zheng Xiao(✉) and PiJun Liang

College of Information Science and Engineering, Hunan University, Hunan, China
{zxiao,pjliang}@hnu.edu.cn

Abstract. Long Short-Term Memory network have been successfully applied to sequence modeling task and obtained great achievements. However, Chinese text contains richer syntactic and semantic information and has strong intrinsic dependency between words and phrases. In this paper, we propose Bidirectional Long Short-Term Memory (BLSTM) with word embedding for Chinese sentiment analysis. BLSTM can learn past and future information and capture stronger dependency relationship. Word embedding mainly extract words' feature from raw characters input and carry important syntactic and semantic information. Experimental results show that our model achieves 91.46 % accuracy for sentiment analysis task.

Keywords: Chinese sentiment analysis · BLSTM · Word embedding

1 Introduction

The rapidly increase and popularity of Chinese social networking platform(such as Weibo, Wei-chat etc.), online-shopping platform (such as Taobao) and the user amount lead to the explosively increasing amount of user generated text available on the Internet, organizing the vast amount of unstructured text data into structured information has become vital important. Data mining or more specifically, text mining techniques are used to extract knowledge from this type of user generated text content. Sentiment analysis is performed to extract the opinion and subjectivity knowledge from online text, formalize this knowledge discovered and analyze it for specific use [1].

Sentiment Analysis can be considered as a classification task. There are three main classification levels in sentient analysis: document-level [2], sentence-level [2,3], and aspect-level [4] sentiment analysis. Document-level sentiment analysis aims to classify an opinion document as expressing a positive or negative opinion or sentiment. It considers the whole document a basic information unit (based on one topic). Sentence-level aims to classify sentiment expressed in each sentence. Aspect-level sentiment analysis aims to classify the sentiment with respect to the specific aspect of entities. This paper will mainly focus on sentence-level sentiment analysis.

© Springer International Publishing AG 2016
X. Sun et al. (Eds.): ICCCS 2016, Part II, LNCS 10040, pp. 601–610, 2016.
DOI: 10.1007/978-3-319-48674-1_53

Using machine learning method to solve classification problem, there are many researchers and relative paper, they mainly focus on features selected and constructed [5–7]. For Chinese sentiment analysis task, Some syntactic cues like part-of-speech (POS), Chinese word segmentation (CWS), Lexicon and their contextual counterparts are commonly used for sentiment analysis problem [8–10]. Many statistical methods have been researched for Chinese text sentiment analysis, including Maximum Entropy model [11], Hidden Markov Model [12], Condition Random Field (CRF) [13], and so on.

Although sentiment analysis researches have been done for many years, it is still a great challenging task, especially for Chinese text. In above method, there are many drawbacks for sentiment analysis classification task. First, they heavily rely on the performances POS, CWS, and emotion words. Second, selecting effective features is critical to achieved great results. In addition, establishing the feature template for statistical methods requires abundant experience, and it is difficult for new researchers to build a successful classification model.

Recently, deep learning (DL) [14] methods has been successful applied to Natural Language Processing (NLP), such as, machine translation, POS, Question answering [15,16], and deep learning models have also been effective in tackling sentiment analysis problem [17–19]. Graves et al. [19] proposed long short-term memory (LSTM) neural network for sequences modeling task. And Socher et al. [17,18] used tree-structured long short-term memory networks to improve semantic representations. That recurrent neural networks are able to retain memory between training examples, allows it to capture relations between words. And it is more promising to apply recurrent neural networks to solve sentiment analysis problem because its variants LSTM having the ability to capture long short-term dependencies [20]. Although there are many researches about sentiment analysis, they are mainly focused on English text. Due to the difference of English and Chinese, Chinese has richer information than English. Chinese is quite different with English in grammatical structure also. That means Chinese focuses more in context. Therefore, putting English sentiment analysis methods and models into Chinese will lead to different result.

In this paper, we introduce bidirectional LSTM model and show its superiority for sentiment analysis problem. Bidirectional long short-term memory networks are able to incorporate contextual information from both past and future inputs. And word embedding can capture semantic relationship between words. Our model relies on neither the feature selected nor Chinese word segment, it can learn from large raw text corpus automatically. We evaluate our model on the online-shopping reviews and Weibo data show that our proposed architecture achieves superior performance.

2 Recurrent Neural Network

2.1 Overview

A recurrent neural network (RNN) is able to process a sequence of arbitrary length by recursively applying a transition function to its internal hidden states for each symbol of the input sequence. The activation of the hidden states at time step t is computed as a function $\mathcal{H}$ of the current input symbol x_t and the previous hidden states h_{t-1}:

$$h_t = \mathcal{H}(Wx_t + Uh_{t-1}) \tag{1}$$

Where $\mathcal{H}$ is common to use state-to-state transition function, usually a logistic sigmoid function or hyperbolic tangent function. W is the input-to-hidden weight matrix, U is the state-to-state recurrent weight matrix.

2.2 Long Short-Term Memory

Difficulties of training an RNN to capture long-term dependencies during training and components of the gradient vector can grow or decay exponentially over long sequences [22,23]. Hochreiter et al. [20] propose LSTM architecture addresses this problem of learning long-term dependencies by introducing a memory cell that is able to preserve state over long periods of time. There are several variants of LSM [24,25]

Figure 1 illustrates a single LSTM memory cell, which consists of a memory cell c_t, an input gate i_t, a forget gate f_t and an output gate o_t. For the LSTM, $\mathcal{H}$ is implemented by the following block functions:

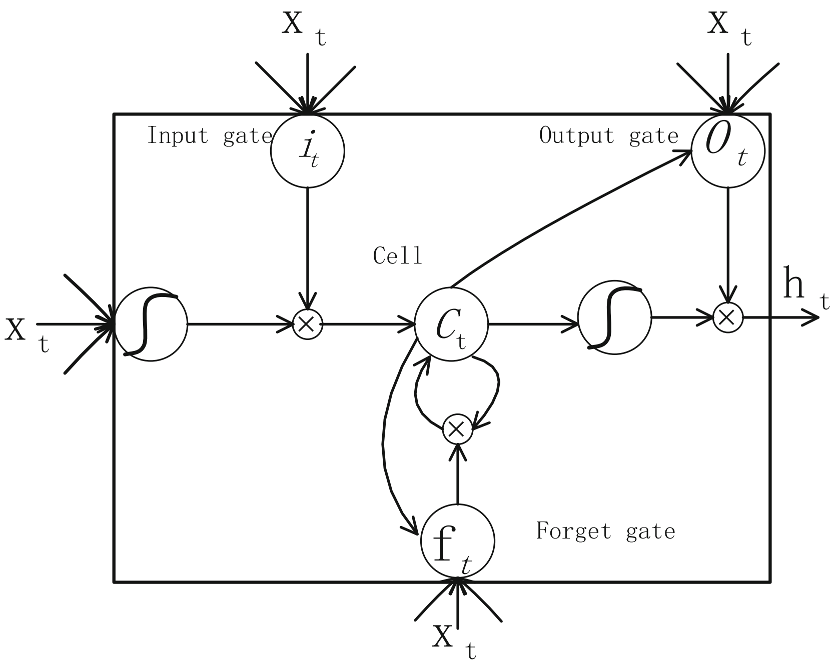

Fig. 1. A single LSTM cell [25]

$$i_t = \sigma(W^{xi}x_t + W^{hi}h_{t-1} + W^{ci}c_{t-1} + b^i) \quad (2)$$
$$f_t = \sigma(W^{xf}x_t + W^{hf}h_{t-1} + W^{cf}c_{t-1} + b^f) \quad (3)$$
$$c_t = f_t \odot c_{t-1} + i_t \odot tanh(W^{xc}x_t + W^{hc}h_{t-1} + b^c) \quad (4)$$
$$o_t = \sigma(W^{xo}x_t + W^{ho}h_{t-1} + W^{co}c_t + b^o) \quad (5)$$
$$h_t = o_t \odot tanh(c_t) \quad (6)$$

Where $x = (x_1, x_2, ..., x_T)$ is the input feature sequence, σ is the logistic function. The symbol $\odot$ represent the element-wise operation, W is the weight matrix and the superscript indicates it is the matrix between two different gates.

3 Model

Now we describe how to use a bidirectional long short-term network to build sentiment analysis model. Figure 2 illustrate BLSTM model for Chinese sentiment analysis.

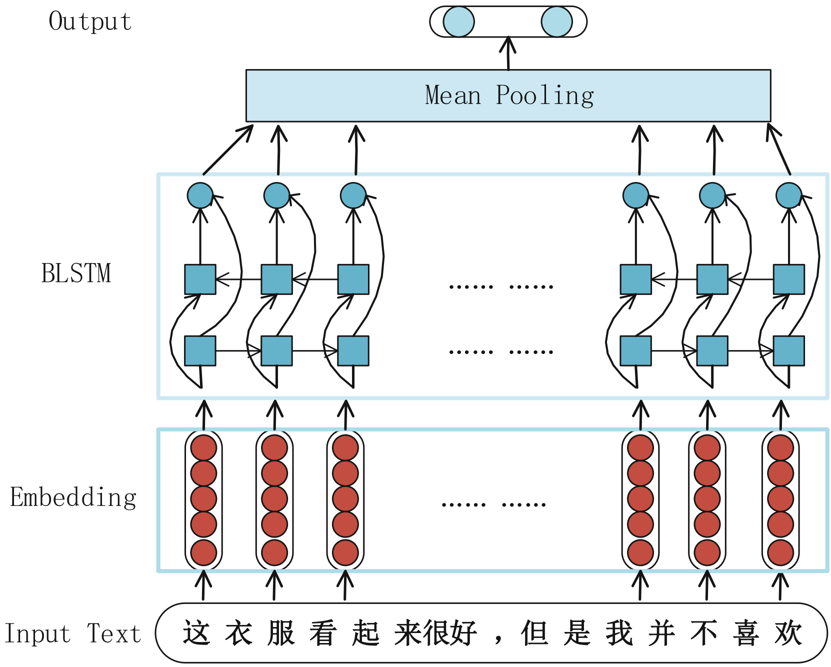

Fig. 2. BLSTM with word embedding sentiment analysis model

Word embedding map the words or phrases from the vocabulary to the vectors of real number in a low-dimensional space than the vocabulary size and it also plays a vital important role for NLP task, since distributed representation or word embedding can carry import syntactic and semantic information [26,27]. And word embedding shows superior performance for Chinese text too [28]. This word embedding layer mainly extracts word feature from raw Chinese characters.

One shortcoming of conventional RNNs is that they are only able to make use of previous context. However, that Chinese has dependency relation with context leads to the result that we need to utilize not only past previous information but also future information. A Bidirectional LSTM consists of two LSTMs that are run in parallel: one on the input sequence and the other on the reverse of the input sequence. At each time step, the hidden state of the Bidirectional LSTM is the concatenation of the forward and backward hidden states. This setup allows the hidden state to capture both past and future information [21].

Figure 3 illustrated a BRNN computes the forward hidden sequence $\overrightarrow{h}$, the backward hidden sequence $\overleftarrow{h}$ and the output sequence y by iterating the backward layer from $t = T$ to 1, the forward layer from $t = 1$ to T and then updating the output layer:

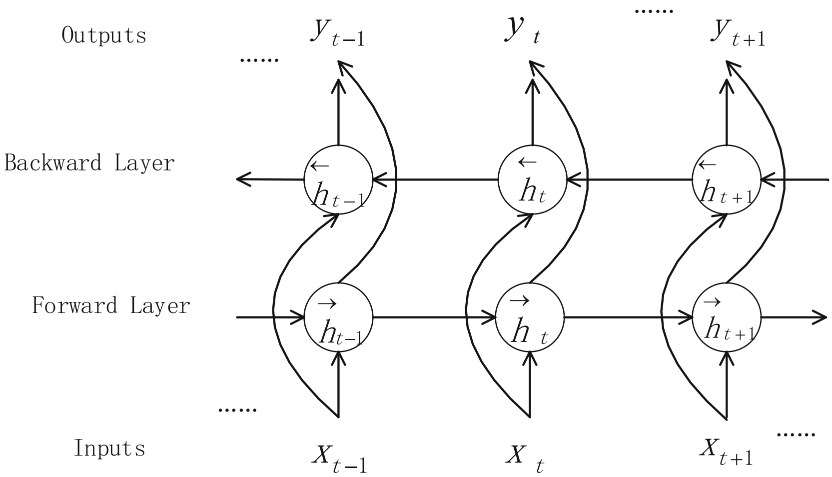

Fig. 3. Bidirectional recurrent neural network

$$\overrightarrow{h}_t = \mathcal{H}(W_{x\overrightarrow{h}}x_t + W_{\overrightarrow{h}\overrightarrow{h}}\overrightarrow{h}_{t-1} + b_{\overrightarrow{h}}) \tag{7}$$

$$\overleftarrow{h}_t = \mathcal{H}(W_{x\overleftarrow{h}}x_t + W_{\overleftarrow{h}\overleftarrow{h}}\overleftarrow{h}_{t-1} + b_{\overleftarrow{h}}) \tag{8}$$

$$y_t = W_{\overrightarrow{h}y}\overrightarrow{h}_t + W_{\overleftarrow{h}y}\overleftarrow{h}_t + b_y \tag{9}$$

Where $y = (y_1, y_2, ..., y_T)$ is the BLSTM sequence output.

In sentiment analysis model, the BLSTM input is the word embedding and $\mathcal{H}$ is the LSTM block transition function. And we wish to predict sentiment label $\widetilde{c}$ from some discrete of class $\mathcal{C}$ (in this paper, $\mathcal{C}$ is the binary sentiment label set: positive or negative). Given the input sequence sample $x = (x_1, x_2, ..., x_T)$, and $y = (y_1, y_2, ..., y_T)$ for the BLSTM output. $z = \frac{1}{T}\sum_i y_i$ is the average over all the timesteps results of the y. $\widetilde{c}$ is predicted by softmax classifier that takes the BLSTM average output z as input:

$$\widetilde{p}_\theta(c|\{x\}_i) = softmax(W^z z + b^z) \tag{10}$$

$$\widetilde{c}_i = argmax\widetilde{p}_\theta(c|\{x\}_i) \tag{11}$$

4 Experiments

We evaluate our model on online-shopping product reviews and Weibo data, all these reviews and data has been tagged by their author as positive and negative. This corpus consists of 13000 reviews, which has 7000 positive comments and we shuffled the positive and negative data. We use the standard train/dev/test splits 9100/950/2950 for sentiment predict task.

In comparing our LSTM/BLSTM model against for Chinese sentiment analysis, we used CRF-based model as baseline. And the CRF-based feature template used the model proposed by Li et al. [13].

In order to get the best performance of BLSTM architecture, we control the BLSTM and LSTM architecture model by varying their number of hidden units. Specifically, for two layers LSTM/BLSTM, the number of parameters are kept the small, the forward and backward transition function are share for BLSTM. Table 1 gives the number of hidden units included in LSTM and BLSTM.

Table 1. LSTM architecture and hidden units

LSTM variant	Hidden units
LSTM	200
2-layer LSTM	128
BLSTM	100
2-layer BLSTM	64

We used back-propagation through time (BPTT) method [29] and AdaGrad [30] with learning rate of 0.8 and minibatch size of 20 to train our model. We use the development set to tune the hyperparameters for our model, and use dropout regularized technology [31] to against overfitting. The experiment result with baseline are summarized in the Table 2.

According to the experiment, we found the best performances for Chinese text sentiment analysis which are obtained by the LSTM and BLSTM architecture topology relatively, In order to investigate the performance of word embedding, we compare the pre-random initial and pre-train word embedding initial using word2vec [25], these embedding had 200-dimensional and were tuned during training epoch. We are summarized our result for BLSTM and LSTM model on sentiment analysis in Fig. 4.

Table 2. The result of the sentiment analysis with different method

Model	Accuracy
CRF-based	88.34 %
LSTM	87.49 %
2-layer LSTM	87.00 %
BLSTM	91.46 %
2-layer BLSTM	90.28 %

Fig. 4. BLSTM sentiment analysis accuracy with different initial ways

From the Fig. 4, we found that the pre-train word embedding initial with well-tuned during training epoch can obtain better performance. Because using pre-train word embedding initial can easily train and tune the hyperparameters, and pre-train word embedding initial have been captured the syntactic and semantic information during training. For pre-random initial word embedding, it is difficult to tune weights and capture dependency between words and phrases than pre-word2vec.

We investigate the effect of sentence length on the performance of the LSTM and BLSTM model for sentiment analysis. In Fig. 5, we show the relationship between accuracy with sentence length, we can find bidirectional long short-term memory network obtain better performance than long short-term memory network, and the LSTM get the best performance with sentence length 10, BLSTM with sentence 20. Because BLSTM can learn previous and feature information and can capture more stronger dependency between words and phrases than LSTM. The result demonstrated that the sentence length limits the performance of the BLSTM/ LSTM on sentiment analysis task as well.

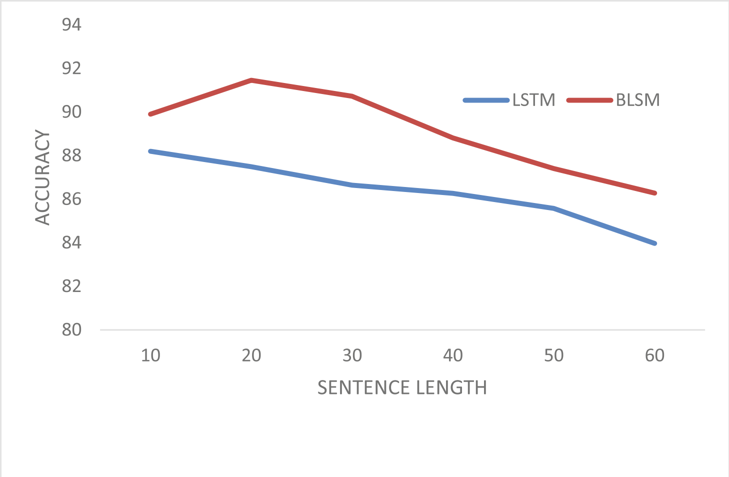

Fig. 5. The effect of the sentence length

5 Conclusion

In this paper, we propose bidirectional long short-term memory network for Chinese text sentiment analysis, and this model take raw characters as input. Word embedding can learned context and syntactic and semantic information during train model, and these information is important for Chinese text sentiment analysis task. The experiment results show it superior performance. Our results suggest that it is promising to use bidirectional long short-term memory network for Chinese text process.

Acknowledgment. The authors thank the anonymous reviewers for their valuable comments and suggestions. The research was partially funded by the National Science Foundation of China (Grant No. 61370095, 61502165), the Open Foundation of Jiangsu Engineering Center of Network Monitoring No. KJR1541, CICAEET fund and PAPD fund, the Scientific Research Fund of Hunan Provincial Education Department (Grant No. 14C0680), and Hunan Normal University of Youth Science Foundation (Grant No. 11404).

References

1. Liu, B.: Sentiment analysis and subjectivity. In: Handbook of Natural Language Processing, pp. 627–666 (2010)
2. Balahur, A., Steinberger, R., Kabadjov, M.: Sentiment analysis in the news. Infrared Phys. Technol. **65**, 94–102 (2014)
3. Long, J., Yu. M., Zhou, M., et al.: Target-dependent twitter sentiment classification. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (2011)

4. Pak, A., Paroubek, P.: Twitter as a corpus for sentiment analysis and opinion mining. In: Proceedings of International Conference on Language Resources and Evaluation (LREc), vol. 10 (2010)
5. Wen, X., Shao, L., Xue, Y., Fang, W.: A rapid learning algorithm for vehicle classification. Inf. Sci. **295**(1), 395–406 (2015)
6. Chen, B., Shu, H., Coatrieux, G., Chen, G., Sun, X., Coatrieux, J.-L.: Color image analysis by quaternion-type moments. J. Math. Imag. Vis. **51**(1), 124–144 (2015)
7. Bin, G., Sheng, V.S., Wang, Z., Ho, D., Osman, S., Li, S.: Incremental learning for - support vector regression. Neural Netw. **67**, 140–150 (2015)
8. Cui, A., Zhang, H., Liu, Y., Zhang, M., Ma, S.: Lexicon-based sentiment analysis on topical chinese microblog messages. In: Li, J., Qi, G., Zhao, D., Nejdl, W., Zheng, H.-T. (eds.) Semantic Web and Web Science, pp. 333–344. Springer, New York (2013)
9. Yuan, B., Liu, Y., Li, H.: Sentiment classification in Chinese microblogs: lexicon-based and learning-based approaches. In: International Proceedings of Economics Development and Research, vol. 68, p. 1 (2013)
10. Wang, D., Li, F.: Sentiment analysis of Chinese microblogs based on layered features. In: Loo, C.K., Yap, K.S., Wong, K.W., Teoh, A., Huang, K. (eds.) ICONIP 2014. LNCS, vol. 8835, pp. 361–368. Springer, Heidelberg (2014). doi:10.1007/978-3-319-12640-1_44
11. Lee, H.Y., Renganathan, H.: Chinese sentiment analysis using maximum entropy. In: Sentiment Analysis Where AI Meets Psychology (SAAIP), vol. 89 (2011)
12. Liu, L., Luo, D., Liu, M., et al.: A self-adaptive hidden Markov model for emotion classification in Chinese microblogs. Math. Prob. Eng. **2015**, 1–8 (2015). doi:10.1155/2015/987189. Article ID 987189
13. Li, J., Hovy, E.H.: Sentiment analysis on the people's daily. In: EMNLP, pp. 467–476 (2014)
14. Bengio, Y., LeCun, Y., Hinton, G.: Deep learning. Nature **521**, 436–444 (2015)
15. Bengio, S., Vinyals, O., Jaitly, N., et al.: Scheduled sampling for sequence prediction with recurrent neural networks. In: Advances in Neural Information Processing Systems, pp. 1171–1179 (2015)
16. Wang, D., Nyberg, E.: A long short-term memory model for answer sentence selection in question answering. In: ACL, July 2015
17. Socher, R., Perelygin, A., Wu, J.Y., et al.: Recursive deep models for semantic compositionality over a sentiment treebank. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1631–1642 (2013)
18. Tai, K.S., Socher, R., Manning, C.D.: Improved semantic representations from tree-structured long short-term memory networks. arXiv preprint arXiv:1503.00075 (2015)
19. Graves, A.: Supervised sequence labelling. In: Graves, A. (ed.) Supervised Sequence Labelling with Recurrent Neural Networks. SCI, vol. 385, pp. 5–13. Springer, Heidelberg (2012)
20. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
21. Graves, A., Jaitly, N., Mohamed, A.-R.: Hybrid speech recognition with dee bidirectional LSTM. In: IEEE Workshop on Automatic Speech Recognition and Under-Standing (ASRU), pp. 273–278 (2013)
22. Hochreiter, S.: The vanishing gradient problem during learning recurrent neural nets, problem solutions. Int. J. Uncertainty Fuzziness Knowl. Based Syst. **6**(02), 107–116 (1998)

23. Bengio, Y., Simard, P., Frasconi, P.: Learning long-term dependencies with gradient descent is difficult. IEEE Trans. Neural Netw. **5**(2), 157–166 (1994)
24. Gers, F.A., Schraudolph, N.N., Schmidhuber, J.: Learning precise timing with LSTM recurrent networks. J. Mach. Learn. Res. **3**, 115–143 (2003)
25. Zaremba, W., Sutskever, I.: Learning to execute. arXiv preprint arXiv:1410.4615 (2014)
26. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Proceedings of NIPS, pp. 3111–3119 (2013)
27. Mikolov, T., Yih, W., Zweig, G.: Linguistic regularities in continuous space word representations. In: Proceedings of HLT-NAACL, pp. 746–751 (2013)
28. Liu, X., Duh, K., Matsumoto, Y., et al.: Learning character representations for Chinese word segmentation. In: NIPS Workshop on Modern Machine Learning and Natural Language Processing (2014)
29. Williams, R.J., Zipser, D.: Gradient-based learning algorithms for recurrent networks, their computational complexity. In: Back-Propagation: Theory, Architectures and Applications, pp. 433–486 (1995)
30. Duchi, J., Hazan, E., Singer, Y.: Adaptive subgradient methods for online learning, stochastic optimization. J. Mach. Learn. Res. **12**, 2121–2159 (2011)
31. Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R.: Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580 (2012)