Towards Silver-tongued Persuasive Recommendation Explanations

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, text generation

ACM Reference Format:
Submission . 2019. Towards Silver-tongued Persuasive Recommendation Explanations. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The sales department in any organization plays a pivotal role in the success of business. With the power of persuasion and deep product knowledge, salespeople bridge the gap between customer needs and the products that fulfill the needs. We've seen many vivid examples of top salesperson who single-handedly boost company-wise sales. This brings an interesting yet challenging question: *Can machines function like a skilled salesman?*

In E-commerce, recommendation systems have made remarkable achievements by accurately predicting user needs and identifying demanded products. State-of-the-art recommendation systems depend on complex machine learning models on heterogeneous data sources. These models function as black boxes, which is a major obstacle towards efficient selling. Aiming to overcome this obstacle, there is growing interest in developing *explainable recommendation* systems to provide intuitive explanations of the recommendation results. In particular, recently many E-commerce sites such as Alibaba and Amazon invest heavily in providing automated recommendation explanations [**?**], motivated by the success of promoting sales using manually-written recommendation reasons. However, these explainable systems are still in the initial stage.

No behavior on earth is more human than selling. To improve user experience and increase user stickness, it is necessary to embody the best sales techniques in explainable recommendations: persuasiveness and knowledge-awareness.

A good salesperson must be *persuasive*, i.e. the ability to convince users to buy or try the recommended items. As persuasiveness directly associates with conversion rates, it should be the primary goal for an explainable recommendation system. To maximize the persuasiveness, it is desirable to compose for each item a catchy and appealing sentence that is enjoyable to read.

A good salesperson must also be *knowledgable*, i.e. comprehend consumer needs and infuse with deep knowledge of products. Knowledge awareness facilitates consumer decision process, leading to increased sales, lower product return rates and reduced enterprise costs. An ideal *knowledge-aware* system highlights each product's strongest features according to different user needs.

In this paper we propose a persuasive and knowledge-aware explainable recommendation system. Given a user's historical behavior and an item's profile, we provide the recommendation reason in natural language. The recommendation reason is persuasive in text style. Knowledge is incorporated through the ceoncept of consumption scenes. A consumption scene is a category of user needs which may take place at a certain place and a certain time. Our explanations describe several key features extracted from the item's profile and elaborate that these features are charming under the user's intended scene in a persuasive manner. Table **??** presents a real output from our system. We can see that for the same wooden bookcase, the generated explanation differ for two distinguished user needs. For example, when the user is looking for modern style house decor, our system decides that the user emphasizes more on functional requirements (printed in blue), such as the storage needs, making the house neat, and so on. Hence, the generated explanation selects features (printed in red) that are highly relevant to such needs, e.g. versatile. On the contrary, when the user is looking for luxury house decor, our system uses key phrases such as extraordinary taste to match specific user needs. The generated explanation selects relevant features such as natural logs and high quality and durable.

It is worthy to clarify that the explanation is post-hoc, as it is not generated from the recommendation model. We allow the flexibility of adopting any recommendation model that runs on a knowledge base of consumption scenes. Construction of the knowledge base is not discussed throughout the paper. The recommendation model delievers recommendations by mapping user intent to a consumption scene.

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

Table 1: An illustrative example

Input: item profile wooden bookcase,scene modern house decor	input: item profile wooden bookcase,scene luxury house decor
一款现代简约木质书架，采用经典的原木设计，由原生态木材拼接而成，成就了天然原木触感，绿色环保，功能多样，满足您不同的收纳需求，让您的家整洁漂亮。 A modern and simple wooden bookcase with a classic log design, made of original wood spliced, which makes the natural wood touch, environmental friendly, versatile to meet your different storage needs, so that your home is neat and beautiful.	这款书架采用好实木打造，选用天然原木，品质优良，坚固耐用。没有过多修饰的装饰，奢华而不张扬，更彰显了不凡的品味。 This bookcase is made of solid wood and is made ofnatural logs. It is of high quality and durable. There is no too much decoration, luxury without being ostentatious, shows the extraordinary taste.

Our work distinguishes from existing explainable recommendation systems. The majority of previous work aims to improve system transparency by devising interpretable models (i.e. factor models [? ? ?]) which lead to feature level explanation. Explanations have been offered as charts [?], word clouds [?], merged phrases in a textual template [? ?], or selected highly useful reviews [?]. It is widely accepted that computers which convey personality are more persuasive [?]. The lack of free text makes an explainable recommendation system less humanly, thus weakens its power to be persuasive.

We notice that there is an emerging trend in studying persuasive systems, including a recent work on transforming given specification of an item to persuasive descriptions [?]. Although they have been been useful in E-commerce sites to attract consumers, they do not satisfy the personalized demands in recommendation systems. They are not able to tailor both persuasive and informative explanations that are relevant to specific user needs, which is precisely the goal of our work.

To develop a persuasive and knowledge-aware system, we face two challenges. The first challenge arises from the insufficiency of training data. Persuasive sentence generation is a highly creative and cognitive process. Deep models have shown promising results in generating readable texts in natural language. However, such success has been fueled by large-scale supervision. It is prohibitively expensive to label persuasive sentences due to the extremely subjective and diverse nature of persuasiveness. Persuasiveness is subjective, because different human judges have varying standards. It is costly to derive a gold standard that is approved by domain experts. Persuasiveness is diverse, because a persuasive sentence can be written in different styles. It is impossible to list all variants of persuasive sentences.

To address the first challenge, we turn to a framework with weak supervision. We program a set of rules with high coverage and low accuracy to generate training sets with noisy labels from external data source that are accessible on the Internet. To resolve the potentially conflicts among rules we adopt a gnerative model with faroring weights that corresponding to imbalanced scene distribution. We experimentally show that weak supervision from externeal blogs outperforms unsupervised methods [?].

The second challenge arises from the structure of dependency between item features and scenes. Some features are universally perceived to be persuasive while some features are attractive only under specific scenes. For example, given a swimsuit, being "high quality and cheap" is good despite of any scene, , having. "a low neckline" under scene "summer beach" is good, under scene "swim competition" is bad. A naive solution is to construct training set for each scene. However, the distribution of scenes is highly skew in user needs. For example, in a total of 30000 scenes built in our in-house scene taxonomy, more than 50% scenes keywords have never appearred in the blogs.

To address the second challenge, we design a deep network following the encoder decoder framework. In the encoder layer, we utilize two modules: a global module to encode universal persuasive patterns on item features, a local module to encode scene-specific relation between persuasiveness and item features. In the decoder layer, we combine the global module and the local module through a copying mechanism. We experimentally validate that the proposed network outperforms state-of-the-art DNs in terms of explanation readibility, persuasiveness and relevance to the intended scene.

Our contributions are three folds.

- In the application level, we present a novel explainable recommendation system. To the best of our knowledge, we are the first to target directly on persuasive and knowledge-aware explanations in natural language by explaining how well the item suits user preferences under different scenes.
- In the model level, we propose a deep network model with a global module to learn universal persuasive patterns and a local module to learn scene-specific persuasive patterns.
- In the evaluation level, we design the experimental scheme as well as some easy-to-implement merits to evaluate the persuasiveness and effectiveness of our explanations.

This paper is organized as follows. We introduce 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

We briefly survey two lines of related work: explainable recommendation systems and creative text generation systems.

2.1 Explainable Recommendation

The importance and urgence of explainable recommendation is assured for the foreseeable future [?]. To date, the majority or explainable recommendation systems resort to explainable models that generate recommendations in interpretable schemes. These fall

broadly into explainable shallow models [? ?], explainable latent factor models [? ? ?], explainable neural models [?], explainable sequential models [?]. The explanations provided are byproducts of the models and hence are conviently expressed as phrases (i.e. the factor-level sentiments or review fragments [? ? ?]), charts or structured information (e.g. attention weights on each item in the same session [?]) and rules (i.e. item or user associations [? ?]).

A strategy that departs from the explainable models is to provide post-hoc explanations, which are not generated by the recommendation model itself. Hence, the predictive accuracy of the black-box recommendation model is maintained, whilst the user experience is optimized. The explanations often utilize intermediate results of the recommendation model, for example, explanations in [?] are extracted by training association rules on the outputs of a latent factor model.

Explanations can benefit recommender systems on a number of aspects. As defined in [?], there are seven categories of recommendation explanations, working respectively to improve the degree of system transparency, scrutability, trust, effectiveness, persuasiveness, efficiency and satisfaction. From the viewpoint of E-commerce sites, *persuasiveness* and *effectiveness* are the most desirable aspects. The majority of existing literature (i.e. explainable models) improves transparency by exposing the way the recommendation engine works. Post-hoc explanations are not necessarily transparent. It is a relatively unexplored area as how to generate explanations that persuade consumers to try a recommended item (persuasiveness) and explain why they may or may not want to have a try (effectiveness). Our work concentrates directly on effectiveness and persuasiveness.

Though existing methods present various types of explaination, most of them are based on predefined forms, which are considered secondary to natural language explanations. Generating explanations in natural language is still in its initial phase. The recent work [?] generates text reviews that combine user opinions on different factors of an item. Our work is more personalized as our description targets to the user's intended consumption scene.

2.2 Creative Text Generation

Neural Network Language Model (NNLM) [?] marks the start of the modern era in text generation using neural networks. To capture long-term dependencies that are not modeled in NNLM, recurrent neural network language model (RNNLM) [] is developed. Later, variants of RNN have been extensively studied, including long short-term memory (LSTM) [] and gated recurrent unit (GRU) []. Nowadays, deep neural networks have achieved remarkable success in machine translation [? ? ?]. To dispense sequential computation in recurrence and convolutions, Transformer [?] - a new simple network arthictecture solely based on attention mechanisms has been proposed recently.

In creative text generation, deep networks have also shownn promising performances, including poetry creation [? ? ? ? ? ?]. In the E-commerce domain, creative text generation also attracts attentions. However, the goal of existing research [? ?] and business applications [?] is to generate item descriptions and slogans. The lack of personalization and relevance to user needs makes it infeasible to apply them to explainable recommendation systems.

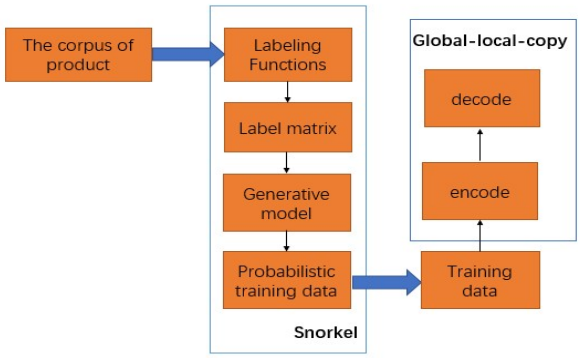

Figure 1: System Architecture

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. 2. 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. 3. **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 fraction candidates that the labeling function emits a non-zero label

[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 2: 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 3: 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

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.

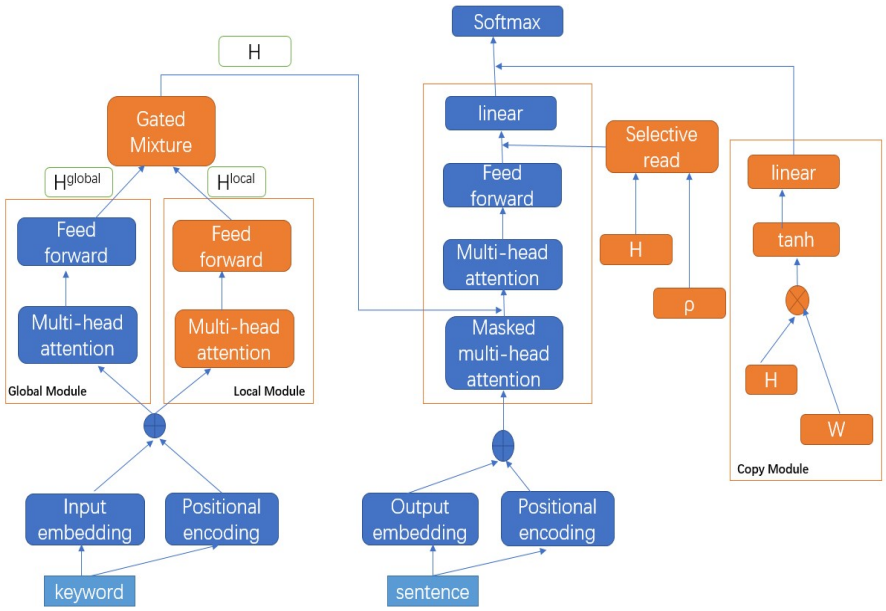

Figure 2: Global-Local-Copy Model

Table 4: Training data format

Input	Output
创意，纸巾盒，欧式	一款欧式风范桦木纸巾盒，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 5: 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. 4.

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. 5. The score of each manual evaluation metrics ranges from 0 to 5 with the higher score the better, see Tab. 6 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 6: 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

Table 7: 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 8: 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

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. 7. 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 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. 8. 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. 9. 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 9: 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。