Automatic Generation of Personalized Comment Based on User Profile

Wenhuan Zeng[1]*, **Abulikemu Abuduweili**[2]*, **Lei Li**[3], **Pengcheng Yang**[4]
[1]School of Mathematical Sciences, Peking University
[2]State Key Lab of Advanced Optical Communication System and Networks,
School of EECS, Peking University
[3]School of Computer Science and Technology, Xidian University
[4]MOE Key Lab of Computational Linguistics, School of EECS, Peking University
{zengwenhuan, abduwali}@pku.edu.cn
tobiaslee@foxmail.com, yang_pc@pku.edu.cn

Abstract

Comments on social media are very diverse, in terms of content, style and vocabulary, which make generating comments much more challenging than other existing natural language generation (NLG) tasks. Besides, since different user has different expression habits, it is necessary to take the user's profile into consideration when generating comments. In this paper, we introduce the task of automatic generation of personalized comment (AGPC) for social media. The dataset we used contains tens of thousands of users' real comments and corresponding user profiles collected from Weibo, one of the most popular social media in China. We propose Personalized Comment Generation Network (PCGN) for AGPC. The model utilizes user feature embedding with a gated memory and attends to user description to model personality of users. In addition, external user representation is taken into consideration during the decoding to enhance the comments generation. Experimental results show that our model can generate natural, human-like and personalized comments.[1]

1 Introduction

Nowadays, social media is gradually becoming a mainstream communication tool. People tend to share their ideas with others by commenting, reposting or clicking *like* on posts in social media. Among these behaviors, *comment* plays a significant role in the communication between posters and readers. Automatically generate personalized comments (AGPC) can be useful due to the following reasons. First, AGPC helps readers express their ideas more easily, thus make them engage more actively in the platform. Second, bloggers can capture different attitudes to the event from multiple users with diverse backgrounds. Lastly, the platform can also benefit from the increasing interactive rate.

Despite its great applicability, the AGPC task faces two important problems: whether can we achieve it and how to implement it? The *Social Differentiation Theory* proposed by Riley and Riley (1959) proved the feasibility of building a universal model to automatically generate personalized comments based on part of users' data. The *Individual Differences Theory* pointed by Hovland et al. (1953) answers the second question by introducing the significance of users' background, which inspires us to incorporating user profile into comments generation process. More specifically, the user profile consists of demographic features (for example, where does the user live), individual description and the common word dictionary extracted from user's comment history. There are few works exploring the comments generation problem. Zheng et al. (2017) first paid attention to generating comments for news articles by proposing a gated attention neural network model (GANN) to address the contextual relevance and the diversity of comments. Similarly, Qin et al. (2018) introduced the task of automatic news article commenting and released a large scale Chinese corpus. Nevertheless, AGPC is a more challenging task, since it not only requires generating relevant comments given the blog text, but also needs the consideration of the diverse users' background.

In this paper, we propose a novel task, automatically generating personalized comment based on user profile. We build the bridge between user profiles and social media comments based on a large-scale and high-quality Chinese dataset. We elaborately design a generative model based on sequence-to-sequence (Seq2Seq) framework. A gated memory module is utilized to model the user

*Equal Contribution.

[1]Source codes of this paper are available at https://github.com/Walleclipse/AGPC

personality. Besides, during the decoding process, the model attends to user description to enhance the comments generation process. In addition, the vocabulary distribution of generated word is adapted by considering the external user representation.

Our main contributions are as follows:

- We propose the task of automatic generating personalized comment with exploiting user profile.
- We design a novel model to incorporate the personalized style in large-scale comment generation. The model has three novel mechanisms: user feature embedding with gated memory, blog-user co-attention, and an external personality expression.
- Experimental results show that the proposed method outperforms various competitive baselines by a large margin. With novel mechanisms to exploit user information, the generated comments are more diverse and informative. We believe that our work can benefit future work on developing personalized and human-like NLG model.

2 Personalized Comments Dataset

We introduce the dataset as follows:

Data Preparation We collect short text posts from Weibo, one of the most popular social media platform in China, which has hundreds of millions of active users. Each instance in the dataset has province, city, gender, age, marital status, individual description of user's, comment added by user and homologous blog content. Figure 1 visually shows a sample instance. We tokenized all text like individual description, comment and blog content into words, using a popular python library Jieba[2]. To facilitate the model to learn valid information from the dataset, we removed @, url, expressions in the text, and unified Chinese into simplified characters. Discrete variables such as province, city, gender and marital status were treated uniformly by one-hot coding. To ensure the quality of text, we filtered out samples with less than two words in the variable of comment and blog content. Besides, in order to learn user-specific expression habits, we retain users with 50 or more records. The resulting dataset contains 4,463,767 comments on 1,495,822 blog posts by 32,719 users.

[2]https://github.com/fxsjy/jieba

UID: 215803	**Age:** 24	**Birthday:** 1994-01-21
Gender: 女 Female	**Province:** 上海 Shanghai	**City:** 未知 NULL
Individual Description: 笨鸟一直飞 Practice makes prefect		
Blog: 医生开了新药，吃了胃会不舒服。。要是所有的事情都是梦就好了 The doctor prescribed the new medicine which let my stomach uncomfortable. If only everything were a dream.		
Comment: 一切都会好起来的 Everything will be ok.		

Figure 1: A data example in personalized comment dataset. Corresponding English translation is provided.

Statistic	User	Comment	Microblog
Train	32,719	2,659,870	1,450,948
Dev	24,739	69,659	27,822
Test	20,157	43,866	17,052
Total	32,719	4,463,767	1,495,822

Table 1: Sample size of three datasets

Data Statistics We split the corpus into training, validation and testing set according to the microblog. To avoid overfitting, the records of the same microblog will not appear in the above three sets simultaneously. Table 1 displays the detail sample size of user, comment and blog about training set, validation set and testing set. Each user in the resulting dataset has an average of 56 samples. The average lengths of blog post, comment and individual description are 50, 11 and 9 words, respectively. The particular statistics of each experimental dataset are shown in Table 2.

Average length	Train	Dev	Test	Total
ID	8.84	9.04	8.83	8.85
Comment	11.28	11.32	11.86	11.28
Microblog	49.67	47.95	50.30	49.65

Table 2: Statistics of text variables. Individual description, abbreviated ID.

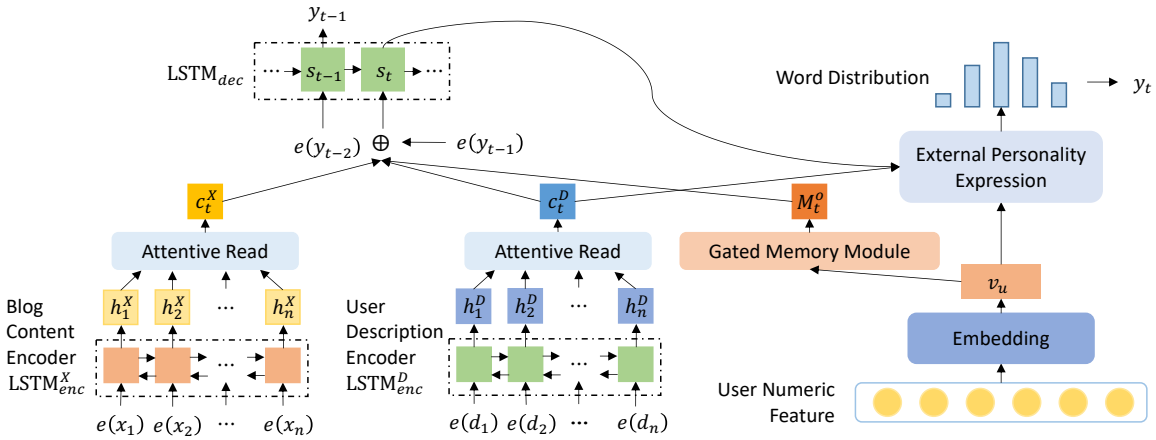

Figure 2: Personalized comment generation network

3 Personalized Comment Generation Network

Given a blog $X = (x_1, x_2, \cdots, x_n)$ and a user profile $U = \{F, D\}$, where $F = (f_1, f_2, \cdots, f_k)$ denotes the user's numeric feature (for example, age, city, gender) and $D = (d_1, d_2, \cdots, d_l)$ denotes the user's individual description, the AGPC aims at generating comment $Y = (y_1, y_2, \cdots, y_m)$ that is coherent with blog X and user U. Figure 2 presents an overview of our proposed model, which is elaborated on in detail as follows.

3.1 Encoder-Decoder Framework

Our model is based on the encoder-decoder framework of the general sequence-to-sequence (Seq2Seq) model (Sutskever et al., 2014). The encoder converts the blog sequence $X = (x_1, x_2, \cdots, x_n)$ to hidden representations $h^X = (h_1^X, h_2^X, \cdots, h_n^X)$ by a bi-directional Long Short-Term Memory (LSTM) cell (Hochreiter and Schmidhuber, 1997):

$$h_t^X = \text{LSTM}_{\text{enc}}^{\text{X}}(h_{t-1}^X, x_t) \tag{1}$$

The decoder takes the embedding of a previously decoded word $e(y_{t-1})$ and a blog context vector c_t^X as input to update its state s_t:

$$s_t = \text{LSTM}_{\text{dec}}(s_{t-1}, [c_t^X; e(y_{t-1})]) \tag{2}$$

where $[\cdot;\cdot]$ denotes vector concatenation. The context vector c_t^X is a weighted sum of encoder's hidden states, which carries key information of the input post (Bahdanau et al., 2014). Finally, the decoder samples a word y_t from the output probability distribution as follows

$$y_t \sim \text{softmax}(\mathbf{W_o} s_t) \tag{3}$$

where $\boldsymbol{W_o}$ is a weight matrix to be learned. The model is trained via maximizing the log-likelihood of ground-truth $Y^* = (y_1^*, \cdots, y_n^*)$ and the objective function is defined as

$$\mathcal{L} = -\sum_{t=1}^{n} \log\Big(p(y_t^* | y_{<t}^*, X, U)\Big) \tag{4}$$

3.2 User Feature Embedding with Gated Memory

To encode the information in user profile, we map user's numeric feature F to a dense vector v_u through a fully-connected layer. Intuitively, v_u can be treated as a user feature embedding denotes the character of the user. However, if the user feature embedding is static during decoding, the grammatical correctness of sentences generated may be sacrificed as argued in Ghosh et al. (2017). To tackle this problem, we design an gated memory module to dynamically express personality during decoding, inspired by Zhou et al. (2018). Specifically, we maintain a internal personality state during the generation process. At each time step, the personality state decays by a certain amount. Once the decoding process is completed, the personality state is supposed to decay to zero, which indicates that the personality is completely expressed. Formally, at each time step t, the model computes an update gate ${g^u}_t$ according to the current state of the decoder s_t. The initial personality state M_0 is set as user feature embedding v_u. Hence, the personality state M_t is erased by a certain amount (by g_t^u) at each step. This process is described as

$$g_t^u = \text{sigmoid}(\mathbf{W_g^u} s_t) \tag{5}$$

$$M_0 = v_u \tag{6}$$

$$M_t = g_t^u \otimes M_{t-1}, \quad t > 0 \tag{7}$$

where $\otimes$ denotes element-wise multiplication. Besides, the model should decide how much atten-

tion should be paid to the personality state at each time step. Thus, output gate g_t^o is introduced to control the information flow by considering the previous decoder state s_{t-1}, previous target word $e(y_{t-1})$ and the current context vector c_t^X

$$g_t^o = \text{sigmoid}(\mathbf{W_g^o}[s_{t-1}; e(y_{t-1}); c_t^X]). \quad (8)$$

By an element-wise multiplication of g_t^o and M_t, we can obtain adequate personality information M_t^o for current decoding step

$$M_t^o = g_t^o \otimes M_t. \quad (9)$$

3.3 Blog-User Co-Attention

Individual description is another important information source when generating personalized comments. For example, a user with individual description "只爱朱一龙" (I only love Yilong Zhu[3]), tends to writes a positive and adoring comments on the microblog related to Zhu. Motivated by this, we propose Blog-user co-attention to model the interactions between user description and blog content. More specifically, we encode the user's individual description $D = (d_1, d_2, \cdots, d_l)$ to hidden states $(h_1^D, h_2^D, \cdots, h_l^D)$ via another LSTM

$$h_t^D = \text{LSTM}_{\text{enc}}^{\text{D}}(h_{t-1}^D, d_t) \quad (10)$$

We can obtain a description context vector c_t^D by attentively reading the hidden states of user description,

$$c_t^D = \sum_j \alpha_{tj} h_j^D \quad (11)$$

$$\alpha_{tj} = \text{softmax}(e_{tj}) \quad (12)$$

$$e_{tj} = s_{t-1} \mathbf{W_a} h_j^D \quad (13)$$

where e_{tj} is a alignment score (Bahdanau et al., 2014). Similarly, we can get the blog content vector c_t^X. Finally, the context vector c_t is a concatenation of c_t^X and c_t^D, in order provide more comprehensive information of user's personality

$$c_t = [c_t^X; c_t^D] \quad (14)$$

Therefore, the state update mechanism in Eq.(2) is modified to

$$s_t = \text{LSTM}_{\text{dec}}(s_{t-1}, [c_t; e(y_{t-1}); M_t^o]) \quad (15)$$

[3]A famous Chinese star.

3.4 External Personality Expression

In the gated memory module, the correlation between the change of the internal personality state and selection of a word is implicit. To fully exploit the user information when selecting words for generation, we first compute a user representation r_t^u with user feature embedding and user description context.

$$r_t^u = \mathbf{W_r}[v_u; c_t^D] \quad (16)$$

where $\mathbf{W_r}$ is a weight matrix to align user representation dimention.

The final word is then sampled from output distribution based on the concatenation of decoder state s_t and r_t^u as

$$\tilde{y_t} \sim \text{softmax}(\mathbf{W_{\tilde{o}_t}}[\text{s}_\text{t}; \text{r}_\text{t}^\text{u}]) \quad (17)$$

where $\mathbf{W_{\tilde{o}_t}}$ is a learnable weight matrix.

4 Experiments

4.1 Implementation

The blog content encoder and comment decoder are both 2-layer bi-LSTM with 512 hidden units for each layer. The user's personality description encoder is a single layer bi-LSTM with 200 hidden units. The word embedding size is set to 300 and vocabulary size is set to 40,000. The embedding size of user's numeric feature is set to 100.

We adopted beam search and set beam size to 10 to promote diversity of generated comments. We used SGD optimizer with batch size set to 128 and the learning rate is 0.001.

To further enrich the information provided by user description, we collected most common k words in user historical comments ($k = 20$ in our experiment). We concatenate the common words with the user individual description. Therefore, we can obtain more information about users' expression style. The model using concatenated user description is named PCGN with common words (PCGN+ComWord).

4.2 Baseline

We implemented a general Seq2Seq model (Sutskever et al., 2014) and a user embedding model (Seq2Seq+Emb) proposed by Li et al. (2016) as our baselines. The latter model embeds user numeric features into a dense vector and feeds it as extra input into decoder at every time step.

Method	PPL	B-2	METEOR
Seq2Seq	32.47	0.071	0.070
Seq2Seq+Emb	31.13	0.084	0.079
PCGN	27.94	0.162	0.132
PCGN+ComWord	**24.48**	**0.193**	**0.151**

Table 3: Automatic evaluation results of different methods. **PPL** denotes perplexity and **B-2** denotes BLEU-2. Best results are shown in bold.

Method	PPL	B-2
Seq2Seq	32.47	0.071
+ Mem	30.73 (-1.74)	0.099 (+0.028)
+ CoAtt	27.12 (-3.61)	0.147 (+0.078)
+ External	27.94 (+0.82)	0.162 (+0.015)

Table 4: Incremental experiment results of proposed model. Performance on METEOR is similar to B-2. **Mem** denotes gated memory, **CoAtt** denotes blog-user co-attention and **External** denotes external personality expression

4.3 Evaluation Result

Metrics: We use BLEU-2 (Papineni et al., 2002) and METEOR (Banerjee and Lavie, 2005) to evaluate overlap between outputs and references. Besides, perplexity is also provided.

Results: The results are shown in Table 3. As can be seen, PCGN model with common words obtains the best performance on perplexity, BLEU-2 and METEOR. Note that the performance of Seq2Seq is extremely low, since the user profile is not taken into consideration during the generation, resulting repetitive responses. In contrast, with the help of three proposed mechanism (gated memory, blog-user co-attention and external personality expression), our model can utilize user information effectively, thus is capable of generating diverse and relevant comments for the same blog. Further, we conducted incremental experiments to study the effect of proposed mechanisms by adding them incrementally, as shown in Table 4. It can be found that all three mechanism help generate more diverse comments, while blog-user co-attention mechanism contributes most improvements. An interesting finding is that external personality expression mechanism causes the decay on perplexity. We speculate that the modification on word distribution by personality influence the fluency of generated comments.

5 Related Work

This paper focuses on comments generation task, which can be further divided into generating a comment according to the structure data (Mei et al., 2015), text data (Qin et al., 2018), image (Vinyal et al., 2015) and video (Ma et al., 2018a), separately.

There are many works exploring the problem of text-based comment generation. Qin et al. (2018) contributed a high-quality corpus for article comment generation problem. Zheng et al. (2017) proposed a gated attention neural network model (GANN) to generate comments for news article, which addressed the contextual relevance and the diversity of comments. To alleviate the dependence on large parallel corpus, Ma et al. (2018b) designed an unsupervised neural topic model based on retrieval technique. However, these works focus on generating comments on news text, while comments on social media are much more diverse and personal-specific.

In terms of the technique for modeling user character, the existing works on machine commenting only utilized part of users' information. Ni and McAuley (2018) proposed to learn a latent representation of users by utilizing history information. Lin et al. (2018) acquired readers' general attitude to event mentioned by article through its upvote count. Compared to the indirection information obtained from history or indicator, user features in user profile, like demographic factors, can provide more comprehensive and specific information, and thus should be paid more attention to when generating comments. Sharing the same idea that user personality counts, Luo et al. (2018) proposed personalized MemN2N to explore personalized goal-oriented dialog systems. Equipped with a profile model to learn user representation and a preference model learning user preferences, the model is capable of generating high quality responses. In this paper, we focus on modeling personality in a different scenario, where the generated comments is supposed to be general and diverse.

6 Conclusion

In this paper, we introduce the task of automatic generating personalized comment. We also propose Personality Comment Generation Network (PCGN) to model the personality influence in comment generation. The PCGN model utilized

gated memory for user feature embedding, blog-user co-attention, and external personality representation to generate comments in personalized style. Evaluation results show that PCGN outpeforms baseline models by a large margin. With the help of three proposed mechanisms, the generated comments are more fluent and diverse.

References

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

Satanjeev Banerjee and Alon Lavie. 2005. Meteor: An automatic metric for mt evaluation with improved correlation with human judgments. In *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*, pages 65–72.

Sayan Ghosh, Mathieu Chollet, Eugene Laksana, Louis-Philippe Morency, and Stefan Scherer. 2017. Affect-lm: A neural language model for customizable affective text generation. *arXiv preprint arXiv:1704.06851*.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8):1735–1780.

Carl I. Hovland, Irving L. Janis, and Harold H. Kelley. 1953. *Communication and Persuasion:Psychological Studies of Opinion Change*. Yale University Press.

Jiwei Li, Michel Galley, Chris Brockett, Georgios P Spithourakis, Jianfeng Gao, and Bill Dolan. 2016. A persona-based neural conversation model. *arXiv preprint arXiv:1603.06155*.

Zhaojiang Lin, Genta Indra Winata, and Pascale Fung. 2018. Learning comment generation by leveraging user-generated data. *CoRR*, abs/1810.12264.

Liangchen Luo, Wenhao Huang, Qi Zeng, Zaiqing Nie, and Xu Sun. 2018. Learning personalized end-to-end goal-oriented dialog.

Shuming Ma, Lei Cui, Damai Dai, Furu Wei, and Xu Sun. 2018a. Livebot: Generating live video comments based on visual and textual contexts. *CoRR*, abs/1809.04938.

Shuming Ma, Lei Cui, Furu Wei, and Xu Sun. 2018b. Unsupervised machine commenting with neural variational topic model. *arXiv preprint arXiv:1809.04960*.

Hongyuan Mei, Mohit Bansal, and Matthew R. Walter. 2015. What to talk about and how? selective generation using lstms with coarse-to-fine alignment. *arXiv preprint arXiv:1509.00838*.

Jianmo Ni and Julian McAuley. 2018. Personalized review generation by expanding phrases and attending on aspect-aware representations. In *In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pages 706–711.

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*, pages 311–318. Association for Computational Linguistics.

Lianhui Qin, Lemao Liu, Wei Bi, Yan Wang, Xiaojiang Liu, Zhiting Hu, Hai Zhao, and Shuming Shi. 2018. Automatic article commenting: the task and dataset. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 2: Short Papers*, pages 151–156.

J. W. Riley and Matilda White Riley. 1959. Mass communication and the social system.

Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112.

Oriol Vinyal, Alexander Toshev, Samy Bengio, and Dumitru Erhan. 2015. Show and tell: A neural image caption generator. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3156–3164.

Hai-Tao Zheng, Wei Wang, Wang Chen, and Arun Kumar Sangaiah. 2017. Automatic generation of news comments based on gated attention neural networks. 6.

Hao Zhou, Minlie Huang, Tianyang Zhang, Xiaoyan Zhu, and Bing Liu. 2018. Emotional chatting machine: Emotional conversation generation with internal and external memory. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

A Case Study

We present some generated cases in Figure 4, 5. There are multiple users (corresponding profiles are shown in Figure 3) that are suitable for generating comments. Seq2Seq generates same comments for the same blog, while PCGN can generate personalized comment conditioned on given user. According to the user profile, U1 adores Yilong Zhu very much. Therefore, U1 tends to express her affection in comments when responses to blogs related to Yilong Zhu. For users whose individual descriptions can not offer helpful information or there is missing value for individual description, the PCGN model pays more attention to numeric features and learns representation from similar seen users.

User	Age	Gender	Province	City	Individual Description
U1	24	女 Female	其他 Others	NULL	只爱朱一龙 I only love Yilong Zhu
U2	23	女 Female	黑龙江 Heilong Jiang	NULL	努力成为更好的自己 Become a better me
U3	20	女 Female	浙江 Zhejiang	宁波 Ningbo	NULL

Figure 3: Part of user profile of case study users. In order to protect user privacy, the birthday variable is not shown here.

Blog
朱一龙 温柔的力量 [超话] # [小仙女] # 朱一龙并肩前行 # 遇见你，是旷野的风闯进心房，是眉间的闯进眼眶。想把一切与你分享，清晨的暖阳、浩瀚的夜空、过去的美好、 未来的相伴。「朱一龙」 Yilong Zhu Gentle power [super topic]# walking side by side with Yilong Zhu#The moment I met you seems like the wind of the wilderness, breaking into my heart. The moment I met you seems like the snow between the eyebrows, blending into the eyes. I want to share everything with you, the warm sun in the morning, the vast night sky, the beauty of the past and the companion in the future.
Comments
Seq2Seq: # 朱一龙温柔的力量 [超话] # # 朱一龙并肩前行 # #Yilong Zhu Gentle power [super topic]# #Yilong Zhu, move forward together#
PCGN U1: 『朱一龙』甜有 100 种方式，吃糖，还有每天 99 次的想你。 There is one hunderd ways of sweetness, have a candy and miss you 99 times a day.
PCGN U2: #朱一龙温柔的力量 [超话] # #朱一龙 并肩前行 # 朱一龙 | ZYL #Yilong Zhu Gentle power [super topic]# #Yilong Zhu, move forward together# Yilong Zhu| ZYL
PCGN U3: 『 朱一龙 』愿你一直如少年，干净纯粹心安， 看透不美好却相信美好 I hope that you are always young, with a clean and pure heart, always believing something beautiful

Figure 4: Generated comments based on blog of different users. Since Seq2Seq model does not take user profile into consideration, it generates same comments for the same blog.

Blog
我的真朋友# 运用日剧和漫画的镜头切割，芭莎特别策划打造视觉大片，将三位主演的剧中人物关系呈现在视觉大片里，让你在放映前先睹为快! Angelababy 发型/刘雪亮 Angelababy 化妆/春楠 邓伦妆发/李健成 朱一龙妆发/李鹏坤 #My true friend # Using the lens of Japanese TV dramas and comics, Bazaar specially plans to create visual blockbusters, and present the relationship among the three main characters in the photo, which will give you a sneak of the movie before its showing! Angelababy Hairstyle / Liu Xueliang Angelababy Makeup / Chun Nan Lun Deng makeup hair / Jiancheng Li Yilong Zhu makeup hair / Pengkun Li
Comments
Seq2Seq: # angelababy [超话] # # angelababy [super topic] #
PCGN U1: # 朱一龙[超话] # # 朱一 龙井然 # 期待井然哥哥 Yilong Zhu[super topic]# #朱一龙井然# Looking forward to Jingan brother
PCGN U2: 期待 期待 looking forward to
PCGN U3: 期待邓伦 looking forward to Lun Deng

Figure 5: Generated comments based on blog of different users.