AspeRa: Aspect-Based Rating Prediction Based on User Reviews

Elena Tutubalina[1,2,4], **Valentin Malykh**[1,3], **Anton Alekseev**[1], **Ilya Shenbin**[1], **Sergey Nikolenko**[1,4]
[1]St. Petersburg Department of the Steklov Mathematical Institute
27 Fontanka, St.Petersburg, Russian Federation, 191023
[2]Kazan Federal University, 18 Kremlyovskaya, Kazan, Russian Federation, 420008
[3]Moscow Institute of Physics and Technology, 9 bld. 7 Instituski per., Dolgoprudny, Russia, 141700
[4]Neuromation OU, Tallinn, 10111 Estonia
ElVTutubalina@kpfu.ru

Abstract

We propose a novel Aspect-based Rating Prediction model (AspeRa) that estimates user rating based on review texts for the items. It is based on aspect extraction with neural networks and combines the advantages of deep learning and topic modeling. It is mainly designed for recommendations, but an important secondary goal of AspeRa is to discover coherent aspects of reviews that can be used to explain predictions or for user profiling. We conduct a comprehensive empirical study of AspeRa, showing that it outperforms state-of-the-art models in terms of recommendation quality and produces interpretable aspects. This paper is an abridged version of our work (Nikolenko et al., 2019).

1 Introduction

As the scale of online services and the Web itself grows, recommender systems increasingly attempt to utilize texts available online, either as items for recommendation or as their descriptions. In this work, we introduce novel deep learning methods for making recommendations with full-text items, aiming to learn interpretable user representations that reflect user preferences and at the same time help predict ratings. We propose a novel Aspect-based Rating Prediction Model (AspeRa) for aspect-based representation learning for items by encoding word-occurrence statistics into word embeddings and applying dimensionality reduction to extract the most important aspects that are used for the user-item rating estimation. We investigate how and in what settings such neural autoencoders can be applied to content-based recommendations for text items.

2 AspeRa Model

The *AspeRa* model combines the advantages of deep learning (end-to-end learning, spatial text representation) and topic modeling (interpretable topics) for text-based recommendation systems. The model receives as input two reviews at once, treating both identically. Each review is embedded with self-attention to produce two vectors, one for author (user) features and the other for item features. These two vectors are used to predict a rating corresponding to the review. All vectors are forced to belong to the same feature space. The embedding is produced by the Neural Attention-Based Aspect Extraction Model (ABAE) (He et al., 2017). As in topic modeling or clustering, with ABAE the designer can determine a finite number of topics/clusters/aspects, and the goal is to find out for every document to which extent it satisfies each topics/aspects. From a bird's eye view, ABAE is an autoencoder. The main feature of ABAE is the reconstruction loss between bag-of-words embeddings used as the sentence representation and a linear combination of aspect embeddings. A sentence embedding is additionally weighted by *self-attention*, an attention mechanism where the values are word embeddings.

The first step in ABAE is to compute the embedding $\mathbf{z}_s \in \mathbb{R}^d$ for a sentence s; below we call it a text embedding: $\mathbf{z}_s = \sum_{i=1}^n a_i \mathbf{e}_{w_i}$, where $\mathbf{e}_{w_i}$ is a word embedding for a word w_i, $e \in \mathbb{R}^d$. As word vectors the authors use *word2vec* embeddings trained with the skip-gram model (Mikolov et al., 2013b). Attention weights a_i are computed as a multiplicative self-attention model: $a_i = \text{softmax}(\mathbf{e}_{w_i}^\top A \mathbf{y}_s)$, where $\mathbf{y}_s$ is the average of word embeddings in a sentence, $\mathbf{y}_s = \sum_{i=1}^n \mathbf{e}_{w_i}$, and $A \in \mathbb{R}^{d \times d}$ is the learned attention model. The second step is to compute the aspect-based sentence representation $\mathbf{r}_s \in \mathbb{R}^d$ from an

Table 1: Two sets of AspeRa hyperparameters (for models with different initialization strategies).

Settings	AspeRa (GloVe)	AspeRa (SGNS)
Embeddings	GloVe	SGNS
Optimization alg.	Adam	Adam
# aspects	11	10
Hidden layer dim.	256	64
# epochs	20	18
# words per sample	256	224

Table 2: Performance of text-based and collaborative rating prediction models.

Model	MSE	
	Instant Videos	Toys & Games
NMF	0.946	0.821
DeepCoNN	0.943	0.851
Attn+CNN	0.936	-
SVD	0.904	0.788
HFT	0.888	0.784
TransRev	0.884	0.784
NARRE	-	0.769
AspeRa (GloVe)	0.870	0.730
AspeRa (SGNS)	**0.660**	**0.571**

aspect embeddings matrix $T \in \mathbb{R}^{k \times d}$, where k is the number of aspects: $\mathbf{p}_s = \text{softmax}(W\mathbf{z}_s + \mathbf{b})$, where $\mathbf{p}_s \in \mathbb{R}^k$ is the vector of probability weights over k aspect embeddings, $\mathbf{r}_s = T^\top \mathbf{p}_s$, and $W \in \mathbb{R}^{k \times d}$, $\mathbf{b} \in \mathbb{R}^k$ are the parameters of a multi-class logistic regression model.

The proposed model's architecture includes an embedder, which provides text and reconstruction embeddings for an object similar to ABAE. The intuition behind this separation of user and item embedding is as follows: there are some features (aspects) important in an item for a user, but the item also has other features. Hence, we want to extract user aspects from a user's reviews as well as item aspects from an item's reviews. The resulting embedding is conditioned on aspect representation of the reviews. The model contains four embedders in total, one pair of user and item embedders for two reviews being considered at once. First each review is paired with another review of the same user, grouping by users and shuffling the reviews inside a group; then with another review of the same item. Thus, the training set gives rise to only twice as many pairs as reviews available for training. The rating score for the first review in a pair is used to train the rating predictor (*MSE*). There are two losses in *AspeRa*: MSE for rating prediction and MaxMargin loss to put user and item embeddings in the same space. The MSE loss assumes that rating is predicted as the dot product of user and item embeddings for a review: $MSE = \frac{1}{N}\sum_{j=1}^{N}(\mathbf{z}_j^{u\top}\mathbf{z}_j^i - r_j)^2$, where $\mathbf{z}_j^u$ is a text embedding for the author of review j, $\mathbf{z}_j^i$ is a text embedding for the item j is about, and r_j is the true rating associated with j. Max-margin loss aims to project all user and item embeddings into the same feature (aspect) space.

3 Experimental Evaluation

We evaluated the proposed model on *Amazon Instant Videos 5-core reviews* and *Amazon Toys and Games 5-core reviews* (He and McAuley, 2016; McAuley et al., 2015). We randomly split each dataset into 10% test set and 90% training set, with 10% of the training set used as a validation set for tuning hyperparameters. We compare two word embeddings for *AspeRa*: *GloVe* (Pennington et al., 2014) and *word2vec* (Mikolov et al., 2013a; Mikolov et al., 2013c) based on the train set of reviews (see Table 1).

We evaluate the performance of *AspeRa* in comparison to state-of-the-art models: NMF (Zhang et al., 2006), DeepCoNN (Zheng et al., 2017), Attn+CNN (Seo et al., 2017), SVD (Koren et al., 2009), HFT (McAuley and Leskovec, 2013), NARRE (Chen et al., 2018), and TransRev (Garca-Durn et al., 2018). Table 2 compares the best Mean Square Error (MSE) of *AspeRa* and other models for rating prediction. Results of existing models were adopted from (Garca-Durn et al., 2018; Chen et al., 2018) with the ratio 80:10:10. Note that while *AspeRa* with generic *GloVe* word embeddings still works better than any other model, adding custom word embeddings trained on the same type of texts improves the results greatly.

4 Conclusion

We have introduced a novel approach to learning rating- and text-aware recommender systems based on ABAE, metric learning, and autoencoder-enriched learning. Our approach jointly learns interpretable user and item representations. It is expectedly harder to tune to achieve better quality, but the final model performs better at rating prediction and almost on par at aspects coherence with other state-of-the-art approaches. Our results can also be viewed as part of the research effort to analyze and interpret deep neural networks, a very important recent trend (Kádár et al., 2017; Radford et al., 2017).

Acknowledgements. This research was done at the Samsung-PDMI Joint AI Center at PDMI RAS and was supported by Samsung Research.

References

Chong Chen, Min Zhang, Yiqun Liu, and Shaoping Ma. 2018. Neural attentional rating regression with review-level explanations. In *Proceedings of the 2018 World Wide Web Conference*, WWW'18, pages 1583–1592, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.

Alberto Garca-Durn, Roberto Gonzalez, Daniel Ooro-Rubio, Mathias Niepert, and Hui Li. 2018. Transrev: Modeling reviews as translations from users to items. *CoRR*, abs/1801.10095.

Ruining He and Julian McAuley. 2016. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In *proceedings of the 25th international conference on world wide web*, pages 507–517. International World Wide Web Conferences Steering Committee.

Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2017. An unsupervised neural attention model for aspect extraction. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 388–397.

Akos Kádár, Grzegorz Chrupała, and Afra Alishahi. 2017. Representation of linguistic form and function in recurrent neural networks. *Computational Linguistics*, 43(4):761–780.

Yehuda Koren, Robert M. Bell, and Chris Volinsky. 2009. Matrix factorization techniques for recommender systems. *IEEE Computer*, 42(8):30–37.

J. McAuley and J. Leskovec. 2013. Hidden factors and hidden topics: understanding rating dimensions with review text. In *Proceedings of the 7th ACM conference on Recommender systems*, pages 165–172.

Julian McAuley, Christopher Targett, Qinfeng Shi, and Anton Van Den Hengel. 2015. Image-based recommendations on styles and substitutes. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 43–52. ACM.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013b. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013c. Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 3111–3119. Curran Associates, Inc.

Sergey I. Nikolenko, Elena Tutubalina, Valentin Malykh, Ilya Shenbin, and Anton Alekseev. 2019. Aspera: Aspect-based rating prediction model. In Leif Azzopardi, Benno Stein, Norbert Fuhr, Philipp Mayr, Claudia Hauff, and Djoerd Hiemstra, editors, *Advances in Information Retrieval*, pages 163–171, Cham. Springer International Publishing.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar. Association for Computational Linguistics.

Alec Radford, Rafal Jozefowicz, and Ilya Sutskever. 2017. Learning to generate reviews and discovering sentiment. *arXiv preprint arXiv:1704.01444*.

Sungyong Seo, Jing Huang, Hao Yang, and Yan Liu. 2017. Representation learning of users and items for review rating prediction using attention-based convolutional neural network. In *3rd International Workshop on Machine Learning Methods for Recommender Systems (MLRec)(SDM17)*.

Sheng Zhang, Weihong Wang, James Ford, and Fillia Makedon. 2006. Learning from incomplete ratings using non-negative matrix factorization. In *Proceedings of the 2006 SIAM international conference on data mining*, pages 549–553. SIAM.

Lei Zheng, Vahid Noroozi, and Philip S Yu. 2017. Joint deep modeling of users and items using reviews for recommendation. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pages 425–434. ACM.