3. Aspect correlation across all reviews is defined as:

   $P_{review} = \frac{1}{k}\sum_{i=1}^{k} P_{r_i^*, r_i}$ where r_i^* and r_i are the predicted and ground-truth rating vectors for aspect A_i across all the reviews. $P_{r_i^*, r_i}$ is the Pearson correlation between two vectors r_i^* and r_i.

4. Normalized discounted cumulative gain is defined as:

   $$n\mathrm{DCG}_{aspect} = \frac{1}{|D_{test}|}\sum_{d=1}^{|D_{test}|}\frac{DCG_d}{IDCG_d}$$

   where $DCG_d = \sum_{i=1}^{k}\frac{2^{r_{di}} - 1}{\log(i+1)}$ is that highly relevant documents appearing lower in a search result list should be penalized as the graded relevance value is reduced logarithmically proportional to the position of the result. The discounted CG accumulated at a particular rank position d, r_{di} is the predicted aspect rating for aspect A_i, $IDCG_d$ is computed the same as the DCG_d but it uses aspect ratings in the ground-truth.

We evaluate on two experimental cases, including aspect features based on bag of words and aspect features based on word vectors. In each the experimental case, the models used the same data set, we perform 5 times for training and testing, and report the mean value of metrics. In each time, we select randomly 75% of given reviews to train, the remaining 25% of given reviews to test. In Table 6, we show the mean value of four metrics for each method.

Table 6. Comparison with other models

Aspect feature	Method	Δ_{aspect}	P_{aspect}	P_{review}	$n\mathrm{DCG}_{aspect}$
Bag of words ($\|V\| = 3941$)	Global Prediction	0.825	0.316	0.569	0.705
	LRR	0.718	0.363	0.638	0.736
	Our LRMNN	0.723	0.451	0.632	0.756
	PRank	0.439	0.624	0.743	0.898
Word vectors ($\|V\| = 3941$)	LRR	0.743	0.403	0.654	0.844
	Our LRMNN	0.712	0.468	0.644	0.928
	PRank	0.412	0.631	0.777	0.925
Word vectors ($\|V\| = 29349$)	LRR	0.738	0.406	0.659	0.823
	Our LRMNN	**0.705**	**0.488**	**0.711**	**0.913**
	PRank	0.409	0.635	0.779	0.924

We can see that when aspect features present based on bag of words, although our model does not performs better than LRR model on Δ_{aspect} and P_{review} but it performs better than Global Prediction on P_{aspect}, P_{review} and $n\mathrm{DCG}_{aspect}$ and it also performs better than LRR on P_{aspect} and $n\mathrm{DCG}_{aspect}$. The model PRank performs best in all metrics. However, this model is fully supervised (i.e. both aspect ratings and overall rating) while others are not. When aspect features represent based on word vectors with dimensional number is 200 and the dictionary size of words is $|V| = 3941$ or $|V|=29349$, we see that all models perform better when they use aspect features present based on bag of words. For each model, our model LRMNN performs better than LRR on Δ_{aspect}, P_{aspect} and $n\mathrm{DCG}_{aspect}$, the model Global Prediction only depend on bag of words that does not depend on the dimensional word vector, so we do not evaluate it in this experimental case. For each the dictionary size of words, we use $|V| = 29349$ for metrics better than $|V| = 3941$.

5.4.2. *Evaluation on overall weighting aspects*

Since the given data have not the ground-truth aspect weights of aspects, we only can evaluate them indirectly through overall rating prediction. The combination of the predicted results of aspect rating and the inference results of aspect weight help us to predict overall rating.

We denote $\alpha_{(GlobalPrediction+LSAWs)}$ is the vector of the overall aspect weights which is computed by the method LSAWs; α_{PRR} is the vector of the overall aspect weights and is computed by LRR algorithm; α_{LRMNN} is the vector of the overall aspect weights which is computed by the method LRMNN.

We evaluate the quality of these overall aspect weights through three cases of overall rating prediction: (1) $PRank + \alpha_{(GlobalPrediction+LSAWs)}$, (2) $PRank + \alpha_{PRR}$, (3) $PRank + \alpha_{LRMNN}$ We choose the two popular metrics, Mean Absolute Error (MAE) and Root mean square error to measure the differences of overall rating.

1. MAE is defined as:

$$MSE = \frac{1}{|D_{test}|}\sum_{d=1}^{|D_{test}|}\left|O_{d^*} - O_d\right|$$

2. RMSE is defined as:

$$RMSE = \sqrt{\frac{1}{|D_{test}|}\sum_{d=1}^{|D_{test}|}(O_{d^*} - O_d)^2}$$

Where O_{d^*} denotes the ground-truth overall rating for review *d*, O_d denotes the prediction overall rating for review *d*. The smaller MSE or $RMSE$ value means a better performance.

Table 7. The differences of overall ratings predicted with ground-true overall ratings

Algorithm	MSE	$RMSE$
$PRank + \alpha_{(GlobalPrediction+LSAWs)}$	0.292	0.384
$PRank + \alpha_{PRR}$	0.260	0.363
$PRank + \alpha_{LRMNN}$	0.278	0.352

In Table 7 shows results of these evaluations. We see that the overall aspect weights are identified based on the combination of Global Prediction and LSAWs gives the worst result. For the metric MSE, the LRR gives the smallest differences of overall ratings predicted. For the metric $RMSE$, the our LRMNN gives the smallest mean. So with this metric it indicates the overall aspect weights (i.e. important aspects) are identified from our model LRMNN is the best.

6. Conclusion

In this paper, we have proposed a latent aspect mining framework for aspect based sentiment analysis from textual review data. The framework includes four main sub-tasks, i.e aspect term extraction, aspect category detection, aspect rating detection for each review, and indentify overall aspect weights for all reviews. Through experimental results, we see that aspect vectors represented by averaging word vectors are more effective than represented by a bag of word model. In most of the metrics used in experiment, our model LRMNN outperforms Global Prediction and LRR model.

References

[1] T. Wong, W. Lam, "Hot item mining and summarization from multiple auction web sites," in Proceedings of the 5th IEEE International Conference on Data Mining, IEEE Computer Society, 2005, pp. 797–800.

[2] W. Jin and H. H. Ho, “A novel lexicalized HMM-based learning framework for web opinion mining,” in 26th International Conference on Machine Learning, 2009, pp. 465–472.

[3] F. Li, C. Han, M. Huang, X. Zhu, Y.-J. Xia, S. Zhang, H. Yu, "Structure-aware review mining and summarization," in Proceedings of the 23rd International Conference on Computational Linguistics, 2010, pp. 653–661

[4] G. Ganu, N. Elhadad, A. Marian, "Beyond the stars: Improving rating predictions using review text content," in Proceedings of 12th International Workshop on the Web and Databases, WebDB, 2009, pp. 1–6.

[5] S. Kiritchenko, X. Zhu, C. Cherry, S. Mohammad, "Detecting aspects and sentiment in customer reviews," in: SemEval@COLING, The Association for Computer Linguistics, 2014, pp. 437–442.

[6] J. J. McAuley, J. Leskovec, D. Jurafsky, “Learning attitudes and attributes from multi-aspect reviews,” in: 12th IEEE International Conference on Data Mining, 2012, pp. 1020–1025

[7] X. Ding, B. Liu, P. S. Yu, “A holistic lexicon-based approach to opinion mining”, in: WSDM, ACM, 2008, pp. 231–240.

[8] W. Xu, Y. Tan, “Semi-supervised target-oriented sentiment classification”, Neurocomputing 337 (2019) 120 – 128

[9] B. Snyder, R. Barzilay, “Multiple aspect ranking using the good grief algorithm,” in: Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, April 22-27, 2007, Rochester, New York, USA, 2007, pp. 300–307.

[10] W. Wang, H. Wang, Y. Song, “Ranking product aspects through sentiment analysis of online reviews,” Journal of Experimental & Theoretical Artificial Intelligence 29 (2) (2017) 227–246.
[11] Y. Liu, J.-W. Bi, Z.-P. Fan, “Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory,” Information Fusion 36 (2017) 149 – 16.
[12] X. Li, L. Bing, P. Li, W. Lam, Z. Yang, “Aspect term extraction with history attention and selective transformation,” in: Proceedings of the 27th International Joint Conference on Artificial Intelligence, IJCAI’18, AAAI Press, 2018, pp. 4194–4200.
[13] J. Zhang, G. Xu, X. Wang, X. Sun, T. Huang, “Syntax-aware representation for aspect term extraction,” Advances in Knowledge Discovery and Data Mining, Springer International Publishing, Cham, 2019, pp. 123–134.
[14] S. Poria, E. Cambria, A. F. Gelbukh, “Aspect extraction for opinion mining with a deep convolutional neural network,” Knowl.-Based Syst. 108 (2016) 42–49.
[15] Chuhan Wu and Fangzhao Wu and Sixing Wu and Zhigang Yuan and Yongfeng Huang, “A hybrid unsupervised method for aspect term and opinion target extraction,” Knowl.-Based Syst. 148 (2018) 66–73.
[16] Huaishao Luo, Tianrui Li, Bing Liu, Bin Wang and Herwig Unger, “Improving Aspect Term Extraction With Bidirectional Dependency Tree Representation,” TASLP, 2019, 27(7):1201-1212.
[17] Xinjie Zhou, Xiaojun Wan* and Jianguo Xiao, “Representation Learning for Aspect Category Detection in Online Reviews”, in Proceedings of AAAI, 2015, pp.417-423.
[18] S Movahedi, E Ghadery, H Faili, A Shakery, “Aspect Category Detection via Topic-Attention Network,” CoRR abs/1901.01183 (2019)
[19] H.Wang, Y.Lu and Ch.Zhai, “Latent Aspect Rating Analysis on Review Text Data: A Rating Regression Approach,” In Proceedings of SIGKDD, 2010, pp.168-176.
[20] Y.Xu, T.Lin and W.Lam, “Latent Aspect Mining via Exploring Sparsity and Intrinsic Information,” In Proceedings of CIKM, 2014, pp.879-888.
[21] [21] Andrew L. Maas and Andrew Y. Ng, “A Probabilistic Model for Semantic Word Vectors,” In Proceedings of NIPS, 2011,
[22] Z.Zha, J.Yu, J.Tang, M.Wang and T.Chua, “Product Aspect Ranking and Its Applications,” IEEE Transactions on Knowledge and Data Engineering, vol. 26, no.5, pp.1211-1224, 2014.
[23] T. Mikolov, I. Sutskever, K. Chen, G. Corrado and J. Dean. “Distributed Representations of Words and Phrases and their Compositionality,” In Proceedings of NIPS, 2013, pp. 1-9.
[24] Y Bengio, R Ducharme, P Vincent, and C Jauvin, “A neural probabilistic language model,” Journal of Machine Learning Research, 3(6):1137–1155, 2003.
[25] W. Yih, K. Toutanova, J. Platt and C. Meek, “Learning Discriminative Projections for Text Similarity Measures,” Proceedings of the Fifteenth Conference on Computational Natural Language Learning, 2011, pp. 247-256.
[26] J Pennington, R Socher and CD Manning, “Glove: Global Vectors for Word Representation,” EMNLP, 2014, 1532-1543
[27] Sugato Basu, “Semi-supervised Clustering by Seeding,” In ICML, 2002.
[28] Y. Lu, C. Zhai, and N. Sundaresan, “Rated aspect summarization of short comments,” In Proceedings of WWW, 2009, pp. 131–140.
[29] K.Crammer and Y.Singer, “Pranking with ranking,” In Proceedings of NIPS, 2001, pp.641-647.
[30] Duc-Hong Pham and Anh-Cuong Le and Thi Kim Chung Le, “A Least Square based Model for Rating Aspects and Identifying Important Aspects on Review Text Data,” NICS, 2016, pp. 16-18.