Automatic Generation of Natural Language Explanations

Felipe Costa
Aalborg Universitet
Aalborg, Denmark
fcosta@cs.aau.dk

Sixun Ouyang
Insight Centre for Data Analytics
University College Dublin, Ireland
sixun.ouyang@insight-centre.org

Peter Dolog
Aalborg Universitet
Aalborg, Denmark
dolog@cs.aau.dk

Aonghus Lawlor
Insight Centre for Data Analytics
University College Dublin, Ireland
aonghus.lawlor@insight-centre.org

ABSTRACT
An interesting challenge for explainable recommender systems is to provide successful interpretation of recommendations using structured sentences. It is well known that user-generated reviews, have strong influence on the users' decision. Recent techniques exploit user reviews to generate natural language explanations. In this paper, we propose a character-level attention-enhanced long short-term memory model to generate natural language explanations. We empirically evaluated this network using two real-world review datasets. The generated text present readable and similar to a real user's writing, due to the ability of reproducing negation, misspellings, and domain-specific vocabulary.

ACM Classification Keywords
I.2.7 Natural Language Processing: Language generation; I.5.1 Models: Neural nets

Author Keywords
Recommender systems; Natural Language Generation; Explainability; Explanations; Neural Network

INTRODUCTION
A recommender system should provide accurate and relevant recommendations, but a good recommendation must be supported by interpretation. The explanation is the key factor to gain the trust of the user. An interpretable system has significant influence on a user's decision [6], and users tend to trust the opinion of others, especially when they describe personal experiences [3].

Current explainable recommendations propose to mine user's reviews to generate explanations. Nonetheless, they lack generating natural language expressions, hence the sentences are produced in a modular way. We aim to generate natural language explanations from reviews, aligning explanations and textual features such as aspects and sentiments, which influence the recommendation of different items. We exploit deep neural networks at character-level to generate explanations. These networks have recently shown good performance to generate sentences as presented by Karpathy *et. al.*, who use a variant of LSTM cells to generate text [2]. Karpathy *et. al.* presented encoding rating vectors of reviews in the training phase, allowing the system to calculate the probability of the next character based on the given rating. Later, Dong *et. al.* presented an efficient method to generate the next word in a sequence when it is added an attention mechanism, improving the performance for long textual sequences [1].

In this paper, we propose a character-level attention-enhanced long short-term memory (LSTM) model to generate personalized natural language explanations based on user-generated reviews. The model is trained using two real-world datasets: BeerAdvocate [5] and Amazon book reviews [1]. The datasets present user reviews describing their opinion about items in natural language. The explanations are adaptively composed by an encoder-side context vector, because our model learns soft alignments between generated characters and user-item relations, for example, ratings from a user to an item.

INTERPRETATION MODEL
The character-level explanation model presents (1) three modules: LSTM network, attention layer, and generator module; and (2) two input sources: review text and concatenated word embeddings of user, item, and rating, as presented in Fig. 1. First, users and items embeddings are learned from *doc2vec* model, where characters of reviews are encoded as one-hot vectors, corresponding to the input time-steps of LSTM network. Second, the embeddings are concatenated with the outputs of LSTM, which are inputs for the following attention layer. Finally, the generator module produce sentences as explanations using outputs from the attention layer.

- **LSTM network** LSTM is an enhanced recurrent neural network (RNN) where information is transmitted from a neuron to the next neuron, and the corresponding neuron in the next layer simultaneously, as presented in Fig. 1. LSTM

IUI'18, Mar 07–11, 2018, Tokyo, Japan
ACM 978-1-4503-5571-1/18/03...$15.00
DOI: https://doi.org/10.1145/3180308.3180366

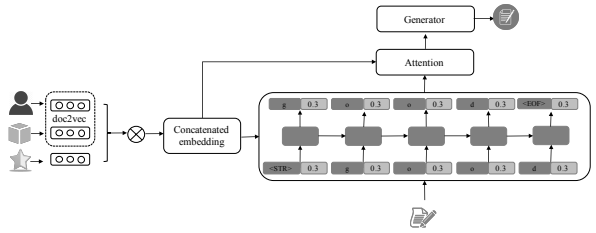

Figure 1. Personalized Explanation Generation Model Architecture

Rating	Text
1	i was not a little to read the first book, i did not like the story. i would not recommend it.
2	i was not interested with the story line and the story was a little slow.
3	the characters are always good. it was a good story.
4	i love the story, i would recommend this book to anyone
5	i love the story and the story line. i would recommend it to anyone who want to read the next book.

Figure 2. Rating Text Samples, from poorly rated (1) to highly rated (5).

was introduced to solve the long-term dependency problem, which causes vanishing gradient in conventional RNN[2].

- **Attention mechanism** The attention mechanism, adaptively learns soft alignments c_t between character dependencies H_t and attention inputs a. Eq. 1 formally defines the new character dependencies using attention layer $H_t^{attention}$ [1].

$$c_t = \sum_i^a \frac{\exp(\tanh(W_s \odot [H_t, a_i]))}{\sum \exp(\tanh(W_s \odot [H_t, a_i]))} a_i$$
$$H_t^{attention} = \tanh(W_1 \odot c_t + W_2 \odot H_t) \quad (1)$$

- **Generating Text** The explanation is generated character by character. The characters are given by maximizing the *softmax* conditional probability p, based on the new character dependencies $H_t^{attention}$ [1], as presented in Eq. 2

$$p = \text{softmax}(H_t^{attention} \odot W + b), \quad char = \arg\max p \quad (2)$$

RESULTS

The model was evaluated using two real-world datasets: BeerAdvocate and Amazon book reviews. The first experiment presents generated explanations given by the rating as attention mechanism to generate explanations with different sentiments, as presented in Fig. 2.

The second experiment generates explanations for particular user-item pairs presented in Fig. 3, where the user opinion about an item is generated in natural language. Finally, evaluating the generated explanations based on readability metrics in Fig. 4. The readability metrics [4] measure how understandable the generated text is, where lower values correspond to an easy and understandable text.

CONCLUSION

The work provides preliminary results in automatically generating natural language explanations. The model differs from recent works [1, 6], due to the use of attention layer combined with character-level LSTM. The proposed model improves the performance and allow to generate more accurate and trustworthy explanations aligned to the user's taste.

We would like to improve the model considering: (1) personalizing explanations to benefit the users' preferences based on their expressed sentiments; and (2) testing the model in larger and more varied review domains such as hotels and restaurants.

Dataset	(User, Item)	Explanation
Amazon Books	(9163, 11021)	i love this series. i can't wait for the next book. i love the characters and the story line. i was so glad that the story was a little longer. i would recommend this book to anyone who enjoy a good mystery.
BeerAdvocate	(shivtim, 2023)	poured from a bottle into a pint glass. a: pours a dark brown with a small head. s - smells of caramel and chocolate. t - a bit of a caramel malt and a little bit of coffee. m- medium body with a solid carbonation. d - medium bodied with a smooth mouthfeel. i can taste the sweetness and a bit of caramel and a little bit of a bit of alcohol.

Figure 3. Generated Text Samples

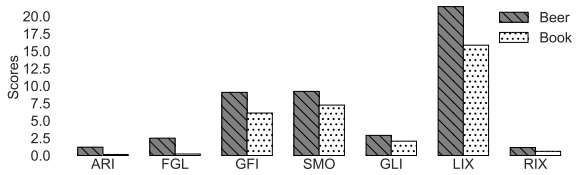

Figure 4. Readability Score of Explainations

ACKNOWLEDGMENTS

This work is supported by Science Foundation Ireland through the Insight Centre for Data Analytics under grant number SFI/12/RC/2289, and Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq (grant# 206065/2014-0).

REFERENCES

1. Li Dong, Shaohan Huang, Furu Wei, Mirella Lapata, Ming Zhou, and Ke Xu. Learning to generate product reviews from attributes *(EACL'17)*.
2. Andrej Karpathy, Justin Johnson, and Fei-Fei Li. 2015. Visualizing and Understanding Recurrent Networks. *CoRR* abs/1506.02078 (2015).
3. Bart P Knijnenburg, Martijn C Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. 2012. Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction* 22, 4-5 (2012), 441–504.
4. Suraj Maharjan, John Arevalo, Manuel Montes, Fabio A González, and Thamar Solorio. 2017. A Multi-task Approach to Predict Likability of Books *(EACL'17)*, Vol. 1. 1217–1227.
5. Julian John McAuley and Jure Leskovec. 2013. From Amateurs to Connoisseurs: Modeling the Evolution of User Expertise Through Online Reviews *(WWW '13)*. ACM, 897–908.
6. Sungyong Seo, Jing Huang, Hao Yang, and Yan Liu. 2017. Interpretable Convolutional Neural Networks with Dual Local and Global Attention for Review Rating Prediction *(RecSys '17)*. 297–305.