Towards Silver-tongued Selling: Product Snippet Generation for Consumption Context Queries

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 Selling: Product Snippet Generation for Consumption Context Queries. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

No behavior on earth is more human than selling. Ever since the dawn of economy, salespeople has played a pivotal role in matching customer needs with appropriate products. We have seen, in history, many vivid examples of top salesperson who single-handedly builds a business empire. Having a strong sales team is crucial to the success of a company. This brings an interesting yet challenging question: *Can a machine function like a skilled salesperson?*

The first step to replace the sales department in E-commerce platforms is to mimic the sales presentation that connects products and consumers. This means a tailored message for each product to the consumers from their specific context. Here the context can be interpreted as the consumption context under which the products will be utilized. We found many consumers like to search for products using the consumption contexts as keywords. For example, people search "swimsuits" with consumption context keyword "summer vacations on beach". Personalization of the message based on the target context and product is important, so that it can be affiliated to a retrieval or a recommendation system.

Therefore, the problem being studied in this paper is described like this. Given a consumption context,[1] a product profile (which consists of the product name and a set of its attributes), our goal is to generate a natural language sentence that is (1) **relevant** to the consumption context, (2) **informative** about the product, (3) and organized in a **persuasive** manner.

Figure. **??** gives an illustrative example of the ideal output. It is worthy to point out that the output satisfies three essential properties. It selects an arbitrary number of attributes that convey most information of the product, and transform their values in an expression that achieves maximal relevance to the consumption context. Finally, the power of persuasion is enhanced by a catchy and appealing sentence that is enjoyable to read.

The problem falls in the broad class of language generation from data. Recently, end-to-end neural frameworks [5, 7, 21, 22] have shown promising progress in this field. End-to-end frameworks have the advantage that errors do not accumulate across separate stages, i.e. in choosing the appropriate attributes and expressing the attribute values. However, neural frameworks require massive training data. Since we emphasize particularly on relevance and persuasiveness, we face a major challenge of insufficient training data. We notice that there is an emerging trend in studying persuasive systems, including a recent work on writing persuasive product descriptions in fashion domain [10]. While their work is the most similar to ours, due to the same data scarcity problem, they propose an unsupervised framework and the performance is secondary.

To address the insufficiency of training data, **at the data level**, we turn to a framework with weak supervision. Weak supervised learning implement unsupervised models or/and manually constructed heuristics on unlabelled data. When it comes to labeling persuasive sentences, there is no sophisticated unsupervised model to discriminate persuasive and non-persuasive text. In order to generate a set of rules that cover a considerably large number of possible variations, we use external data sources (i.e. blogs and advertisements) to discover syntactic sequential rules and features for persuasive texts.

Weak supervision is inherently incomplete and inaccurate. Furthermore we should solve the structural dependency between item attributes and consumption context. Some attributes are universally perceived to be persuasive while others are attractive only under specific contexts. For example, given a swimsuit, being "high quality and cheap" is good despite of any context, , having. "a low neckline" under context "summer beach" is good, under context "swim competition" is bad. Therefore, **at the model level**, we embed a

[1] We consider the context to be explicitly provided. It can also be extracted, e.g. from user profiles, which is beyond the scope of this paper.

Conference'17, July 2017, Washington, DC, USA
© 2019 Association for Computing Machinery.
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

global-local component for skew attribute-context distribution, and a copying component for out-of-vocabulary references.

Our solution **at the knowledge level** is to incorporate knowledge to understand high-level themes of a consumption context. We experiment with different knowledge representations extracted from data sources other than the training data (i.e. from search logs in our E-commerce platforms). Thus our knowledge-rich framework summarizes the connection between consumption needs and products at different granularities, which further dispose of the data scarcity problem.

Our contributions are three folds.

- We study a novel problem of generating persuasive, relevant and informative product snippet for different consumption contexts. As this process is highly creative and cognitive, it will beneficial for other AI topics. For example, snippet generation is the first step, which can be later extended to negotiation.
- We present data-level, model-level and knowledge-level solutions to handle with the problem of insufficient training data.
- We design and present a set of novel experimental scheme as well as some easy-to-implement merits to evaluate the effectiveness of our system.

This paper is organized as follows. We introduce the related work in Sec. 2. In Sec. **??**, we first introduce the architecture of our system and describe the weak supervision data set. In Sec. **??** we discribe our model with its global-local-copy component. 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 research related to our work, language generation and learning with weak supervision.

2.1 Language Generation

Natural Language Generation (NLG) task has always been one of the most widely studied problems in the area of natural language processing. in particular, we identify two types of NLG tasks: data-to-document generation and creative text generation.

Data-to-document generation is a classic NLG task. Given some structured data, such as a table, data-to-document generation task is to produce text, such as a sentence or a paragraph, that adequately and fluently describes the input data. Applications of data-to-document generation techniques include automatically generating product review [2**?**], game report [21], dialogues [22], biography [7], and captioning an image [5].

Early systems usually consist of two separate stages: a content selection stage to decide "what to say", and a surface realization stage to decide "how to say". The recent success of Deep Neural Network (DNN) models [14] has galvanized research on end to end systems that blur the distinction between the two stages. Most of the DNN systems employ an encoder-decoder framework. Frequently adopted encoders include Multi-Level Perception (MLP) [1, 21**?**] or a hierarchical form of LSTM [22]. Copying is considered to be effective to manage out-of-vocabulary table references. For example, several copying mechanism are compared in [1, 21]. Beyond simple copying, refer to a referent using varied mention forms is additionally studied in [22]. After a template text with data slots is generated, a delayed copy mechanism is proposed in [8] to fill in the slots with proper data records. In the decoder layers, RNN [21] and LSTM [22] are common choices. Other network structure may replace the encoder-decoder framework as the method of choice. For example, a feed-forward neural language model with attention based word copying is presented in [7] which takes aggregated word and table embeddings as input.

From a commercial point of view, **creative text generation** tasks have received considerably more attention. In these tasks, the generated text must reveal more human characteristics. Applications of creative text generation include computational humor (i.e. joke recognition [6] and pun generation [18]), poetry generation (i.e. Chines poetry generation [20, 23] and English poetry generation [4]) and slogan generation [17], and so on.

Again, DNN method is appealing when supervision is accessible. For example, most state-of-the-art research on poetry generation is based on encoder-decoder framework [4, 20, 23]. However, training collections is difficult to obtain for other types of creative text, due to the inherent complexity of cognitive process. In this case, unsupervised method has become the main stream method. Most of them are heavily dependent on syntactic templates, e.g. word substitution [11, 15, 18] and can only generate short headline style sentences or slogans. A recent work [10] explores the possibility of generating a complete persuasive sentence by an unsupervised approach.

To summarize, there are fruitful results yielded by NLG results, however, our work is different from existing work on the following two aspects. (1) While most NLG tasks focus on the output's fluency and fidelity to references, we emphasize on the persuasiveness and relevance of the output. This is because our system aims to maximize user experience in the E-commerce ecosystem. In a sense, our system resembles post-hoc explainable recommendation systems, i.e. systems that provide explanations which are not generated by the recommendation model itself. This type of explanations can greatly improve the degree of system effectiveness, persuasiveness, efficiency and satisfaction [16]. Generating explanations in natural language is still in its initial phase. (2) End-to-end DNN models require large amounts of training data to obtain promising results. When it is impossible to generate labeled corpus, creative text generation systems often resort to unsupervised approaches. Our work attempts to exploit the superior learning power of DNNs by utilizing weak supervisions.

2.2 Weak Supervision

Supervised learning is firmly established as a standard technology across all machine learning related tasks when labelled ground truth is available. However, labelling is a labor costly procedure. To address the data scarcity issue, Recently, a significant amount of research has been proposed that aims to utilize weak supervision, which is the opposite of strong supervision [24]. The collection of weak supervision can be obtained by either an unsupervised (and possibly worse performing) model, or a set of manually constructed heuristics. Examples of the former category includes training a fully feedforward NN model for information retrieval on documents

Figure 1: Example persuasive snippets for a bookcase under different consumption contexts.

Input	Product Profile	书架，木质，自然，高质量…… bookcase, wooden, natural, high quality,...	
	Consumption Context	现代家居 modern house decor	奢华家居 luxury house decor
Output	Persuasive Relevant Informative Explanation	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.

labeled by a simple ranking model such as BM25 [3]. Examples of the latter category includes generating training sets by aggregating expert-derived rules [13].

As weak supervisions are often incomplete (i.e. only a small fraction of training set is labelled), inexact (i.e. only coarse-grained labels are given), and/or inaccurate (i.e. given labels are not always correct), model adaptions are necessary to optimize performance. However, this is not fully explored in the DNN literature, especially for natural language generation. Most previous works simply treat weak supervision signals as labels.

3 WEAK SUPERVISION

In this section, we describe in detail the method to generate a training corpus where an input of consumption context and an item profile is associated with an output of relevant, informative and persuasive sentence. As our system is built on an E-commerce platform, we have the fortune to access a large scale product description datasets. That is, a paragraph is associated with each product (name and attributes). However, this data set does not guarantee that the product description is relative to any consumption context, or informative about the product, or written in a persuasive manner.

We take a boolean retrieval model as our labeler to filter relevant and informative descriptions. First, we build a lexicon consisting of consumption context terms. If the output product description does not contain at least one context term or it does not contain at least one product attribute in the input, the pair is deleted from our corpus.

The identification of persuasive descriptions is more challenging. On one hand, annotating such a massive amount of instances is costly. On the other hand, recognition of persuasive sentences is highly subjective. Therefore, we adopt “pseudo-labeling” by fitting these descriptions to a set of patterns derived from an external data set. The data set is crawled from blogs that we beleve to be advertisements? We use a combination of a set of manually constructed heuristics and a set of automatically extracted rules.

3.1 Manually Constructed Heuristics

3.2 Automatically Extracted Rules

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.

Manual labeling is very time consuming, so we use the Snorkel [13] 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 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 [19] 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.

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

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.

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

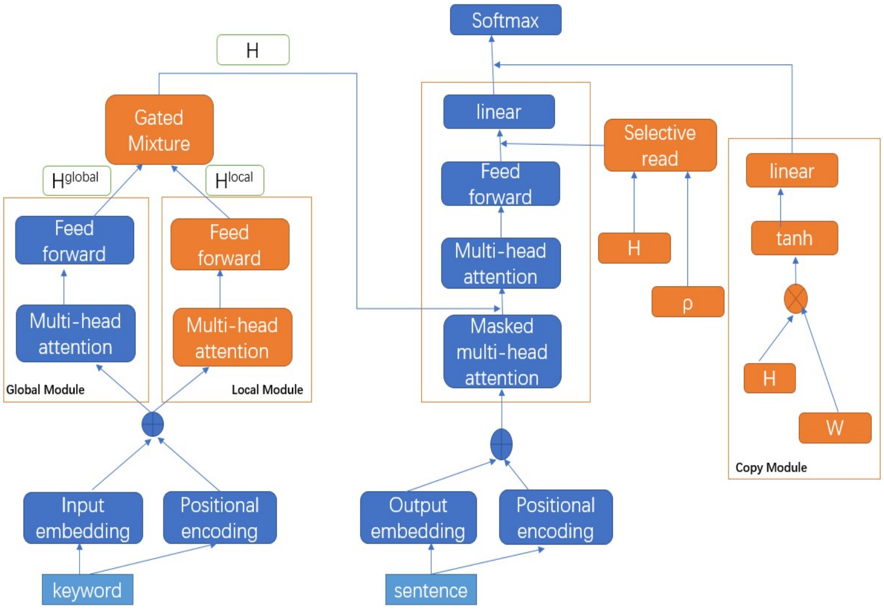

Figure 2: Global-Local-Copy Model

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。

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 [9] and BLEU [12] 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 4: Manual Evaluation

Evaluation Metric	Description
Fluency [20]	Does the sentence read smoothly and fluently?
Catchyness [10]	Is the description attractive,catchy?
Relatedness [10]	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?

4.5 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

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

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

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. 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.

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

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

REFERENCES

[1] J. Bao, D. Tang, N. Duan, Z. Yan, M. Zhou, and T. Zhao. 2019. Text Generation From Tables. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 27, 2 (Feb 2019), 311–320.

[2] Felipe Costa, Sixun Ouyang, Peter Dolog, and Aonghus Lawlor. 2018. Automatic Generation of Natural Language Explanations. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion*. ACM, New York, NY, USA, 57:1–57:2.

[3] Mostafa Dehghani, Hamed Zamani, Aliaksei Severyn, Jaap Kamps, and W. Bruce Croft. 2017. Neural Ranking Models with Weak Supervision. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17)*. ACM, New York, NY, USA, 65–74.

[4] Marjan Ghazvininejad, Xing Shi, Yejin Choi, and Kevin Knight. 2016. Generating Topical Poetry. In *Conference on Empirical Methods in Natural Language Processing*. 1183–1191.

[5] Andrej Karpathy and Li Fei-Fei. 2015. Deep visual-semantic alignments for generating image descriptions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3128–3137.

[6] Chloe Kiddon and Yuriy Brun. 2011. That's what she said: double entendre identification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*. Association for Computational Linguistics, 89–94.

[7] Remi Lebret, David Grangier, and Michael Auli. 2016. Neural Text Generation from Structured Data with Application to the Biography Domain. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. 1203–1213.

[8] Liunian Li and Xiaojun Wan. 2018. Point Precisely: Towards Ensuring the Precision of Data in Generated Texts Using Delayed Copy Mechanism. In *Proceedings of the 27th International Conference on Computational Linguistics*. 1044–1055.

[9] Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. *Text Summarization Branches Out* (2004).

[10] Vitobha Munigala, Abhijit Mishra, Srikanth G Tamilselvam, Shreya Khare, Riddhiman Dasgupta, and Anush Sankaran. 2018. PersuAIDE! An Adaptive Persuasive

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。

Text Generation System for Fashion Domain. In *Companion of the The Web Conference 2018 on The Web Conference 2018*. International World Wide Web Conferences Steering Committee, 335–342.

[11] Gozde Ozbal, Daniele Pighin, and Carlo Strapparava. 2013. BRAINSUP: Brainstorming support for creative sentence generation. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Vol. 1. 1446–1455.

[12] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics*. Association for Computational Linguistics, 311–318.

[13] Alexander Ratner, Stephen H Bach, Henry Ehrenberg, Jason Fries, Sen Wu, and Christopher Ré. 2017. Snorkel: Rapid training data creation with weak supervision. *Proceedings of the VLDB Endowment* 11, 3 (2017), 269–282.

[14] Ilya Sutskever, James Martens, and Geoffrey E Hinton. 2011. Generating text with recurrent neural networks. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*. 1017–1024.

[15] Stamatina Thomaidou, Ismini Lourentzou, Panagiotis Katsivelis-Perakis, and Michalis Vazirgiannis. 2013. Automated snippet generation for online advertising. In *Proceedings of the 22nd ACM international conference on Conference on information &#38; knowledge management (CIKM ’13)*. ACM, New York, NY, USA, 1841–1844. https://doi.org/10.1145/2505515.2507876

[16] Nava Tintarev and Judith Masthoff. 2011. *Designing and Evaluating Explanations for Recommender Systems*. Springer US, Boston, MA, 479–510. https://doi.org/10.1007/978-0-387-85820-3_15

[17] Polona Tomavsic, M Znidarvsic, and Gregor Papa. 2014. Implementation of a slogan generator. In *Proceedings of 5th International Conference on Computational Creativity, Ljubljana, Slovenia*, Vol. 301. 340–343.

[18] Alessandro Valitutti, Hannu Toivonen, Antoine Doucet, and Jukka M Toivanen. 2013. "Let Everything Turn Well in Your Wife": Generation of Adult Humor Using Lexical Constraints. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Vol. 2. 243–248.

[19] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*. 5998–6008.

[20] Zhe Wang, Wei He, Hua Wu, Haiyang Wu, Wei Li, Haifeng Wang, and Enhong Chen. 2016. Chinese Poetry Generation with Planning based Neural Network. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. 1051–1060.

[21] Sam Wiseman, Stuart Shieber, and Alexander Rush. 2017. Challenges in Data-to-Document Generation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. 2253–2263.

[22] Zichao Yang, Phil Blunsom, Chris Dyer, and Wang Ling. 2017. Reference-Aware Language Models. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. 1850–1859.

[23] Xingxing Zhang and Mirella Lapata. 2014. Chinese Poetry Generation with Recurrent Neural Networks. In *Conference on Empirical Methods in Natural Language Processing*. 670–680.

[24] Zhi-Hua Zhou. 2018. A brief introduction to weakly supervised learning. *National Science Review* 5, 1 (2018), 44–53.