Persuasively Explainable Recommendation

Submission

ABSTRACT

In the e-commerce environment, it is a personalized service to recommend suitable products according to the shopping scene of the user. We hope to provide users with an explainable and persuasive recommendation reason when recommending products, that is, why they should recommend this product, and motivate online purchasers to make successful purchases through persuasive descriptions. In this contribution we present our system by combinning weak supervision frame with generate model. We first select the persuasive sentences from corpus as the training data of our generate model through weak supervision. Then through our proposed model yield persuasive sentence. We conduct comprehensive experiments on real sets. Compared with state-of-the-art methods, our framework produces sentences with higher ROUGE and BLEU scores and more attractive and persuasive.

KEYWORDS

persuasive, explainable,recommendation

ACM Reference Format:

Submission . 2018. Persuasively Explainable Recommendation. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

To provide accurate recommendation, state-of-the-art recommendation systems depend on complex machine learning models (i.e. deep neural networks) on heterogeneous data sources. These models function as black boxes, which is a major obstacle to improving user experience and increasing user stickness. Aiming to overcome this obstacle, there is growing interest in devoloping *explainable recommendation* systems to provide intuitive explanations of the recommendation results. In particular,motivated by the success of promoting sales using manually-written recommendation reasons, recently many E-commerce sites such as Alibaba and Amazon invest heavily in providing automated recommendation explanations [].

Explanations can benefit recommender systems on a number of aspects. As defined in [**?**], there are seven categories of recommendation explanations. From the viewpoint of E-commerce sites, *pursuasiveness* and *effectiveness* are the two most desirable aspects. *Persuasiveness* is the ability to convinces users to buy the recommended items. *Effectiveness* is the power to help users make better decisions. As these are properties that directly associat to conversion rates, an industry-level explainable recommendation system must focus on generating persuasive and effective explanations.

Existing explanable

Persuasive explanation For example, explaining how the system works can in Pursuasive explanation If users understand how the item suits under different contexts, he/she will make better decisions.

On one hand, existing work on explainable recommendation, pursuasiveness, mostly on enhancing transparency

On the other hand, there is a recent trend on generating pursuasive texts,

To the best of our knowledge, we are the first. input: user behavior.

relavance

challenge (1) lack of training data (2) relevance (3) pursuasive

(1) weak supervision (2)

contribution (1) model (2) application (3) evaluation

With the rapid development of e-commerce, more and more customers choose to purchase goods on the Internet. In mass goods, how to attract customers' attention through product description and strive for customers' stay time is a means to stand out from a large number of merchants. At present, the generation of product descriptions mainly relies on manual generation, and there is a problem that the amount of products covered is small and the cost is high. In recent years, deep learning has made breakthroughs in many fields such as image, natural language processing, and information retrieval. If we can generate persuasive product descriptions through deep learning, we can save costs and also improve the quality of content. Because a person's knowledge is limited, the generated vocabulary is limited. By learning a large amount of data, the machine-generated content can far exceed one's information.

Two challenges arise in generating persuasive and explainable recommendations for recommended products automatically .

The first challenge lies in getting training data. It is almost impossible to get a tag corpus, the data is a description of the product, and the tag is whether these descriptions are persuasive. Only the corpus contains descriptions of the products, but there is no guarantee that these sentences are persuasive. We need to choose persuasive sentences from these corpus as our training set.

The second challenge is related to the description of the scene. In our corpus, there are many sentences with only product descriptions, no descriptions of related scenes, but with the description of the scene is the sentence we really want. We need to design new module for this purpose.

Our goal is to generate persuasive recommendation reason containing scene descriptions when recommending products. To solve the first challenge, we use a weak supervision framework to select persuasive sentences. Weak supervision frameword mark the data without the user to manually mark any training data. In this way, we do not cost time to label mass of data manually.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00

https://doi.org/10.1145/nnnnnnn.nnnnnnn

To solve the second challenge, we propose a global local module: a specific module that focuses on the description of the scene. It has two parts, the global part is to learn the text description of all products on all texts, and the local part is to learn the scene-specific description through the local module. Through the weighted sum of these two parts, the generated sentence also has a description of the scene while describing the product.

In academia, explainable recommendation belongs to a broader branch of explainable AI, which is an important and urgent issue in the whole artificial intelligence community.

Our contributions are three folds. (1) We take the problem of generating persuasive descriptions as sequence to sequence model . The source text are scene and attribute about the product, the target text are persuasive description within scene about the product. (2) we use a weak supervision framework to select persuasive sentences as the training set of the model, which avoids a lot of manual labeling work. (3) Our system introduces a global-local module, so that the automatically generated description sentences for products can be described with scenes.

This paper is organized as follows. We briefly survey the related work in Sec. 2. In Sec. 3, we first introduce the architecture of our system and describe the weak supervision and our global-local-copy model. We present and analyze the experimental results on a real data set in Sec. 4. We conclude our work and suggest future directions in Sec. 5.

2 RELATED WORK

2.1 Explainable Recommendation

In recent years, there have been many studies on natural language processing, including poetry creation [? ? ? ? ? ?],machine translation [? ? ?], and so on. A persuasive and explainable recommendation reason for the generation of goods, similar to the above two questions, is a sequence model. In fact, most ML research usually focuses on predictive tasks, but rarely provides explanations for them. However, many customers need to rely on the recommendation reasons provided by the system to be firm in their determination to purchase the item.

The traditional explainable recommendation method has content-based recommendation and collaborative filtering. [? ?] used tags as content features for recommendations and generate corresponding explanations. Collaborative filtering is divided into user-based collaborative filtering and item-based collaborative filtering. [? ? ?] provided an explanation using the ratings of its neighbors or its similar items. In recent years, many models have attempted to use textual information to generate textual explanations for recommendations. One is to extract commodity features as an explanation, such as [?] used topic models to explain recommendations, [?] found the related entities and gave conceptual explanation, the other is to generate complete sentences as an explanation, such as [?] generated textual explanation based on fixed templates, [?] based on linguistic creativity to generate various forms of sentences.

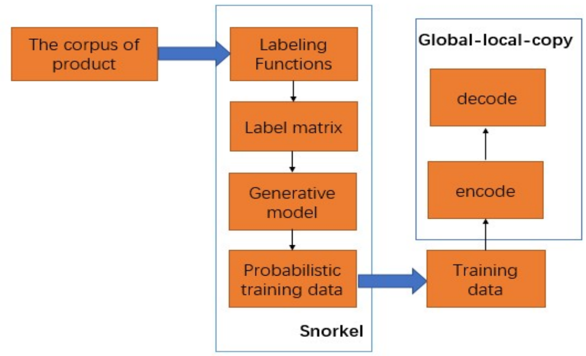

Figure 1: System Architecture

2.2 Creative Text Generation

3 SYSTEM ARCHITECTURE

We take the corpus of product as the input of our system, and produces a persuasive sentence with the scene description of the product. Fig. 1 shows an overview of the proposed system architecture with two major steps: (1) Given the corpus of product,the first step is to select the persuasive sentences as the training data of our model, (2) identify the scene name, product name, cpv data [1] from the selected sentence as the input of our model,then yield persuasive sentence. In this section, we first introduce the weak supervision method for selecting the persuasive sentences. We then present our global-local-copy model in detail.

3.1 Resources Used

We use the list of product recommendation reasons as our dataset. The corpus are generated by high quality person.But the quality of the original dataset is far from ideal, there are many recommended reasons are even the original title of the product, so we need to filter the training data.

3.2 Weak Supervision

Manual labeling is very time consuming, so we use the Snorkel [?] weak supervision method to mark the data without the user to manually mark any training data. Rather than hand-labeling training data, users of Snorkel write labeling functions(LF), which allow them to express various weak supervision sources such as patterns, heuristics, external knowledge bases, and more. We wrote ten labeling functions based on the characteristics of persuasive sentences, is shown in Tab. 1. Among them, the labels of the first five functions are positive and the rest are negative.

Next, Snorkel automatically learns a generative model over the labeling functions,the output of Snorkel is a set of probabilistic labels. The statistics about the resulting label matrix is shown in Tab. 2. **Coverage** is the fraction of candidates that the labeling function emits a non-zero label for. **Overlap** is the fraction candidates that the labeling function emits a non-zero label for and that another labeling function emits a non-zero label for. **Conflict** is the

[1] Cpv is a collection of values of the attributes of the product. Here, only the value of the product attributes is in the sentence it can be extracted.

Table 1: Labeling Functions

Labeling Functions	Description
is_neat	Sentence is neat
has_modal	Sentence has modal particle
four_word	Sentence contains a four-word structure
dot_word	The comma is followed by "让/使/为/给" or verbs
end_exclamation	Sentence ends with an exclamation point
no_adj_and_adv	Sentence has no adjectives and adverbs
other_words	Sentence contains characters other than Chinese, English, numbers, and specified symbols (。，？！、；：).
tree_depths	the depth of the dependency tree is greater than 10
clause_num	the number of clauses is greater than 10
token_num	the number of word segments is greater than 10

Table 2: Statistics about the resulting label matrix

LFs	Coverage	Overlaps	Conflicts
is_neat	0.075185	0.060664	0.040715
has_modal	0.022520	0.019763	0.004743
four_word	0.418368	0.333411	0.061301
dot_word	0.607374	0.411911	0.118444
end_exclamation	0.070130	0.061328	0.010403
no_adj_and_adv	0.113256	0.086246	0.063460
other_words	0.060238	0.052460	0.049377
tree_depths	0.004969	0.004637	0.004564
clause_num	0.022300	0.022194	0.022154
token_num	0.103537	0.077849	0.056611

fraction candidates that the labeling function emits a non-zero label for and that another labeling function emits a conflicting non-zero label for. We choose sentences with probabilistic labels are bigger than 0.5 and the words are less than 50 as the training set of our model.

3.3 Background: Transformer

Transformer [?] is a network architecture based solely on an attention mechanism, dispensing with recurrence and convolutions entirely. Transformer have an encoder-decoder structure and both the encoder and decoder are composed of a stack of $N = 6$ identical layers.

Encoder: Each layer has two sub-layers. The first is a multi-head self-attention mechanism, and the second is a fully connected feed-forward network.

Decoder: In addition to the two sub-layers in each encoder layer, the decoder inserts a third sub-layer, which performs multi-head attention over the output of the encoder stack.

3.4 Global-Local-Copy Model

Global-Local-Copy model is comprised of three modules which is based on Transformer architecture. Fig. 2 illustrates the detailed model structure. Global-Local-Copy model is also with encoder-decoder structure. The encoder consists of global module and local module and the copy module is in decoder part.

Our goal is to generate persuasive sentences with scene descriptions based on the scenes, products, and attributes given by the user. In our training set, some sentences have only product descriptions, no descriptions of related scenes, and some sentences are product descriptions in different scenarios. We use a global module to learn text descriptions of all products on all texts and learn scene-specific description through local modules. We want the output sentence to contain user-supplied input, so we also add the copy module to our model.

Encoder: We produce a global encoding H^{global} of X using a global encode part of Transformer and the local encoding is H^{local}. The outputs of the two modules are combined through a mixture layer to yield a global-local encoding H of X. The left of Fig. 2 illustrates the global-local modules encoder.

$$\mathbf{H} = \beta^s \mathbf{H}^{local} + (1 - \beta^s)\mathbf{H}^{global}. \tag{1}$$

Here, the scalar β is a learned parameter between 0 and 1 that is specific to the scenario s.

Decoder: The copy module is in decoder module, the probability of generating any target word y_t, is given by the mixture of probabilities as follows

$$p(y_t) = p(y_t, g) + p(y_t, c) \tag{2}$$

where g stands for the generate-mode, and c the copy mode. the right of Fig. 2 illustrates the copy module decoder. H is global-local encoding the above-mentioned, $\zeta(y)$ is the weighted sum of hidden states H corresponding to y, referred to as selective read in the right of Fig. 2.

$$\zeta(y) = \sum_{\tau=1}^{T} \rho_\tau \mathbf{h}_\tau \tag{3}$$

$$\rho_\tau = \begin{cases} \frac{1}{K} p(x_\tau, \mathbf{c}|\mathbf{H}), & x_\tau = y_t \\ 0, & otherwise \end{cases} \tag{4}$$

where K is equal to the number of positions with source keywords in the target sentence, τ is the index of source keywords, T is the number of keywords, t is the index of word in target sentence, and $p(x_\tau, \mathbf{c}|\mathbf{H})$ is the probability of the source keyword be copied in target sentence.

The score of each mode is calculated:

Generate-Mode: first connect the output of the feed forward part of the transformer method and selective-read, and then $p(y_t, g)$ is calculated through the full connection.

Copy-Mode: first calculate $\sigma(\mathbf{HW})$, σ is a non-linear activation function, here using the *tanh* function. Next $p(y_t, c)$ is calculated through the full connection.

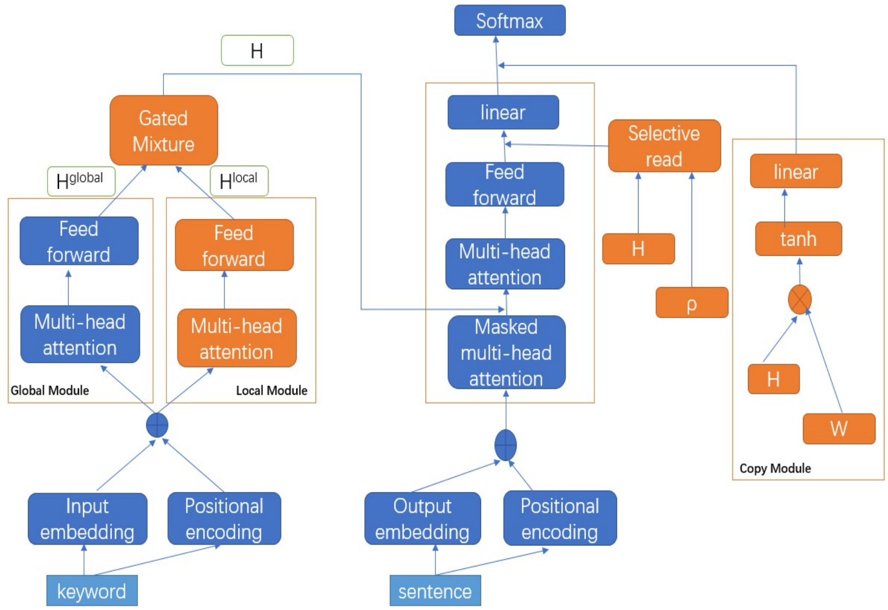

Figure 2: Global-Local-Copy Model

Table 3: Training data format

Input	Output
创意，纸巾盒，欧式	一款欧式风范榉木纸巾盒，盒身采用创意撞色设计，不仅能放杂物，还能作为桌面摆设，大中小三种尺寸可选，适合多种场合使用。

Table 4: Manual Evaluation

Evaluation Metric	Description
Fluency [?]	Does the sentence read smoothly and fluently?
Catchyness [?]	Is the description attractive,catchy?
Relatedness [?]	Is the description semantically related to the target scene?
Completeness	Is the description contains the corresponding scene, product and attribute?
Informative	Is the description informative?

4 EXPERIMENTAL SETUP

4.1 Dataset

In this paper,we focus on two sub-scenarios under the home: creative home and simple home. We select the description of the products in these two scenarios from the list of product recommendation reasons. We collected 150,743 sentences related to these products, after weak supervision, left 103,612 sentences. We chose sentences which keyword input only appears once as the test set. Training data format is shown in Tab. 3.

4.2 Comparative Method

4.3 Training

We take the words from source side of corpus as the input vocabulary and chose the words from target side of corpus which word frequency greater than 20 as the output vocabulary. The dimension of word embedding and hidden units are both 512,the minibatch was set to be 64. The parameter of global-local module β is initialized by 0.5, the parameter W in copy module is randomly initialized and the the parameter p is initialized by zero.

4.4 Evaluation

There are no direct evaluation metrics so that evaluate text generation system is difficult. We choose ROUGE [?] and BLEU [?] metrics that are popularly used for generation tasks (especially Machine Translation and Summarization). These two metrics are both based on references, but there are thousands of ways to generate an appropriate sentence for a specific product,the limited references are impossible to cover all the correct results. So,we we use five evaluation standards for human evaluators to check the quality of the generated descriptions on a small test dataset of 30 instances. The manual evaluation metrics are listed in Tab. 4. The score of each manual evaluation metrics ranges from 0 to 5 with the higher score the better, see Tab. 5 for more detailed Grading Rules. All the generated sentences are evaluated by 5 experts and the rating scores are averaged as the final score.

Table 5: Manual Evaluation details

Evaluation Metric	Score	Description
Fluency	0	Not at all smooth
	1-4	how many places are not smooth minus how many points
	5	Very smooth
Catchyness	0-5	The ratio of attractive words in total words multiply by 5
Relatedness	0	Completely unrelated to the scene
	1	none
	2	Refer to the scene
	3-5	how many descriptions related to the scene, add how many points
Completeness	0	No input at all
	1	none
	2	Contains an input keyword
	3	Contains two input keyword
	4	There's no third word involved, but it's relevant
	5	Completely contains
Informative	0	No information at all
	1	It's describing the product
	2-5	how much information about the product, add how many points

4.5 Results

We report the experimental results for our two approaches, i.e. global-local model and global-local-copy model. The difference between two models is former has no copy module. We compare our models with the Transformer method. Results are reported on the test data of 1472 instances, used for automatic evaluation and a held-out set of 32 instances, used for manual evaluation. The source keyword of test data for automatic evaluation are never appeared in train data. We choose the source keyword of test data that have scene name, product name and only one cpv value for manual evaluation.

From the perspective of considering our system as another machine translation system that converts some keywords of product(the scene name, product name, cpv data) into a persuasive product description with scene, we have results shown in Tab. 6. Popular machine translation and summarization metrics BLEU and ROUGE are used. There are four different ROUGE measures: ROUGE-N, ROUGE-L, ROUGE-W, and ROUGE-S, depending on the textual units to be compared. As can be seen from the results, our two methods are superior to Transformer in every indicator. Explain that both the global-local module and the copy module have a positive impact on the model. Because these two metrics are both based on references, and the copy module is aim to let the output sentence contain user-supplied input, so the results of global-local-copy model is better than global-local model.

From the perspective of human psychology of persuasive product descriptions, we manually evaluated the generated descriptions

Table 6: Automatic Evaluation Metrics

Metrics	Transformer	Global-local	Global-local-copy
ROUGE-1	0.3933	0.4050	0.4054
ROUGE-2	0.1319	0.1446	0.1488
ROUGE-3	0.0643	0.0740	0.0777
ROUGE-4	0.0424	0.0514	0.0521
ROUGE-L	0.3259	0.3373	0.3423
ROUGE-W	0.1491	0.1552	0.1585
ROUGE-S*	0.1628	0.1762	0.1784
BLEU-1	0.2964	0.3056	0.3096
BLEU-2	0.1556	0.1671	0.1729
BLEU-3	0.0807	0.0926	0.0977
BLEU-4	0.0522	0.0632	0.0654

Table 7: Manual Evaluation Metrics

Metrics	Transformer	Global-local	Global-local-copy
Catchyness	1.2235	1.2455	1.3320
Relatedness	2.4375	2.5000	2.7500
Fluency	3.4375	3.7187	3.9375
Completeness	3.6250	3.9375	3.9062
Informative	3.0312	3.4687	3.4687

using human evaluators. Five different measures were used to evaluate the human subjectiveness: Catchyness, Relatedness, Fluency, Completeness and Informative. It can be evidently observed in Tab. 7. that the proposed system generated more catchy, better related, more fluency sentences compared to the Transformer method. Because our global-local module focuses on the description of the scene, resulting in the generated sentences with more descriptions of the scene, more appealing and more relevant to the scene. What's more, sentences generated by our model contain more input keywords and have more information about product.

For qualitative analysis, we also provide the sentences generated from our system as well as other systems in the Tab. 8. As we can see, the descriptions generated by our systems are competitive or better in terms of creativity, persuasiveness and fluency than the supervised baselines but have less overlap with the reference descriptions. This explains why our system is deemed to have underperformed than the baselines, as per the automatic evaluation scores. In general, the field of creative text generation demands looking beyond simplistic evaluation measures and it is about time that trainable metrics for evaluating persuasive text holistically, including aspects on creativity, coherency, novelty are proposed.

5 CONCLUSION

6 ACKNOWLEDGMENTS

Table 8: Sample generations from different systems along with inputs and reference descriptions

Input	创意，挂钟，奢华
Transformer	创意十足的挂钟，舒适静音的设计，温柔的花纹，灵动而神秘，让你爱坐在客厅的时光里里静静享受质量。
Global-local	创意挂钟，奢华镶钻，奢华镶钻，奢华镶钻。
Global-local-copy	创意十足的大号挂钟，奢华范，奢华独特。
Input	简约，挂钟，精致
Transformer	简约静音挂钟，做工精致，细节精致，高档品质之选。
Global-local	可摇摆的静音挂钟，做工精致，造型独特，简约大气的外形符合你的工作品质生活，静音设计，增加家中的灵动性。
Global-local-copy	这款挂钟，造型简约大方，做工精致，散发着大自然的气息，选用的静音扫描机芯，走时准确，可挂在墙上，方便又不掉色。