Quality-Sensitive Training! Social Advertisement Generation by Leveraging User Click Behavior

Yongzhen Wang
Indiana University Bloomington
Bloomington, Indiana, USA
kuadmu@163.com

Heng Huang
Alibaba Group
Hangzhou, Zhejiang, China
gongchong.hh@taobao.com

Yuliang Yan
Alibaba Group
Hangzhou, Zhejiang, China
yuliang.yyl@alibaba-inc.com

Xiaozhong Liu*
Indiana University Bloomington
Bloomington, Indiana, USA
liu237@indiana.edu

ABSTRACT

Social advertisement has emerged as a viable means to improve purchase sharing in the context of e-commerce. However, humanly generating lots of advertising scripts can be prohibitive to both e-platforms and online sellers, and moreover, developing the desired auto-generator will need substantial gold-standard training samples. In this paper, we put forward a novel seq2seq model to generate social advertisements automatically, in which a quality-sensitive loss function is proposed based on user click behavior to differentiate training samples of varied qualities. Our motivation is to leverage the clickthrough data as a kind of quality indicator to measure the textual fitness of each training sample quantitatively, and only those ground truths that satisfy social media users will be considered the eligible and able to optimize the social advertisement generation. Specifically, under the qualified case, the ground truth should be utilized to supervise the whole training phase as much as possible, whereas in the opposite situation, the generated result ought to preserve the semantics of original input to the greatest extent. Simulation experiments on a large-scale dataset demonstrate that our approach achieves a significant superiority over two existing methods of distant supervision and three state-of-the-art NLG solutions.

CCS CONCEPTS

• **Computing methodologies → Artificial intelligence**; **Natural language processing**; **Natural language generation**; • **Applied computing → Electronic commerce**.

KEYWORDS

Social Advertising; Natural Language Generation; Clickthrough Data; Neural Networks; Sequence-to-Sequence

ACM Reference Format:
Yongzhen Wang, Heng Huang, Yuliang Yan, and Xiaozhong Liu. 2019. Quality-Sensitive Training! Social Advertisement Generation by Leveraging User Click Behavior. In *Proceedings of the 2019 World Wide Web Conference (WWW'19), May 13–17, 2019, San Francisco, CA, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3308558.3313536

1 INTRODUCTION

In the era of e-commerce, the purchase sharing via social networks is an organic form of sale promotion generated by the public[1], when a growing number of people make their buying decisions from the shared product links[2]. For instance, Alibaba users share product information over 150M times per day through social media sites like Weibo and WeChat, on which those users may one-click-share what they bought to relatives or friends conveniently. While these deliveries can be just plain URLs, online stores ideally want to present attractive advertising scripts to not only explain product features adequately, but also convert audiences into potential customers, as shown in Fig. 1. However, writing such eye-catching narratives does require substantial marketing skills. Even though content experts could be employed to create the social advertisements, merchants have to struggle with the cost incurred by an army of commodities.

As a solution to save the workforce, traditional software heavily relies on settled examples to auto-generate product descriptions, which in turn challenges both personalization and copyright. More recently, neural networks enable a data-driven architecture *sequence-to-sequence* (seq2seq) for the *natural language generation* (NLG) [1, 2], in which an encoder reads a sequence of tokens into a context vector, and then from that a decoder yields a sequence of specific outcomes. But unlike the machine translation with an end-to-end nature, a majority of e-commerce sales lack a pair of concise yet appropriate texts to construct the gold-standard training samples with respect to the seq2seq paradigm. Especially in the context of mobile social apps, online sellers tend to serve shoppers with a brief title as well as many detail images to post product information for the sake of *search engine optimization* (SEO), when too lengthy presentations are not suitable for both display and edit on the limited screens of cellphones. Under this circumstance, the

*Corresponding Author

WWW '19, May 13–17, 2019, San Francisco, CA, USA

ACM ISBN 978-1-4503-6674-8/19/05.
https://doi.org/10.1145/3308558.3313536

[1] https://www.bigcommerce.com/ecommerce-answers/what-are-social-share-buttons-and-how-do-they-impact-conversions

[2] https://sproutsocial.com/insights/social-media-ecommerce

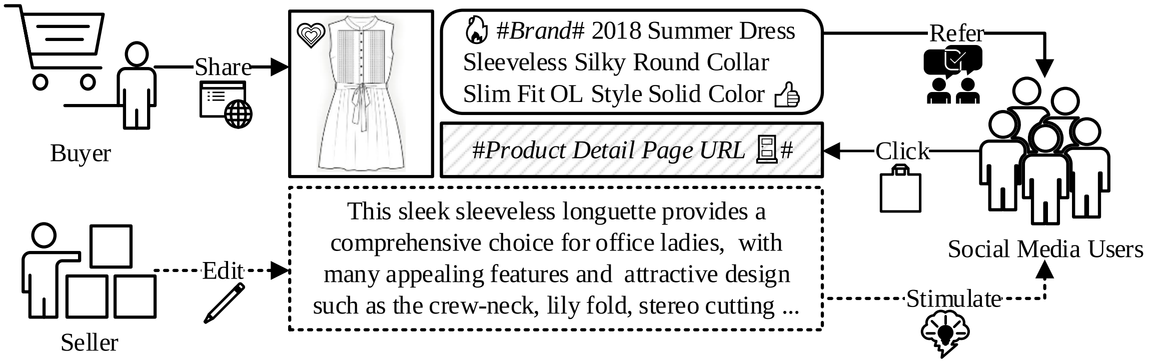
Buyer
Share
#Brand# 2018 Summer Dress Sleeveless Silky Round Collar Slim Fit OL Style Solid Color
Refer
#Product Detail Page URL #
Click
Social Media Users
Seller
Edit
This sleek sleeveless longuette provides a comprehensive choice for office ladies, with many appealing features and attractive design such as the crew-neck, lily fold, stereo cutting ...
Stimulate

Figure 1: The illustration of the purchase sharing, in which a social advertisement (rounded textbox) is prepared for each product title (dashed textbox) accompanied by a raw URL (shaded textbox).

title often merely covers a few key attributes only, and moreover, drawing useful scripts from those photos is far more difficult.

Concerning the seq2seq NLG, a qualified training sample always requires some textual fitness to hold between its input and output, such as the semantic equivalence for machine translation, and the language logic for question answering. To address the qualities of training samples, latest studies try a widely used toolkit ROUGE [21] to calculate word co-occurrences from inputs to outputs exactly [6, 7, 24, 34]. Meanwhile, another strand of works attempt to retrieve approximate outputs as soft templates to offer additional guidelines [5]. Nonetheless, these methods of distant supervision [23] can hardly apply to the *social advertisement generation* (SAG) for the following reasons. First, due to vocabulary variation, counting the lexical overlaps could barely present the correct relationships between product titles (most readily available) and social advertisements (composed by heuristics). Second, the rough advertising templates would inevitably bring about irrelevant product details, which might confuse social media users and thus cause a decrease in the related purchase conversion rate. Last but most important, the performance of a social advertisement is highly dependent on how well it attracts audiences to click-return, and this type of quality assessment should be intrinsically dynamic because of the evolving interest of social media users outside. These reasons motivate us to re-investigate the measurement for textual fitness, and take the user-centric qualities of training samples into consideration.

In this study, we aim at the SAG to envision an innovative NLG application, in which user click behavior is leveraged to estimate the actual textual fitness of a candidate social advertisement to a particular product title. The basic idea lies in that the human-friendly advertising has a greater potential to improve the *clickthrough rate* (CTR) of shared product links for social media sites. In other words, the varying CTR can be utilized as a kind of quality indicator to characterize the matching degrees of corresponding paired texts, for which the desired evaluation involves relevance, diversity, and interestingness three criteria that are not easy to judge by computers. Note that the former two indicate the properties of generated results themselves, whereas the third one reveals the up-to-date preference of social media users accordingly. Specifically, we integrate the source paired texts with their CTR logs to establish a latent semantic space, where any pair of the product title and social advertisement could be mapped into two embeddings respectively so that the textual fitness may be measured therefrom. More formally, this measurement can be interpreted as a certain genre of supervision strength under the view of Internet crowd. With such a distant supervision, we are able to assess the qualities of training samples quantitatively, and the better the textual fitness grows, the higher the quality becomes.

The contribution of this paper is threefold. First, we explore the textual fitness measurement via a latent semantic space that features user click behavior, which enables the user-centric qualities of training samples for the seq2seq NLG. Second, we develop a novel seq2seq model to generate social advertisements automatically, where a quality-sensitive loss function is proposed to differentiate the training samples of varied qualities. Third, we conduct extensive experiments on 1.18M samples with their ground truths extracted by heuristics, and simulation results demonstrate that our approach outperforms two existing methods of distant supervision and three state-of-the-art NLG solutions.

2 RELATED WORK

This study touches on several strands of research within e-commerce NLG, click logs mining, and distant supervision as follows.

E-commerce NLG It is well acknowledged that the NLG offers dynamic improvements for e-commerce by turning product data into reader- and SEO-friendly narratives[3]. As one of the typical applications, Wang et al. [32] present a statistical framework that generates product descriptions from related attributes, in which some selected templates are combined with writing knowledge to decide what and how to say. More recently, Wang et al. [33] propose the first data-driven solution to compress product titles for the sake of SEO, which makes use of a pointer network-based model under the settings of multi-tasks. Subsequently, Gong et al. [15] and Sun et al. [28] further put forward two methods to summarize product titles, where the former utilizes three different categories of features in parallel, and the latter introduces a knowledge encoder as the auxiliary tool. Not long ago, Mathur et al. [22] employ several seq2seq models on a multi-lingual dataset, and thereby create a joint

[3] https://multichannelmerchant.com/blog/natural-language-generation-coming-ecommerce

model to produce product titles in various languages. Nonetheless, these works all conform to one formulation that the information of outputs is entirely contained in that of inputs, which differs from the SAG that parts of the social advertisements are free from or beyond their associated product titles.

Click Logs Mining In the field of information retrieval, click logs mining has proved to be a powerful means toward optimizing search engines [18, 20]. The most well-known model on query-document matching is the latent semantic analysis [9], by which a document or query can be mapped into a low-dimensional concept by the single value decomposition of corresponding document-term or query-term matrix. On this basis, Gao et al. [12, 13] develop a document ranking model trained on clicked query-document pairs, which ranks the documents by the likelihood of given query being a semantics-based translation of each document, with the assumption that the two sources of texts share the same distribution over semantic topics. As a further research, Huang et al. [17] propose a *deep structured semantic model* (DSSM) that projects both queries and documents into a common low-dimensional space respectively. This model is discriminatively trained by maximizing the conditional likelihoods of clicked documents regarding a particular query, and is considered the state-of-the-art in comparison to most previous works. However, despite these positive examples of exploiting click logs, few studies have applied the user click behavior to the NLG domain for either locating or sampling important contents.

Distant Supervision Over the past years, distant supervision has made a great success in a series of classification tasks featuring a weakly labeled training dataset [10, 23]. Inspired by this observation, Cao et al. [6] develop a seq2seq model for extractive summarization, in which each candidate sentence with comparatively high ROUGE-2 [21] score is annotated as a positive training sample. Similarly, Cheng and Lapata [7], Nallapati et al. [24], and Wang et al. [34] all adopt this labeling strategy and further raise three different seq2seq summarizers separately, where the sentence labels serve as ground truths to offer the essential guidelines for training samples. Most recently, Cao et al. [5] utilize *Lucene*[4] to retrieve approximate summaries as pseudo annotations for abstractive summarization, which enhances the textual fitness between the input and output inside each training sample. Unfortunately, although these studies improve the qualities of training samples in diverse ways, they lack an effective mechanism to integrate such kind of supervision information into the training process of the seq2seq paradigm.

In this study, our proposed seq2seq model derives from the established pointer-generator network [25], in which we put forward a quality-sensitive loss function to differentiate the training samples of varied qualities. In particular, the involved quality assessment builds upon the concept of the DSSM [17], which integrates the source paired texts with their CTR logs to develop a latent semantic space for measuring the textual fitness of a candidate social advertisement to a particular product title. To the best of our knowledge, we make the first attempt to address the problem of generating social advertisements automatically, and the practice of applying CTR data into the NLG domain is also less explored. Besides, this work is launched on a simulation environment, where a total of 180K experimental samples are constructed and hence make our results more convincing.

3 METHODOLOGY

This section formulates the SAG into a seq2seq NLG problem, as shown in Fig. 2, where the input and output are respectively a product title (green mark) and its paired social advertisement (red mark). To be detailed, Sect. 3.1 presents the encoder integrating with contextual features, Sect. 3.2 introduces the decoder making use of a pointer-generator mechanism, Sect. 3.3 elaborates how to set up the latent semantic space, and Sect. 3.4 describes the implementations for quality-sensitive training.

3.1 Contextual Encoder with NER

The goal of encoder is to transform an input sequence into a dense representation. Consider a product title of n words $T = \{t_{1:n}\}$, and each token $t_i \in \mathbb{V}$ (vocabulary list) can be represented by a word embedding $\boldsymbol{e}(t_i) \in \mathbb{R}^{d_\mathrm{e}}$. In the NLG domain, previous studies have illustrated the superiority of *bidirectional long-short-term memory* (Bi-LSTM) in presenting sequential data, because of its capability to handle the long-range dependency and avoid the gradient vanishing [3, 26]. The Bi-LSTM encoder consists of both forward and backward two *recurrent neural networks* (RNNs), in which the hidden state at time step i can be denoted as follows:

$$\boldsymbol{h}_i = \left[\overrightarrow{\boldsymbol{h}}_i; \overleftarrow{\boldsymbol{h}}_i\right] \tag{1}$$

where each hidden state $\boldsymbol{h}_i \in \mathbb{R}^{2d_\mathrm{h}}$ can be regarded as a local representation with focusing on the current and surrounding (before $\overrightarrow{\boldsymbol{h}}_{i-1}$ and after $\overleftarrow{\boldsymbol{h}}_{i+1}$, as Eq. 2-3) inputs simultaneously.

$$\overrightarrow{\boldsymbol{h}}_i = \mathrm{LSTM}\Big(\boldsymbol{e}(t_i), \overrightarrow{\boldsymbol{h}}_{i-1}\Big) \tag{2}$$

$$\overleftarrow{\boldsymbol{h}}_i = \mathrm{LSTM}\Big(\boldsymbol{e}(t_i), \overleftarrow{\boldsymbol{h}}_{i+1}\Big) \tag{3}$$

Despite the robustness of Bi-LSTM, as Fig. 1 shows that even though a product title conforms to a temporal sequence in a certain sense, a host of the phrases therein could swap their positions without any divergence, especially for those adjectives such as "*Sleeveless*" and "*Silky*". To better adapt the seq2seq paradigm, we draw out the fine-grained information from product titles via *named entity recognition* (NER) so as to enhance the sequential dependency of encoding stage. This empirical practice is derived from the contextual LSTM of [14], in which the extracted named entities are used as contextual features to capture the segment structures within text sequences. For instance, the *brand-name* is frequently followed by the *season-to-market*, whereas those adjacent modifiers seldom to restrict their appearance orders. More formally, an original token t_i is combined with its named entity $r_i \in \mathbb{V}$ as a composite input $\Big[\boldsymbol{e}(t_i); \boldsymbol{e}(r_i)\Big] \in \mathbb{R}^{2d_e}$ to feed the Bi-LSTM encoder, by which no additional parameters are introduced herefrom.

3.2 Attentional Decoder with Pointer

The decoder is used to generate an output sequence in terms of settled encoded representations. Given the hidden states of encoder

[4] https://lucene.apache.org

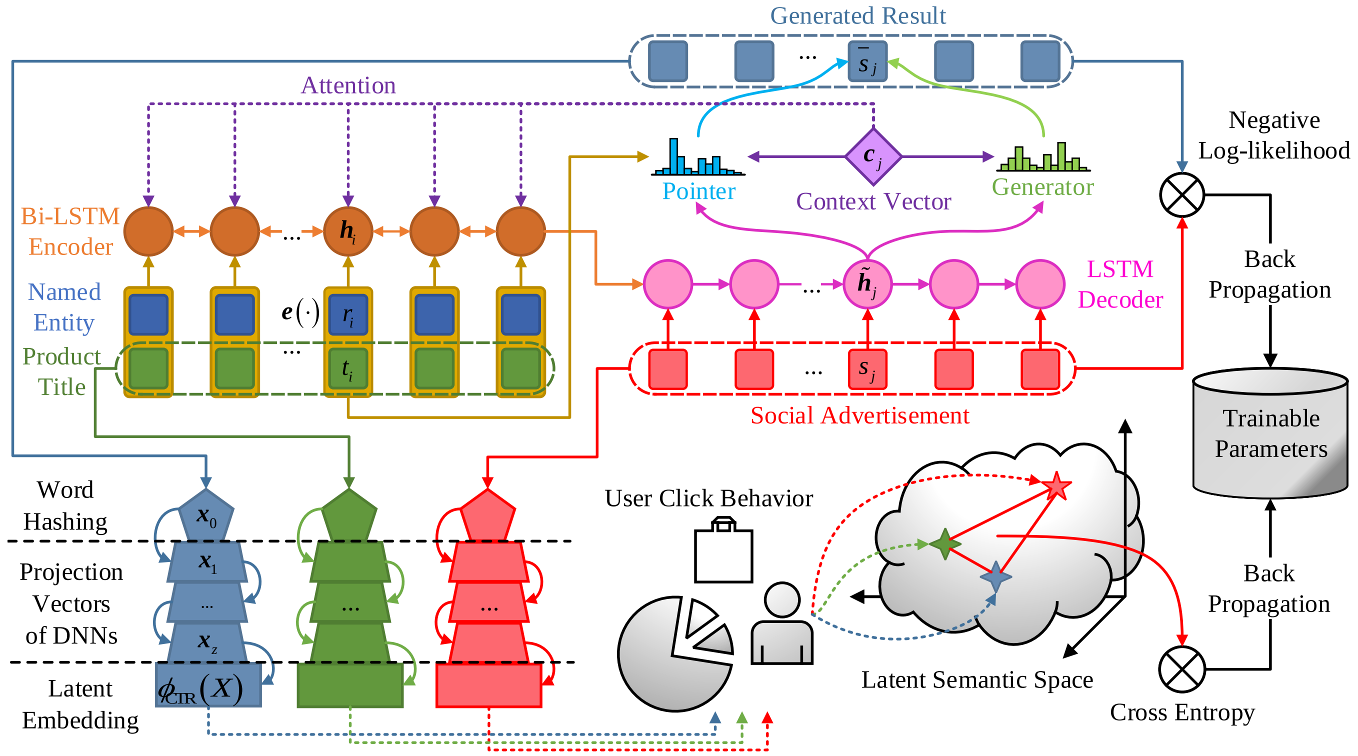

Figure 2: The seq2seq NLG framework formulated by the SAG.

$H = \{\boldsymbol{h}_{1:n}\}$, we apply an attention-based LSTM to sequentially compose a social advertisement $S = \{s_{1:m}\}$. To be specific, each hidden state of decoder $\tilde{\boldsymbol{h}}_j \in \mathbb{R}^{d_\text{h}}$ (as Eq. 5) is made by both the previous state and the last sampled word $\bar{s}_j \in \mathbb{V}$[5]. Next, this hidden state is concatenated with a context vector $\boldsymbol{c}_j \in \mathbb{R}^{2d_\text{h}}$ to feed through a linear layer and produce the vocabulary distribution $\text{Pr}^{\text{voc}}_{j,1:|\mathbb{V}|}$ as below:

$$\text{Pr}^{\text{voc}}_{j,1:|\mathbb{V}|} = \text{Softmax}\Big(\boldsymbol{W}_\text{v}\big[\tilde{\boldsymbol{h}}_j;\boldsymbol{c}_j\big] + \boldsymbol{b}_\text{v}\Big) \tag{4}$$

where $\boldsymbol{W}_\text{v} \in \mathbb{R}^{|\mathbb{V}|\times 3d_\text{h}}$ indicates a learnable matrix, and $\boldsymbol{b}_\text{v} \in \mathbb{R}^{|\mathbb{V}|}$ denotes a bias term. In particular, $\boldsymbol{c}_j$ (as Eq. 6) is the weighted sum of encoder hidden states, and the attention weight $\alpha^i_j \in [0,1]$ (as Eq. 7) demonstrates how much the title token t_i contributes to sample the advertisement word s_j. This mechanism actually acts as an intermediate phase to determine which parts of the encoder to highlight for current samplings [1], in which a bi-linear learner $\boldsymbol{M} \in \mathbb{R}^{d_\text{h}\times 2d_\text{h}}$ is employed to measure such kind of relationship.

$$\tilde{\boldsymbol{h}}_j = \text{LSTM}\Big(\boldsymbol{e}(\bar{s}_j), \tilde{\boldsymbol{h}}_{j-1}\Big) \tag{5}$$

$$\boldsymbol{c}_j = \sum_{i=1}^{n} \alpha^i_j \boldsymbol{h}_i \tag{6}$$

$$\alpha^i_j = \frac{\exp\Big(\tilde{\boldsymbol{h}}_j^\top \boldsymbol{M}\boldsymbol{h}_i\Big)}{\sum_{k=1}^{n}\exp\Big(\tilde{\boldsymbol{h}}_j^\top \boldsymbol{M}\boldsymbol{h}_k\Big)} \tag{7}$$

$$\text{Pr}^{\text{opt}}_j = \text{Sigmoid}\Big(\boldsymbol{W}_\text{o}\big[\tilde{\boldsymbol{h}}_j;\boldsymbol{c}_j;\boldsymbol{e}(\bar{s}_j)\big] + \boldsymbol{b}_\text{o}\Big) \tag{8}$$

Unlike conventional NLG problems, the neologisms including coming trends and emerging fashions bloom in e-commerce every day, which results in a lot of *out-of-vocabulary* (OOV) words necessarily. To this end, we introduce an established pointer from [25] to transport title contents into our target advertisements directly. More formally, a probability $\text{Pr}^{\text{opt}}_j \in \mathbb{R}$ (as Eq. 8, where $\boldsymbol{W}_\text{o} \in \mathbb{R}^{3d_\text{h}+d_\text{e}}$ and $\boldsymbol{b}_\text{o} \in \mathbb{R}$) is utilized to switch on the option for whether making a copy or not, and hence the final distribution $\text{Pr}^{\text{dis}}_{j,1:|\mathbb{V}|}$ is obtained by adding a pointer component $\text{Pr}^{\text{ptr}}_{j,1:|\mathbb{V}|}$ to a generator basis $\text{Pr}^{\text{voc}}_{j,1:|\mathbb{V}|}$ as follows:

$$\text{Pr}^{\text{dis}}_{j,1:|\mathbb{V}|} = \text{Pr}^{\text{opt}}_j \times \text{Pr}^{\text{voc}}_{j,1:|\mathbb{V}|} + \Big(1 - \text{Pr}^{\text{opt}}_j\Big) \times \text{Pr}^{\text{ptr}}_{j,1:|\mathbb{V}|} \tag{9}$$

$$\text{Pr}^{\text{ptr}}_{j,k} = \begin{cases} \alpha^i_j, & \text{if} \quad k = \mathbb{V}_{t_i} \\ 0, & \text{otherwise} \end{cases} \qquad \forall k \in \mathbb{N}_+ \wedge k \leq |\mathbb{V}| \tag{10}$$

where $\mathbb{V}_{t_i} \in \mathbb{N}_+$ returns the index of t_i in $\mathbb{V}$.

[5] If during the training process, $\bar{s}_j$ is the current token of ground truth, or the last emission of decoder otherwise.

3.3 Latent Semantic Space Featuring CTR

Numerous works have confirmed the CTR as one of the most promising channels to access human evaluation on the Internet [20, 31]. In this study, we aim at the SAG to verify this hypothesis, by leveraging user click behavior to measure the textual fitness between product titles and social advertisements.

Generally speaking, a well-designed product title T is pieced by the main characteristics of sale promotion due to length limits. This messy display would certainly affect audiences' returning experiences, for which an appropriate social advertisement S is required to stimulate their shopping impulses effectively. Since evaluating such textual fitness is fairly computer-hard, we take the CTR as an alternative means to estimate the matching degrees of corresponding paired texts, as shown in Eq. 11.

$$Q(T,S) = \text{Cosine}(T,S \mid \text{CTR}) = \frac{\phi_{\text{CTR}}(T)^\top \phi_{\text{CTR}}(S)}{\left\|\phi_{\text{CTR}}(T)\right\| \times \left\|\phi_{\text{CTR}}(S)\right\|} \tag{11}$$

where $\text{Cosine}(T,S \mid \text{CTR})$ represents the affinity of T to S within the latent semantic space $\phi(\cdot)_{\text{CTR}} \in \mathbb{R}^{d_\text{s}}$ featuring the CTR, which equates to promote an ocean of social media users to infer the textual fitness between T and S from semantics. In this manner, the consumer preference for social advertisements can be depicted to help assess the user-centric qualities of training samples quantitatively.

Specifically, we follow the mature practice of [17] to apply a stack of z *deep neural networks* (DNNs) for deriving the latent embedding of a text sequence X (T or S) as below:

$$\boldsymbol{x}_0 = \text{Hashing}(X) \tag{12}$$

$$\boldsymbol{x}_l = \tanh\left(\boldsymbol{W}_l \boldsymbol{x}_{l-1} + \boldsymbol{b}_l\right), \quad \forall l \in \{1, 2, \cdots, z\} \tag{13}$$

$$\phi_{\text{CTR}}(X) = \tanh\left(\boldsymbol{W}_\text{s} \boldsymbol{x}_z + \boldsymbol{b}_\text{s}\right) \tag{14}$$

where $\text{Hashing}(\cdot)$ indicates a function that converts from the one-hot representation X into the letter n-gram denotation $\boldsymbol{x}_0$ [27], $\boldsymbol{x}_{1:z} \in \mathbb{R}^{d_\text{x}}$ represents z projection vectors at different layers of the DNNs, and $\boldsymbol{W}_l \in \mathbb{R}^{d_\text{x} \times d_\text{x}}$, $\boldsymbol{b}_l \in \mathbb{R}^{d_\text{x}}$, $\boldsymbol{W}_\text{s} \in \mathbb{R}^{d_\text{s} \times d_\text{x}}$, $\boldsymbol{b}_\text{s} \in \mathbb{R}^{d_\text{s}}$ denote $2z + 2$ tensors for the linear maps within Eq. 13-14. In practice, the pairwise ranking strategy of [8] is adopted to tune all these parameters, in which a training sample with its own CTR above the threshold $\Theta \in \mathbb{R}$ is selected as a qualified individual (T, S^+), and otherwise as an unqualified one (T, S^-), with the purpose of assigning the qualified higher scores in comparison to the unqualified. More formally, our objective is to minimize the cost function $\mathcal{L}_{\text{rank}}$ defined as follows:

$$\mathcal{L}_{\text{rank}} = \text{Max}\Big(0, \Omega - Q(T, S^+) + Q(T, S^-)\Big) \tag{15}$$

where $\Omega \in \mathbb{R}$ denotes the marginal threshold. Once set up, the optimized latent semantic space will be responsible for inference only, and no longer get updated during training the SAG subsequently.

3.4 Quality-sensitive Loss Function

Conceptually, the seq2seq paradigm can be considered as a kind of conditional language model for the transduction tasks like machine translation [19, 29], in which the ground truths simply play the roles of data labels while their contained semantics are consistently ignored. It is fine if the every training process is quite perfect, but what if a significant portion of the training samples are not satisfactory? In this study, we aim at the SAG to put forward a quality-sensitive loss function $\mathcal{L}_{\text{qs}}$, as shown in Eq. 16.

$$\mathcal{L}_{\text{qs}} = Q(T,S) \times \mathcal{L}_{\text{neg}} + \Big(1 - Q(T,S)\Big) \times \mathcal{L}_{\text{ce}} \tag{16}$$

where $\mathcal{L}_{\text{neg}}$ and $\mathcal{L}_{\text{ce}}$ represent two different losses for the conventional negative log-likelihood (as Eq. 17) and the introduced cross entropy (as Eq. 18) respectively. Note that $\mathcal{L}_{\text{ce}}$ achieves the minimum value if and only if $Q(T,S) = Q(S,\bar{S})$.

$$\mathcal{L}_{\text{neg}} = -\log \Pr(S \mid T, \bar{S}) = -\sum_{j=1}^{m} \log \Pr^{\text{dis}}_{j, \mathbb{V}_{s_j}} \tag{17}$$

$$\begin{aligned} \mathcal{L}_{\text{ce}} = &-Q(T,S) \times \log Q(S,\bar{S}) \\ &- \Big(1 - Q(T,S)\Big) \times \log \Big(1 - Q(S,\bar{S})\Big) \end{aligned} \tag{18}$$

The rationale behind our loss allocation is as follows. Regarding the qualified case $Q(T,S) \uparrow$, the ground truth S should be utilized as much as possible to supervise the whole training phase, and consequently reproduce itself into the generated result $\bar{S}$, namely $\bar{S} \cong S \rightarrow \mathcal{L}_{\text{neg}} \downarrow$. However, in the opposite situation $Q(T,S) \downarrow$, we would like $\bar{S}$ to be as similar as the original input T within the latent semantic space $\phi_{\text{CTR}}(\cdot)$, namely $Q(T,\bar{S}) \cong 1 \rightarrow \bar{S} \cong T \rightarrow Q(T,S) \cong Q(S,\bar{S}) \rightarrow \mathcal{L}_{\text{ce}} \downarrow$. In particular, $\mathcal{L}_{\text{qs}}$ will be immediately reduced to the common quality-insensitive training mode once $Q(T,S) = 1$. To sum up, $Q(T,S)$ is exactly a conditioner for the model outcome $\bar{S}$ to decide between the expected answer S and the trivial response T. Note that $\phi_{\text{CTR}}(\cdot)$ is employed here to not only assess the user-centric qualities of training samples, but also forecast the textual fitness of S and newly observed $\bar{S}$.

In essence, the proposed quality-sensitive training mode equips the seq2seq model with a potential pipeline to exploit source paired texts from the semantic level. Especially for the unqualified case, we cannot expect that the ground truth therein could supply positive word-by-word supervision during the training period, and instead it should serve as a benchmark to restrict the semantic scopes of all results to be produced, i.e., to prevent the decoding stage from being too random and thus avoid $Q(T,S) \geq Q(T,\bar{S})$. It is mainly because that an incompatible social advertisement is often prone to puzzle audiences and further lessen the related purchase conversion rate, for which we only hope that $\bar{S}$ could preserve the semantics of T to the greatest extent. This is somewhat analogous to impose a specific orientation into the common beam search [29] that is frequently embedded at decoder.

4 EXPERIMENT

4.1 Experimental Setup

This section presents the experimental setup for assessing our approach, including 1) the dataset used for training and testing, 2)

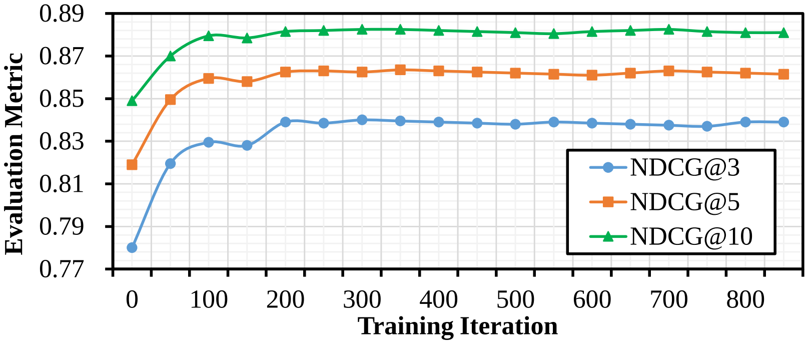

Figure 3: The optimization process of the latent semantic space, where the *normalize discounted cumulative gain* (NDCG) is applied to measure the accuracy of the pairwise ranking strategy.

the implementation details, 3) the evaluation metrics, and 4) the contrast methods.

Dataset We conduct extensive experiments on a dataset created from *Taobao*[6], in which all social advertisements are extracted from the associated product displays on search guide tabs (the heuristic rule: > 100 words and contain corresponding product names). Specifically, we select the clothing category (most actively traded) as our experimental subject and the dataset includes a total of 1.18M samples, where 1M of which are applied to build the latent semantic space, and the remaining 180K are used for developing the proposed seq2seq model. To be detailed, the 180K samples split into approximately 90% for training, 5% for validation, and 5% for testing. Once constructed, these samples are deployed in the context of purchase sharing via an online simulation environment (an instant messaging social app), and their CTR logs within a week are traced down to monitor the user click behavior. As a side note, we choose this dataset for the reason that no other open source data satisfy the condition to access an ocean of the CTR logs.

Implementation We use *Tensorflow*[7] for our implementations, in which the dimensions of word embeddings and hidden states are 128 and 256 respectively. More concretely, a vocabulary list of 70K words is built according to all training samples, and no pre-train operation is applied to the word embeddings in advance. Meanwhile, the maximum time step of encoder is set to 400, whereas that of decoder is fixed to 100 for training, and 35 for both validating and testing. Note that every LSTM parameter is randomly initialized over a uniform distribution within $[-0.02, 0.02]$. Based on these settings, we optimize the proposed seq2seq model using the AdaGrad [11] with a learning rate of 0.001, and perform the mini-batch gradient descent with a batch size of 16 for around 190K iterations (20 epochs in total). Moreover, as for the latent semantic space, we follow the established setup from [17] to utilize a stack of 3-layered DNNs for deriving the hierarchical text features, and the pairwise ranking strategy of [8] is adopted to tune the parameters therein. As can be seen from Fig. 3, the optimized space reaches an accuracy of at least 80% to discriminate the textual fitness for any pair of the product title and social advertisement.

Evaluation We take the widely used toolkit ROUGE [21] to evaluate the generated social advertisements automatically. To this end, we report both ROUGE-1 and ROUGE-2 (lexical overlaps for unigram and bigram) as a way to assess the informativeness, and ROUGE-L (the longest common subsequence) as a means to determine the fluency, in terms of the bytes of ground truths.

In addition, we conduct a human evaluation through the online simulation environment, on which a number of crowdsourcing workers are employed from *Datacrowd*[8] to compare between a list of social advertisements produced by all methods. Specifically, three criteria considered independently are as follows: 1) How *relevant* are the social advertisements to the corresponding product titles? 2) How *diverse* are the social advertisements for readers to easily grasp the expression changes? 3) How *interesting* are the social advertisements for audiences to proactively know about the product features? Note that we do not allow any correlation among workers during this evaluation, and each criterion is assigned with a 2-point scale of 0 (fail) and 1 (pass). By this means, each worker browses a set of 10 social advertisements of the same origin at random per session, and subsequently submits a score to judge every aspect of that paired method. To be exact, a total of 10 workers are responsible for each generated social advertisement, together with the Majority Voting [4] function to aggregate a final decision. Meanwhile, we also ask those workers to mark the ground truths so as to examine the raw qualities of whole experimental samples.

Benchmark To validate the proposed textual fitness, we compare our approach (denoted as $\text{P}_{\cdot\text{CTR}}$) against its four variants. The first one is to eliminate the cross entropy loss from Eq. 16, and thus adopt the common quality-insensitive training mode (denoted as $\text{P}_{\cdot\text{None}}$), which equates to setting $Q(T, S) = 1$ by default, namely no measurement for the textual fitness. The other three are all derived from the heuristic practice of [24], which applies the ROUGE as a substitute means to calculate the textual fitness from inputs to outputs, denoted as $\text{P}_{\cdot\text{ROUGE-1}}$, $\text{P}_{\cdot\text{ROUGE-2}}$, and $\text{P}_{\cdot\text{ROUGE-L}}$ respectively.

To better verify the effectiveness of our approach, we also select five representative NLG methods as another benchmark group, including: 1) Re3Sum: using approximate outputs as soft templates to guide the seq2seq model [5]; 2) Transformer: a sequence transduction model based entirely on the self-attention [30]; 3) PGenerator:

[6] https://www.taobao.com
[7] https://www.tensorflow.org
[8] https://datacrowd.taobao.com

a hybrid pointer-generator network with the trade-off between copying and producing [25]; 4) CopyNet: the first work to incorporate copying into the neural network-based seq2seq learning [16]; and 5) Seq2ASeq: a standard seq2seq model with the attention mechanism [1].

For clarity, $\text{P.}_{\text{ROUGE-1}}$, $\text{P.}_{\text{ROUGE-2}}$, $\text{P.}_{\text{ROUGE-L}}$ and Re3Sum belong to the two categories of existing solutions to address the qualities of training samples, where the former three focus on computing lexical overlaps, and the latter one turns to obtain additional information from similar outputs. Meanwhile, Transformer, PGenerator and CopyNet are three of the state-of-the-art solutions in the NLG domain, each with a different copying fashion to handle the OOV problem. Besides, Seq2ASeq is a classic model that integrates the seq2seq paradigm with the attention mechanism, which has witnessed a great success in various NLG applications like the machine translation.

4.2 Results and Discussion

Table 1 reports the ROUGE comparison between our approach and its four variants, in which P._{CTR} gets the highest score on all three evaluation metrics, whereas $\text{P.}_{\text{ROUGE-1}}$ receives the lowest ranking of the whole group. To be specific, we further draw the following conclusions:

1) P._{None} vs. P._{CTR}: The use of the proposed textual fitness does contribute to improve both the informativeness (ROUGE-1 and ROUGE-2) and fluency (ROUGE-L) of the generated social advertisements.

2) P._{None} vs. $\text{P.}_{\text{ROUGE-1}}$: The unigram overlaps cannot provide enough evidence to characterize the textual fitness for any pair of the product title and social advertisement that might experience a drastic vocabulary variation.

3) $\text{P.}_{\text{ROUGE-1}}$ vs. $\text{P.}_{\text{ROUGE-2}}$ vs. $\text{P.}_{\text{ROUGE-L}}$: The larger-grained word co-occurrences may present a deeper level of natural language understanding to capture the semantic affinity between product titles and social advertisements, and ROUGE-2 is the best choice among all its related versions to calculate the textual fitness, which conforms to the previous studies of [6, 7, 24, 34].

4) $\text{P.}_{\text{ROUGE-2}}$ vs. P._{CTR}: The CTR indeed has a better ability than the literal alignment ROUGE to distinguish whether a candidate social advertisement suits to a particular product title, owing to its reflecting consumer preference for the advertising.

To better illustrate the power of user click behavior, Fig. 4 visualizes the value distributions of various textual fitness over all experimental samples. From this graph, the curve of the proposed measurement is extremely similar to the Gaussian distribution, whereas those of the other counterparts have a relatively narrow range of span, and thus would suffer difficulties to differentiate the training samples of varied qualities accurately. In addition, Table 2 summarizes the correlation tests between the proposed textual fitness and each of ROUGE-1, ROUGE-2, and ROUGE-L separately. It can be seen that ROUGE-2 obtains a bit greater coefficient than the others, which could partly explain the reason why $\text{P.}_{\text{ROUGE-2}}$ ranks the second in Table 1.

Table 3 demonstrates the superiority of our approach over five representative NLG solutions, in which P._{CTR} gains the highest ranking of all six methods. Specifically, P._{CTR} achieves the maximum value on ROUGE-1 and ROUGE-2 simultaneously, but is slightly surpassed by Seq2ASeq and PGenerator respectively on ROUGE-L. However, the two counterparts' leading-edge is not statistically significant, and they both fall far behind our approach in terms of the informativeness (ROUGE-1 and ROUGE-2). To further explore the performance comparison, Table 4 presents the human evaluation on a total of 4,500 testing samples of varied qualities, where each percentage indicates the proportion of crowdsourcing workers who pass the corresponding criterion. Obviously, P._{CTR} still occupies the first place of the whole group, which means that the conclusion drawn by Table 3 is sustained. More concretely, we can reach the following findings:

1) Ground Truth ranks the third only, and therefore is by no means the gold standard for developing an ideal seq2seq model, which is consistent with our above analysis, i.e., constructing qualified training samples toward the SAG is not an easy matter.

2) Nearly no diversity and interestingness (below 5%) exists in the social advertisements generated by Seq2ASeq, CopyNet, and PGenerator. In addition, we also find that these three generic RNN-based solutions repeatedly return several frequent patterns like "*This is the newest*" of evidently sequential dependency, which results in their higher scores on ROUGE-L (fluency) in Table 3.

3) With the ability to draw global dependencies from inputs to outputs, Transformer significantly improves the scores on all three criteria in comparison to the former baselines. However, because of adopting the conventional quality-insensitive training mode, this method will more or less be polluted by those unqualified (what social media users are disinteresting of) training samples, which consequently causes an obvious gap in the interestingness with either P._{CTR} or Ground Truth.

4) While Re3Sum leverages the rough advertising templates to enhance both the diversity and interestingness, it still gets trapped into the poor relevance apparently due to introducing some irrelevant product details.

5) It is noteworthy that P._{CTR} obtains significantly better diversity than Ground Truth. This is partly because that P._{CTR} requires the ground truths to offer additional (user-centric) semantic constraints instead of merely word-by-word supervision for generating the social advertisements, which is particularly useful when facing the training samples with imbalanced textual fitness.

6) There is a large discrepancy in the evaluation results between ROUGE and crowdsourcing. For example, the ranking of Transformer in Table 4 is clearly higher than that in Table 3, and meanwhile it exceeds the lower expectation Seq2ASeq successfully, which to a certain degree proofs the usefulness and importance of human evaluation (user click behavior) in measuring the textual fitness.

4.3 Case Study

To better understand the effect of quality-sensitive training mode, Fig. 5 displays two typical examples from both the qualified and unqualified samples separately. For simplicity, we only compare the top three methods (P._{CTR}, Transformer, Re3Sum) in Table 4 against Ground Truth, and reasonably come to the following analysis:

1) While the rough advertising templates can offer a majority of proper details on sale promotion, Re3Sum still carries the risk to

Table 1: The ROUGE evaluation (%) of testing samples regarding our approach and its four variants.

Method \ Metric	ROUGE-1	ROUGE-2	ROUGE-L	Ranking
$P_{\cdot None}$	23.64*	6.46*	16.38*	2.0000
$P_{\cdot ROUGE\text{-}1}$	23.59*	6.45*	16.20*	1.0000
$P_{\cdot ROUGE\text{-}2}$	24.04*	6.50*	16.51*	3.8333
$P_{\cdot ROUGE\text{-}L}$	23.84*	6.50*	16.49*	3.1667
$P_{\cdot CTR}$	**24.39**	**7.14**	**17.10**	**5.0000**

1) the Ranking is returned by the Friedman test based on ROUGE-1, ROUGE-2, and ROUGE-L, and the higher the better;
2) * indicates the Wilcoxon signed-rank test with $p < 0.01$, in comparison to $P_{\cdot CTR}$.

Table 2: The correlation coefficients between the proposed textual fitness and each of ROUGE-1, ROUGE-2, and ROUGE-L separately.

Metric	ROUGE-1	ROUGE-2	ROUGE-L
the Proposed	0.1748	**0.2482**	0.2309

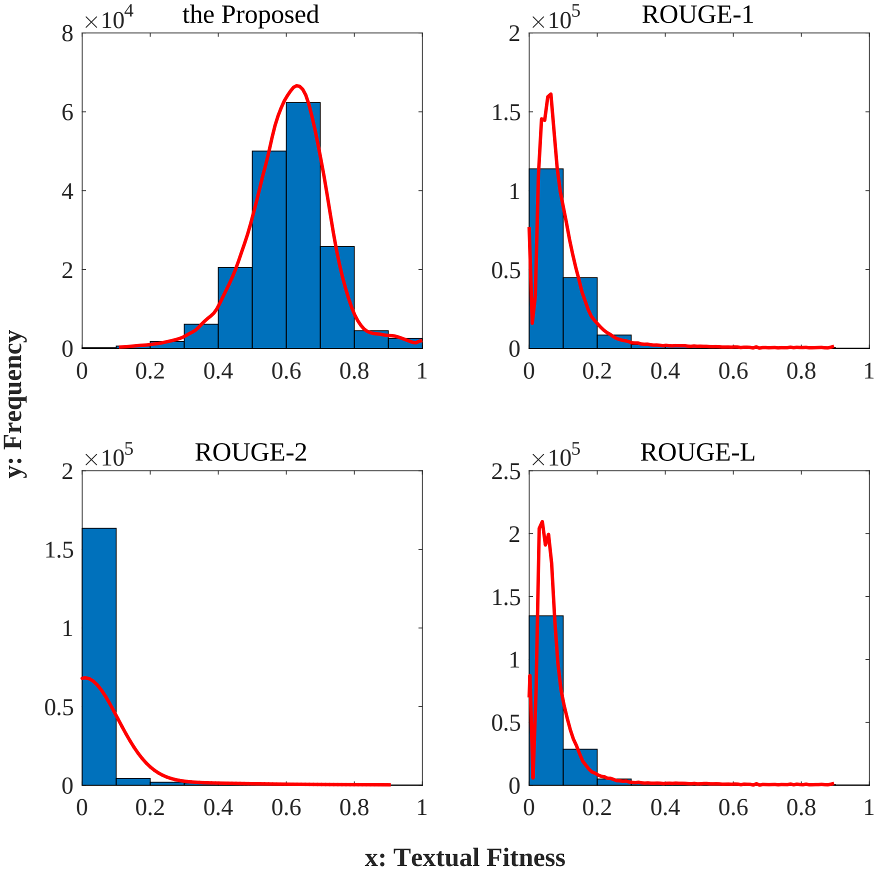

Figure 4: The frequency histograms and kernel density estimations of the proposed textual fitness, ROUGE-1, ROUGE-2, and ROUGE-L over all experimental samples.

Table 3: The ROUGE evaluation (%) of testing samples regarding our approach and five representative counterparts.

Metric Method	ROUGE-1	ROUGE-2	ROUGE-L	Ranking
Seq2ASeq	20.70*	4.37*	17.39	3.6667
CopyNet	20.24*	3.89*	16.79*	2.3333
PGenerator	20.85*	3.82*	**17.42**	3.6667
Transformer	21.60*	3.81*	15.53*	2.0000
Re3Sum	24.18	6.71*	16.69*	4.0000
P$_{\cdot\mathrm{CTR}}$	**24.39**	**7.14**	17.10	**5.3333**

1) the Ranking is returned by the Friedman test based on ROUGE-1, ROUGE-2, and ROUGE-L, and the higher the better;
2) * indicates the Wilcoxon signed-rank test with $p < 0.01$, in comparison to P$_{\cdot\mathrm{CTR}}$.

Table 4: The human evaluation (%) of testing samples regarding our approach and five representative counterparts.

Metric Method	Relevance	Diversity	Interestingness	Ranking
Seq2ASeq	72.52*	2.81*	4.10*	2.2500
CopyNet	25.48*	4.81*	0.37*	1.5000
PGenerator	96.91	2.90*	1.30*	3.2500
Transformer	**98.28**	89.80	53.73*	5.7500
Re3Sum	28.87*	70.69*	44.24*	3.5000
P$_{\cdot\mathrm{CTR}}$	95.41	**92.48**	**79.58**	**6.2500**
Ground Truth	95.74	75.30*	74.15	5.5000

1) the Ranking is returned by the Friedman test based on Relevance, Diversity, and Interestingness, and the higher the better;
2) * indicates the Wilcoxon signed-rank test with $p < 0.01$, in comparison to P$_{\cdot\mathrm{CTR}}$.

replace the given product names with the relevant yet inaccurate ones, just as in both the cases where "*hoodie*" and "*down pant*" respectively change into "*blouse*" and "*trousers*" of much higher generality.

2) Even though the self-attention can capture the correlation between specific products and associated descriptions, Transformer tends to produce some more common modifiers, such as "*round collar design → hoodie*" and "*with whatever T-shirts on → down pant*", which lacks of a few specialized adjectives toward the given product names. The reason is that, in terms of the conventional quality-insensitive training mode, the universal outcomes usually possess a lower damage to affect those unqualified training samples at decoding end.

3) Regarding the qualified case (left), P$_{\cdot\mathrm{CTR}}$ can almost acquire an amount of product details as equal as Ground Truth does. But surprisingly, the former creates a far better narrative than the latter in the opposite situation (right), where Ground Truth yields the correct product names only, whereas P$_{\cdot\mathrm{CTR}}$ generates several customized phrases with respect to the corresponding title contents. For instance, "*fits any figure well → slim*" and "*keeps warming → winter*". To sum up, the proposed quality-sensitive training mode provides the decoder a chance to take original inputs as a kind of backup ground truths and further consume the information therein.

5 CONCLUSION

In this paper, we highlight user click behavior in the context of the SAG via a latent semantic space that features the CTR, and on this basis, measure the textual fitness to quantitatively assess the user-centric qualities of training samples toward the seq2seq NLG. Notably, while social media users' interest varies over time, both the latent semantic space and quality assessment can change correspondingly. More specifically, we develop a novel seq2seq model to generate social advertisements automatically, in which a quality-sensitive loss function is proposed to characterize the quality distinction among training samples. In particular, for the qualified case, the ground truth should be utilized as much as possible to supervise the entire training process, but in the opposite situation, the generated result ought to retain the semantics of original input to the greatest extent. To sum up briefly, after an ocean of training samples are carefully projected into the CTR-based latent semantic space, only those relevant, diverse, and interesting contents will be oversampled to optimize the social advertisement generation. This strategy can be especially important for the purchase sharing concerning social media sites. Extensive experiments with real users demonstrate the validity of the proposed textual fitness, and the superiority of our approach over two representative methods of distant supervision and three state-of-the-art NLG solutions.

In future work, an appealing direction is to generalize our approach for other NLG applications as long as the user behavior information is available. For instance, in an explainable search scenario where the CTR logs are already prepared, we will be able to tailor the proposed seq2seq model so as to enhance the dynamic performance of content service.

Quality Level Training Sample		Qualified ✓	Unqualified ×
Product Title		秋季时尚字母Logo印花连帽套头卫衣 Autumn Fashion Hoodie Printed with Letter Logos	冬款白鹅绒通勤显瘦百搭羽绒裤 Winter Commuter Slim Casual Down Pant Made by White Goose Down
Social Advertisement	Ground Truth	纯色套头卫衣，简单大方，搭配胸前Logo印花，彰显时尚品味。 This is a solid-color simple and elegant hoodie, printed with letter logos on the chest to show your fashion taste.	找了一款羽绒裤的效果图供你们参考。 This is an effect drawing about the down pant for your reference.
	P.CTR	潮范儿十足的一款连帽卫衣，胸前的字母印花设计更显时尚潮流。 This is a hoodie full of fashion style, on which the letter printing design at chest highlights the latest trends.	休闲时尚的百搭羽绒裤，任何身形都可以驾驭，穿起来非常保暖。 This is a casual down pant featuring leisure fashion, which fits any figure well and keeps warming when in wearing.
	Transformer	这款卫衣采用了经典的圆领设计，让你的穿搭变得更加的时尚。 This hoodie has a classic round collar design, which makes your outfit more in fashion.	这款羽绒裤的版型是属于窄管的，随便搭配一件T恤就很好看。 This down pant follows a skinny shape and looks pretty with whatever T-shirts on.
	Re3Sum	遮肉显瘦的一款小衫，穿上就能很好的衬托出一整个人的时尚感。 This blouse helps you look slim and brings out a whole sense of fashion when dressed.	这款长裤采用的是麂皮绒的面料，打造属于时髦达人的冬季服饰。 This pair of trousers is made from suede fabrics to tailor-make winter clothes for fashionistas.

Figure 5: The case study, in which the phrases within each generated social advertisement are marked in red if they are directly related to the corresponding product title.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous referees for their valuable comments and helpful suggestions.

REFERENCES

[1] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* (2014).

[2] Dzmitry Bahdanau, Jan Chorowski, Dmitriy Serdyuk, Philemon Brakel, and Yoshua Bengio. 2016. End-to-end attention-based large vocabulary speech recognition. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*. IEEE, Shanghai, China, 4945–4949.

[3] Yi Bin, Yang Yang, Fumin Shen, Ning Xie, Heng Tao Shen, and Xuelong Li. 2018. Describing video with attention-based bidirectional LSTM. *IEEE Transactions on Cybernetics* PP(99) (2018), 1–11.

[4] Caleb Chen Cao, Jieying She, Yongxin Tong, and Lei Chen. 2012. Whom to Ask? Jury Selection for Decision Making Tasks on Micro-blog Services. *Proceedings of the Vldb Endowment* 5, 11 (2012), 1495–1506.

[5] Ziqiang Cao, Wenjie Li, Sujian Li, and Furu Wei. 2018. Retrieve, rerank and rewrite: Soft template based neural summarization. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. ACL, Melbourne, Australia, 152–161.

[6] Ziqiang Cao, Wenjie Li, Sujian Li, Furu Wei, and Yanran Li. 2016. Attsum: Joint learning of focusing and summarization with neural attention. In *Proceedings of the International Conference on Computational Linguistics*. ACM, Osaka, Japan, 547–556.

[7] Jianpeng Cheng and Mirella Lapata. 2016. Neural summarization by extracting sentences and words. In *Proceedings of the Annual Meeting of the Association for*

Computational Linguistics. ACL, Berlin, Germany, 484–494.

[8] Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. 2011. Natural language processing (almost) from scratch. *Journal of Machine Learning Research* 12, Aug (2011), 2493–2537.

[9] Scott Deerwester, Susan T Dumais, George W Furnas, Thomas K Landauer, and Richard Harshman. 1990. Indexing by latent semantic analysis. *Journal of the American society for information science* 41, 6 (1990), 391–407.

[10] Jan Deriu, Maurice Gonzenbach, Fatih Uzdilli, Aurelien Lucchi, Valeria De Luca, and Martin Jaggi. 2016. Swisscheese at semeval-2016 task 4: Sentiment classification using an ensemble of convolutional neural networks with distant supervision. In *Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics (Workshop)*. ACL, San Diego, CA, 1124–1128.

[11] John Duchi, Elad Hazan, and Yoram Singer. 2011. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research* 12, Jul (2011), 2121–2159.

[12] Jianfeng Gao, Xiaodong He, and Jian-Yun Nie. 2010. Clickthrough-based translation models for web search: from word models to phrase models. In *Proceedings of the International Conference on Information and Knowledge Management*. ACM, New York, USA, 1139–1148.

[13] Jianfeng Gao, Kristina Toutanova, and Wen-tau Yih. 2011. Clickthrough-based latent semantic models for web search. In *Proceedings of the International Conference on Research and Development in Information Retrieval*. ACM, Beijing, China, 675–684.

[14] Shalini Ghosh, Oriol Vinyals, Brian Strope, Scott Roy, Tom Dean, and Larry Heck. 2016. Contextual lstm (clstm) models for large scale nlp tasks. *arXiv preprint arXiv:1602.06291* (2016).

[15] Yu Gong, Xusheng Luo, Kenny Q Zhu, Shichen Liu, and Wenwu Ou. 2018. Automatic generation of Chinese short product titles for mobile display. *arXiv preprint arXiv:1803.11359* (2018).

[16] Jiatao Gu, Zhengdong Lu, Hang Li, and Victor OK Li. 2016. Incorporating copying mechanism in sequence-to-sequence learning. *arXiv preprint arXiv:1603.06393* (2016).

[17] Po-Sen Huang, Xiaodong He, Jianfeng Gao, Li Deng, Alex Acero, and Larry Heck. 2013. Learning deep structured semantic models for web search using clickthrough data. In *Proceedings of the International Conference on Information & Knowledge Management*. ACM, San Francisco, USA, 2333–2338.

[18] Thorsten Joachims. 2002. Optimizing search engines using clickthrough data. In *Proceedings of the International Conference on Knowledge Discovery and Data Mining*. ACM, New York, USA, 133–142.

[19] Nal Kalchbrenner and Phil Blunsom. 2013. Recurrent continuous translation models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. ACL, Seattle, USA, 1700–1709.

[20] Liangda Li, Hongbo Deng, Anlei Dong, Yi Chang, Ricardo Baeza-Yates, and Hongyuan Zha. 2017. Exploring query auto-completion and click logs for contextual-aware web search and query suggestion. In *Proceedings of the International World Wide Web Conferences*. ACM, Western, Australia, 539–548.

[21] Chin-Yew Lin and Eduard Hovy. 2003. Automatic evaluation of summaries using n-gram co-occurrence statistics. In *Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics*. ACL, Stroudsburg, USA, 71–78.

[22] Prashant Mathur, Nicola Ueffing, and Gregor Leusch. 2018. Multi-lingual neural title generation for e-Commerce browse pages. *arXiv preprint arXiv:1804.01041* (2018).

[23] Mike Mintz, Steven Bills, Rion Snow, and Dan Jurafsky. 2009. Distant supervision for relation extraction without labeled data. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. ACL, Stroudsburg, USA, 1003–1011.

[24] Ramesh Nallapati, Feifei Zhai, and Bowen Zhou. 2017. SummaRuNNer: A recurrent neural network based sequence model for extractive summarization of documents. In *Proceedings of the AAAI Conference on Artificial Intelligence*. AAAI, San Francisco, USA, 3075–3081.

[25] Abigail See, Peter J Liu, and Christopher D Manning. 2017. Get to the point: Summarization with pointer-generator networks. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. ACL, Vancouver, Canada, 1073–1083.

[26] Iulian Vlad Serban, Alessandro Sordoni, Ryan Lowe, Laurent Charlin, Joelle Pineau, Aaron C Courville, and Yoshua Bengio. 2017. A hierarchical latent variable encoder-decoder model for generating dialogues. In *Proceedings of the AAAI Conference on Artificial Intelligence*. AAAI, San Francisco, USA, 3295–3301.

[27] Henning Sperr, Jan Niehues, and Alex Waibel. 2013. Letter n-gram-based input encoding for continuous space language models. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. ACL, Sofia, Bulgaria, 30–39.

[28] Fei Sun, Peng Jiang, Hanxiao Sun, Changhua Pei, Wenwu Ou, and Xiaobo Wang. 2018. Multi-source pointer network for product title summarization. *arXiv preprint arXiv:1808.06885* (2018).

[29] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In *Proceedings of the Annual Conference on Neural Information Processing Systems*. MIT Press, MontrÃĺal, Canada, 3104–3112.

[30] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Proceedings of the Annual Conference on Neural Information Processing Systems*. MIT Press, Long Beach, USA, 5998–6008.

[31] Gang Wang, Xinyi Zhang, Shiliang Tang, Christo Wilson, Haitao Zheng, and Ben Y Zhao. 2017. Clickstream user behavior models. *ACM Transactions on the Web* 11, 4 (2017), 21.

[32] Jinpeng Wang, Yutai Hou, Jing Liu, Yunbo Cao, and Chin-Yew Lin. 2017. A statistical framework for product description generation. In *Proceedings of the International Joint Conference on Natural Language Processing (Short Paper)*. ACL, Taipei, 187–192.

[33] Jingang Wang, Junfeng Tian, Long Qiu, Sheng Li, Jun Lang, Luo Si, and Man Lan. 2018. A Multi-task learning approach for improving product title compression with user search log data. In *Proceedings of the AAAI Conference on Artificial Intelligence*. AAAI, New Orleans, USA, 451–458.

[34] Yongzhen Wang, Xiaozhong Liu, and Zheng Gao. 2018. Neural Related Work Summarization with a Joint Context-driven Attention Mechanism. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. ACL, Brussels, Belgium, 1776–1786.