

# Estimating User Interest from Open-Domain Dialogue

Michimasa Inaba      Kenichi Takahashi

Hiroshima City University

3-4-1 Ozukahigashi, Asaminami-ku, Hiroshima, Japan

{inaba, takahashi}@hiroshima-cu.ac.jp

## Abstract

Dialogue personalization is an important issue in the field of open-domain chat-oriented dialogue systems. If these systems could consider their users' interests, user engagement and satisfaction would be greatly improved. This paper proposes a neural network-based method for estimating users' interests from their utterances in chat dialogues to personalize dialogue systems' responses. We introduce a method for effectively extracting topics and user interests from utterances and also propose a pre-training approach that increases learning efficiency. Our experimental results indicate that the proposed model can estimate user's interest more accurately than baseline approaches.

## 1 Introduction

Chat is a very important part of human communication. In fact, it has been reported that it makes up about 62% of all conversations (Koiso et al., 2016). Since chat is also important for human-to-machine communication, studies of dialogue systems that aim to enable open-domain chat have received much attention in recent years (Ritter et al., 2011; Higashinaka et al., 2014; Sordani et al., 2015; Vinyals and Le, 2015; Zhao et al., 2017). In these studies, dialogue personalization is an important issue: if such systems could consider users' experiences and interests when engaging them in a conversation, it would greatly improve user satisfaction. To this end, Hirano et al. extracted predicate-argument structures (Hirano et al., 2015), Zhang and Chai focused on conversational entailment (Zhang and Chai, 2009, 2010) and Bang et al. extracted entity relationships (Bang et al., 2015). These studies aimed to

employ users' utterance histories to generate personalized responses.

In contrast, this study aims to estimate the user's interest in particular topics (e.g., music, fashion, or health) to personalize the dialogue system's responses based on these interests. This would allow it to focus on topics the user is interested in and avoid topics they dislike, enhancing user engagement and satisfaction.

This paper therefore proposes a neural network-based method for estimating users' interests using their utterances in chat dialogues. Our method estimates their levels of interest not only in topics that appear in the dialogues, but also in other topics that have not appeared. Even if a user enjoys talking about the current topic, they will get bored if the system talks about it endlessly. By gauging the user's potential interest in topics that have not directly appeared in the dialogue, the system can expand the discussion to other topics before the user gets bored.

In this study, we use data from human-to-human dialogues because the current performance of chat-oriented dialogue systems is not sufficient for them to talk with humans naturally. We also use textual dialogue data to avoid speech recognition issues. In addition, to estimate the target user's interests independently of the dialogue system's utterances, we only consider their own utterances and ignore those of their dialogue partner.

This paper brings three main contributions, as follows. 1. We propose a topic-specific sentence attention approach that enables topics and user interests to be efficiently extracted from utterances. 2. We develop a method for pre-training our model's utterance encoder, so it learns what topics are related to each target user's utterance. 3. We show experimentally that the proposed sentence attention and pre-training methods can provide high performance when used together.

## 2 Related Work

Many studies related to estimating user interest from text data have targeted social network services (SNS), especially Twitter users. For example, Chen et al. proposed a method of modeling interest using the frequencies of words in tweets by the target user and followers (Chen et al., 2010). Some methods have also been proposed that consider superordinate concepts acquired from knowledge bases. For example, Abel et al. modeled Twitter users using the appearance frequencies of certain named entities (e.g., people, events, or music groups), acquired using OpenCalais<sup>1</sup> (Abel et al., 2011). In addition, some methods have used categories from Wikipedia (Michelson and Macskassy, 2010; Kapanipathi et al., 2014; Zarrinkalam et al., 2015) or DBpedia (Kapanipathi et al., 2011). Several methods have also been proposed that use topic models, such as latent dirichlet allocation (LDA) (Weng et al., 2010; Bhattacharya et al., 2014; Han and Lee, 2016). However, it is difficult to apply such methods directly to dialogue because they assume that users are posting about subjects they are interested in. This is a reasonable assumption for SNS data, but in conversations, people do not always limit themselves to topics they are interested in. For instance, people will play along and discuss subjects the other persons are interested in, even if they are not interested in them, as well.

Other studies have attempted to estimate users’ levels of interest (LOI) from dialogues. Schuller et al. tackled the task of estimating listeners’ interest in a product from dialogues between them and someone introducing a particular product, proposing a support vector machine (SVM)-based method incorporating acoustic and linguistic features (Schuller et al., 2006). In 2010, LOI estimation was selected as a sub-challenge of the INTER-SPEECH Paralinguistic Challenge (Schuller et al., 2006), but there the focus was on single-topic (product) interest estimation from spoken dialogue, not open-domain estimation. In addition, that task considered business dialogues, not chats.

## 3 Model Architecture

The task considered in this paper is as follows. Given an utterance set  $U_s = (u_1, u_2, \dots, u_n)$  ut-

<sup>1</sup><http://www.opencalais.com/>

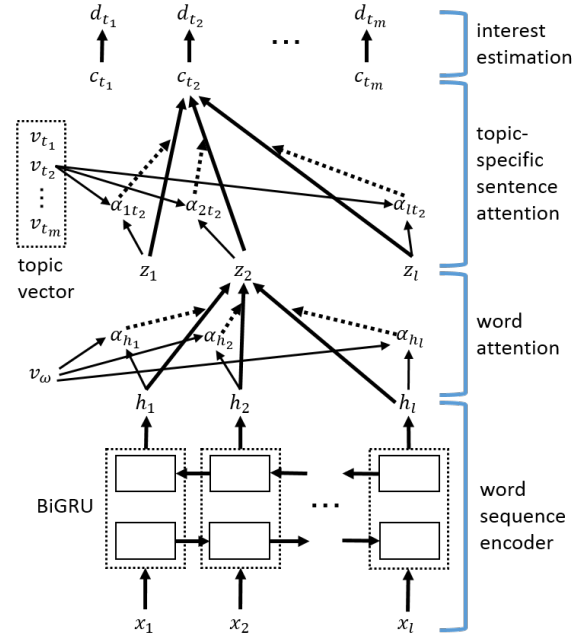


Figure 1: Overview of the proposed interest estimation model.

tered by a speaker  $s$  during dialogues with other speakers, we estimate their degrees of interest  $Y_s = (y_1, y_2, \dots, y_m)$  in topics in a given topic set  $T = (t_1, t_2, \dots, t_m)$ . Here, the  $t_i$  correspond to concrete topics, such as movies or travel while  $y_i$  indicates the speaker’s level of interest in  $t_i$ , on the three-point scale used for the LOI estimation task described in the previous section. Using this scale, the  $y_i$  can take the values 0 (disinterest, indifference and neutrality), 1 (light interest), or 2 (strong interest).

To accurately gauge the speaker’s interest from their utterances, we believe it is important to extract the following two types of information efficiently.

- The topic of each utterance
- How interested the speaker is in the topic

Our proposed interest estimation model extracts this information efficiently and uses a pre-training method to improve learning. Figure 1 presents an overview of our neural network model, which first encodes the word sequence, applies word attention and topic-specific sentence attention, and finally estimates the degrees of interest  $D_s = (d_{t_1}, d_{t_2}, \dots, d_{t_m})$ . The proposed pre-training method is used for the word sequence encoder. The model is described in detail below.

### 3.1 Word Sequence Encoder

The word sequence encoder converts utterances into fixed-length vectors using a recurrent neural network (RNN). First, the words in each utterance are converted into word vectors using Word2vec (Mikolov et al., 2013), giving word vector sequences  $x = (x_1, x_2, \dots, x_l)$ . The RNN encoder uses a hidden bidirectional-GRU (BiGRU) layer, which consists of a forward GRU that reads from  $x_1$  to  $x_l$  in order and a backward GRU that reads from  $x_l$  to  $x_1$  in reverse order. The forward GRU computes the forward hidden states  $\vec{h}_i$  as follows.

$$\vec{h}_i = \overrightarrow{GRU}(x_i, \vec{h}_{i-1}) \quad (1)$$

The backward GRU calculates the backward hidden states  $\overleftarrow{h}_i$  in a similar way. By combining the outputs of both GRUs, we obtain the objective hidden state  $h_i$ :

$$h_i = [\vec{h}_i : \overleftarrow{h}_i] \quad (2)$$

where  $[:]$  represents vector concatenation.

### 3.2 Topic Classification Pre-Training

Estimating the user’s level of interest in each topic requires first assessing the topic of each utterance. Since this is not given explicitly, the model must infer this information from the utterance set and degrees of interest in each topic, so the learning difficulty is high. In this study, based on the idea of pre-training (Erhan et al., 2010), we introduce a new pre-training method for the sentence topic classification task to the word sequence encoder. The important point to note about this task is that the topic classes involved are identical to those in the topic set  $Y_s$ . This helps to reduce the difficulty of learning to estimate the relationships between utterances and topics and allows the model to focus on interest estimation during the main training phase.

During pre-training, the classification probability  $p$  for each topic is calculated as follows, based on the output  $h_l$  of the BiGRU after inputting the last word vector  $x_l$ . (Word attention, as described in the next section, is not used in pre-training.)

$$p = \text{softmax}(W_c h_l + b_c) \quad (3)$$

where  $W_c$  and  $b_c$  are parameters for topic classification. The cross-entropy is used as the loss function during pre-training.

### 3.3 Word Attention

Based on an idea from Yang et al., we also included word attention in our model (Yang et al., 2016). Word attention is based on the idea that all words do not contribute equally to the desired result and involves using an attention mechanism to weight each word differently. The resulting utterance vector  $z$  is obtained as follows.

$$v_{hi} = \tanh(W_\omega h_i + b_\omega) \quad (4)$$

$$\alpha_{hi} = \frac{\exp(v_{hi}^\top v_\omega)}{\sum_i \exp(v_{hi}^\top v_\omega)} \quad (5)$$

$$z = \sum_i \alpha_{hi} h_i \quad (6)$$

where  $W_\omega$  and  $b_\omega$  are parameters. Unlike the original attention mechanism used in neural translation (Bahdanau et al., 2015) and neural dialogue (Shang et al., 2015) models, the word attention mechanism uses a common parameter, called context vector  $v_\omega$  to calculate weight  $\alpha_i$  for each hidden state.  $v_\omega$  is a high-level representation for calculating word importance and, like the model’s other parameters, is randomly initialized and then optimized.

### 3.4 Topic Specific Sentence Attention

Our model uses a word sequence encoder with word attention to convert the utterance set  $U_s = (u_1, u_2, \dots, u_n)$  into the utterance vector set  $Z_s = (z_1, z_2, \dots, z_n)$ . It then extracts information for estimating the level of interest in each topic from  $Z_s$ , but, as with word attention, not all utterances contribute equally. Yang et al. proposed a sentence attention mechanism that takes the same approach as for word attention, but, since it uses only one parameter to calculate sentence importance (similar to the context vector  $v_\omega$  for word attention), it is not capable of topic-specific estimation. This is because the important utterances in a given utterance set differ from topic to topic. For example, “I jog every morning” is probably useful for estimating interest in topics, such as sports or health, but not in, say, computers or vehicles.

In this study, we therefore propose a new topic-specific sentence attention approach. The topic vector  $v_{t_i}$  represents the importance of each sentence for topic  $t_i$ , and the associated content vector  $c_{t_i}$  is calculated as follows.

$$v_j = \tanh(W_r z_j + b_r) \quad (7)$$

$$\alpha_{jt_i} = \frac{\exp(v_j^T v_{t_i})}{\sum_j \exp(v_j^T v_{t_i})} \quad (8)$$

$$c_{t_i} = \sum_j \alpha_{jt_i} z_j \quad (9)$$

Here,  $W_r$  and  $b_r$  are shared, topic-independent parameters. The topic vector  $v_{t_i}$  is randomly initialized and then optimized during training.

### 3.5 Interest Estimation

We then use the content vector  $c_{jt_i}$  to compute the degree of interest  $d_{t_i}$  in topic  $t_i$  as follows.

$$d_{t_i} = \tanh(W_{t_i} c_{t_i} + b_{t_i}) + 1 \quad (10)$$

Here, the parameters  $W_{t_i}$  and  $b_{t_i}$  estimate the overall degree of interest in the topics  $t_i$ , and it is different for each topic after optimization. Finally, one is added, so that  $d_{t_i}$  uses the same 0 to 2 range as the correct values  $y_i$ .

During training, we use the mean squared error (MSE) between the correct answer  $y_i$  and  $d_{t_i}$  as the loss function:

$$L = \frac{1}{n} \sum_i (y_i - d_{t_i})^2 \quad (11)$$

## 4 Experiments

We conducted a series of experiments to evaluate the proposed method’s performance. For these, we created a dataset based on logs of one-to-one text chats between human subjects and the results of questionnaires answered by each subject. We also tested several baseline methods for comparison purposes.

### 4.1 Datasets

We asked each subject to first fill out a questionnaire about their interests and then engage in text chats in Japanese with partners they had not previously been acquainted with. We recruited 163 subjects via the CrowdWorks<sup>2</sup> crowd-sourcing site. The subjects were asked to rate their levels of interest in the 24 topic categories shown in Table 1 using a three-point scale discussed in Section 3. These topics were selected based on the categories used by Yahoo! Chiebukuro<sup>3</sup>, a Japanese question-and-answer site, focusing on topics that are likely to appear in one-to-one dialogues between strangers.

<sup>2</sup><https://crowdworks.jp/>

<sup>3</sup><https://chiebukuro.yahoo.co.jp/>

Table 1: Topic Categories

Travel	Movies	Celebrities
Music	Reading	Anime / Manga
Games	Computers	Home Appliances
Beauty	Fashion	Sports / Exercise
Health	School	Outdoor Activities
Housing	Housekeeping	Marriage / Love
Animals	Family	Cooking / Meal
Vehicles	History	Politics / Economy

Table 2: Example dialogue (translated by authors)

A	対話を開始します。よろしくお願ひします。 Let’s start a conversation. Nice to meet you.
B	はい、よろしくお願ひします。 Hi, nice to meet you.
A	何かご趣味はありますか？ What are your hobbies?
B	最近ハペット中心の生活になつて居るのでペットが趣味になりますね。 Currently, I am living a pet-centered lifestyle. So, raising pets is my hobby.
A	何を飼つてらっしゃるのですか？ Which pets do you have?
B	猫を飼つて居ます。3匹居るのでにぎやかですよ。 I have three cats and they are lively.
A	3匹ですか、いいですね！雑種ですか？ Three cats. That sounds great! Are they mixed breed?
B	はい、全部雑種です。手がかからなくて楽ですね。何か動物は飼つて居ますか？ Yes, they are all mixed breed cats. They are low-maintenance and easy to keep. Do you have any animals?

Each dialogue lasted for one hour and was conducted via Skype instant messaging. We only instructed the subjects to “Please try to find things you and your partner are both interested in and then try to broaden your conversation about these subjects.” We gave the subjects no specific instructions as to the intended content or topics of their conversations. Table 2 shows an example dialogue between subject A and B.

All the utterances in the chat data were then classified by subject. Each data point consisted of all the data about one subject, namely their chat ut-

Table 3: Data Statistics

Number of users (data points)	163
Number of dialogues	408
Number of utterances	49029
Avg. number of strong interest topics	11.48
Avg. number of light interest topics	7.30
Avg. number of neutral topics	5.21

terances and questionnaire responses (correspond to  $U_s$  and  $Y_s$  defined in Section 3). The data was evaluated using 10-fold cross-validation and their statistics are shown in Table 3.

## 4.2 Settings

Word2Vec (Mikolov et al., 2013) was trained using 100 GB of Twitter data with 200 embedding cells, a minimum word frequency of 10, and a skip-gram window size of 5.

The word sequence encoder was a single-layer BiGRU RNN with 200 input cells and 400 output cells. The word and sentence attention layers had 400 input and output cells while the estimation layer had 400 input cells and 1 output cell. The model was trained using Adam (Kingma and Ba, 2015).

During pre-training, we used questions and answers from the Yahoo! Chiebukuro Data (2nd edition)<sup>4</sup> for each topic. All topics were equally covered, and a total of 770k sentences were used for training while 2400 sentences (100 for each topic) were used for testing. After pre-training, the topic classification accuracy for the test data was 0.755.

## 4.3 Evaluation

When using the proposed method as part of a dialogue system, it is effective to select the best topic from those available for the system to generate an appropriate response. Therefore, in this experiment, each topic was ranked based on the estimated degree of interest  $d_{t_i}$ , and the methods were evaluated based on whether the topics the user was interested in should have been ranked higher or the other topics ranked lower. The rankings were evaluated using the normalized discounted cumulative gain ( $NDCG$ ), a widely used metric in the field of information retrieval. This gives values between 0 and 1, with higher values indicating more accurate

<sup>4</sup>[http://www.nii.ac.jp/dsc/idr/yahoo/chiebkr2/Y\\_chiebukuro.html](http://www.nii.ac.jp/dsc/idr/yahoo/chiebkr2/Y_chiebukuro.html)

ranking predictions and is calculated as follows.

$$NDCG@k = \frac{DCG_k}{IDCG_k} \quad (12)$$

$$DCG_k = rel_i + \sum_{i=2}^K \frac{rel_i}{\log_2 i} \quad (13)$$

Here,  $k$  is the number of top-ranked objects used for the  $NDCG$  calculation, and  $rel_i$  is the graded relevance of the result at position  $i$ , which was given by the degree of interest  $Y_s$  in this experiment. The ideal  $DCG$  ( $IDCG$ ) is the  $DCG$  if the ranking list had been correctly ordered by relevance.

In addition, to evaluate the accuracy of the estimated degrees of interest in each topic, we also calculated the MSEs between the results of each method and the correct answers.

## 4.4 Baseline Methods

To evaluate the proposed model, we also conducted experiments using the following three modified models.

### Without Pre-Training

To evaluate the effectiveness of topic classification pre-training, we tested our model without this step. Instead, the word sequence encoder was randomly initialized and then trained. This model was otherwise identical to the proposed method.

### Without Sentence Attention

To evaluate the effectiveness of topic-specific sentence attention, we tried instead using max-pooling to obtain the content vector. Again, this model was identical to the proposed method.

### Without Pre-Training or Sentence Attention

This model combined the two modifications mentioned above: it did not use topic classification pre-training and used max-pooling to obtain the content vectors, but was otherwise identical to the proposed method.

We also compared our model’s performance to those of the following two baseline methods.

### Topic Frequency

The first baseline was based on a method, proposed by Abel et al., that identifies the named entities (such as people, events, or

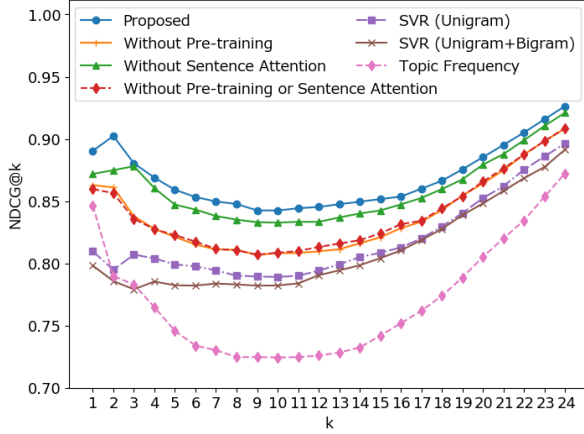


Figure 2: NDCG@k results for all methods, for k between 1 and 24.

Table 4: Mean Squared Error

Proposed	<b>0.533</b>
Without Pre-Training	0.580
Without Sentence Attention	0.561
Without Pre-Training or Sentence Attention	0.568
SVR (unigram)	0.597
SVR (unigram + bigram)	0.611

music groups) associated with words in the user’s tweets using OpenCalais and models the user’s interests using a named entity frequency vector (Abel et al., 2011). However, as we used Japanese dialogues, we could not use OpenCalais, so we instead used the topic classifier described in Section 3.2. Since this classifier is trained for classification for sentences and not for words, we employed sentence level topic frequency. The topic frequency was used to gauge the user’s interest, and the topics were ranked in frequency order.

## SVR

The second baseline method used support vector regression (SVR) to estimate the degree of interest. We conducted experiments using only unigrams, and using both unigrams and bigrams. We used the RBF kernel function. The SVR models were trained for each topic individually and then used to estimate the degrees of interest.

## 4.5 Results

Figure 2 shows the  $NDCG$  results for the topics ranked in the top  $k$ . These indicate that the proposed method performed better than the other methods for all values of  $k$ . Comparing the performances of the methods that used pre-training (“Proposed” and “Without Sentence Attention”) with those of the ones that did not (“Without Pre-Training” and “Without Pre-Training or Sentence Attention”) indicates that the proposed pre-training step was effective. On the other hand, a method that used sentence attention (“Without Pre-Training”) showed nearly the same results as one that did not (“Without Pre-Training or Sentence Attention”), although the latter did achieve higher  $NDCG$ s for  $k \geq 5$ . This indicates that using sentence attention alone does not improve performance. However, the proposed method performed better than the method without sentence attention, confirming that sentence attention is useful, but only if it is used in conjunction with pre-training.

Turning now to the SVR-based methods, we observe that using only unigram features worked better than using both unigrams and bigrams, although both methods were still inferior to the neural network-based methods, including the proposed method.

When  $k = 1$ , the topic frequency baseline achieved higher  $NDCG$ s than the SVR-based methods, because it correctly noted that users were strongly interested in the topics they spoke about most frequently. However, these results were still inferior to those of the neural network-based methods. Furthermore, it presented the worst  $NDCG$  results among all the methods for  $k \geq 4$ , due to speakers sometimes talking about subjects they were not interested in, as discussed in Section 2.

Table 4 shows the MSEs between the degree of interest results for each method and the correct answers (excluding Topic Frequency, which cannot output the degree of interest). The proposed method gave significantly smaller MSE value, indicating that its estimates were the most accurate. In addition, the “Without Pre-Training” method showed the lowest performance of all the neural network-based methods, also indicating that the proposed sentence attention approach is not effective without also using pre-training.

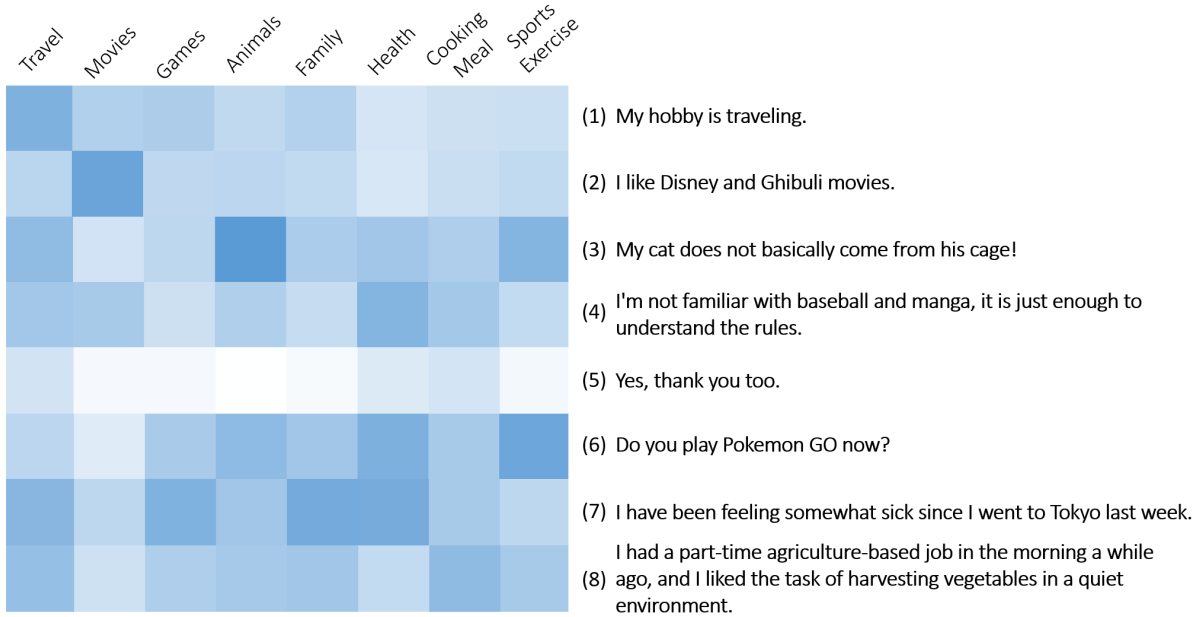


Figure 3: Visualization of Attention (Proposed)

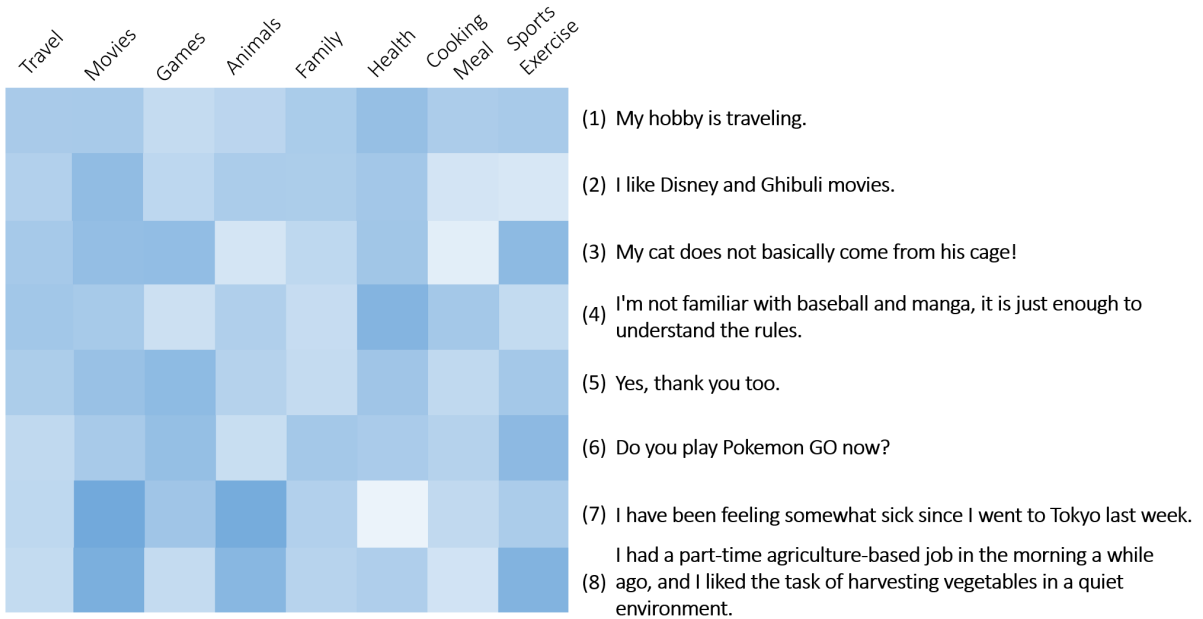


Figure 4: Visualization of Attention (Without Pre-Training)

#### 4.6 Discussion

The experimental results discussed in the previous section indicate that it is important to use the proposed pre-training and sentence attention steps together. To analyze the sentence attention mechanism further, we visualized the sentence weights  $\alpha_{jt_i}$  given by equation (8) for selected topics and utterances. Figures 3 and 4 show the sentence

weights with and without pre-training, respectively. Here, darker cells indicate higher  $\alpha_{jt_i}$  values.

Figure 3 shows that the sentence weights for the topics corresponding to the actual meaning of the sentence are high. (1), (2) and (3) are easy-to-understand examples. The topics related to each utterance take the highest weights. In addition,

utterance (4) includes sports-related words, such as “baseball” and “rule”, but the weight of the “Sports / Exercise” topic is not high because the utterance did not indicate such an interest on the part of the speaker. Thus, the sentence weights do not simply reflect the topics of the words, but also the user’s level of interest in the topic. Interestingly, although the utterance (6) refers to the smartphone game “Pokemon GO”, the weight of the “Game” topic is not very high, but those of the “Sports/Exercise” and “Health” topics are both high. Pokemon GO is interesting to people who do not usually play games, and this appears to be reflected in the results. On the other hand, utterance (7) shows high weights for several topics that intuitively appear to be unrelated to the utterance itself.

The sentence weights shown in Figure 4 often do not correspond to the topics or meanings of the utterances. For example, utterance (5) is not important for interest estimation and its weights in Figure 3 are small. However, in Figure 4, all weights are relatively high. Similarly, utterances (7) and (8) show high weights for unrelated topics.

The above results confirm that the pre-training step is important for learning the topic-specific sentence attention correctly. Without pre-training, the model must learn the relationships between utterances and topics by starting from a clean slate, and the difficulty of this task makes harder to determine the appropriate results. The experimental results in the previous section show that pre-training makes this task easier and improves performance. With proper training, topic-specific sentence attention then enabled the proposed method to achieve the best performance.

## 5 Conclusion

In this paper, we have presented a neural network-based method for estimating users’ levels of interest in a pre-determined list of topics based on their utterances in chat dialogues. The proposed method first encodes utterances by using BiGRU and considering word attention, a set of utterance vectors was obtained. It then uses these to generate content vectors corresponding to each topic via topic-specific sentence attention. Finally, it uses the content vectors to estimate the user’s degree of interest in each topic. The utterance encoder is pre-trained to classify sentences by topic before

the whole model is trained. Our experimental results showed that the proposed method can estimate degrees of interest in topics more accurately than baseline methods. In addition, we found that it was most effecting to use topic-specific sentence attention and topic classification pre-training in combination.

In future work, we plan to apply the proposed method to a dialogue system and conduct dialogue experiments with human users. Even if we can estimate which topics a user is interested in, generating and selecting concrete utterances remains a challenging problem. For example, users who are interested in sports are not equally interested in all of them: someone may be interested in football but not in golf, for instance. We therefore plan to develop an appropriate way of incorporating the proposed method into such a system.

## Acknowledgements

The Yahoo! Chiebukuro Data (2nd edition) provided to National Institute of Informatics by Yahoo Japan Corporation was used in this study.

This study received a grant of JSPS Grants-in-aid for Scientific Research 16H05880.

## References

- Fabian Abel, Qi Gao, Geert-Jan Houben, and Ke Tao. 2011. Analyzing user modeling on twitter for personalized news recommendations. *User Modeling, Adaption and Personalization* pages 1–12.
- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. *Proc. ICLR*.
- Jeesoo Bang, Hyungjong Noh, Yonghee Kim, and Gary Geunbae Lee. 2015. Example-based chat-oriented dialogue system with personalized long-term memory. In *IEEE International Conference on Big Data and Smart Computing (BigComp)*. pages 238–243.
- Parantapa Bhattacharya, Muhammad Bilal Zafar, Niloy Ganguly, Saptarshi Ghosh, and Krishna P Gummadi. 2014. Inferring user interests in the twitter social network. In *Proceedings of the 8th ACM Conference on Recommender systems*. ACM, pages 357–360.
- Jilin Chen, Rowan Nairn, Les Nelson, Michael Bernstein, and Ed Chi. 2010. Short and tweet: experiments on recommending content from information streams. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, pages 1185–1194.



- Dumitru Erhan, Yoshua Bengio, Aaron Courville, Pierre-Antoine Manzagol, Pascal Vincent, and Samy Bengio. 2010. Why does unsupervised pre-training help deep learning? *Journal of Machine Learning Research* 11(Feb):625–660.
- Jonghyun Han and Hyunju Lee. 2016. Characterizing the interests of social media users: Refinement of a topic model for incorporating heterogeneous media. *Information Sciences* 358:112–128.
- Ryuichiro Higashinaka, Kenji Imamura, Toyomi Meguro, Chiaki Miyazaki, Nozomi Kobayashi, Hiroaki Sugiyama, Toru Hirano, Toshiro Makino, and Yoshihiro Matsuo. 2014. Towards an open-domain conversational system fully based on natural language processing. In *COLING*. pages 928–939.
- Toru Hirano, Nozomi Kobayashi, Ryuichiro Higashinaka, Toshiro Makino, and Yoshihiro Matsuo. 2015. User information extraction for personalized dialogue systems. *The 19th Workshop on the Semantics and Pragmatics of Dialogue (SemDial)* pages 67–75.
- Pavan Kapanipathi, Prateek Jain, Chitra Venkataramani, and Amit Sheth. 2014. User interests identification on twitter using a hierarchical knowledge base. In *European Semantic Web Conference*. Springer, pages 99–113.
- Pavan Kapanipathi, Fabrizio Orlandi, Amit P Sheth, and Alexandre Passant. 2011. Personalized filtering of the twitter stream pages 6–13.
- Diederik P Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. *Proceedings of the 3rd International Conference on Learning Representations*.
- Hanae Koiso, Tomoyuki Tsuchiya, Ryoko Watanabe, Daisuke Yokomori, Masao Aizawa, and Yasuharu Den. 2016. Survey of conversational behavior: Towards the design of a balanced corpus of everyday japanese conversation. In *LREC 2016*.
- Matthew Michelson and Sofus A Macskassy. 2010. Discovering users’ topics of interest on twitter: a first look. In *Proceedings of the fourth workshop on Analytics for noisy unstructured text data*. ACM, pages 73–80.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. pages 3111–3119.
- Alan Ritter, Colin Cherry, and William B Dolan. 2011. Data-driven response generation in social media. In *Proceedings of the conference on empirical methods in natural language processing*. pages 583–593.
- Björn W Schuller, Niels Köhler, Ronald Müller, and Gerhard Rigoll. 2006. Recognition of interest in human conversational speech. In *INTERSPEECH*.
- Lifeng Shang, Zhengdong Lu, and Hang Li. 2015. Neural Responding Machine for Short Text Conversation. *Proceedings of the 53th Annual Meeting of Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing* pages 1577–1586.
- Alessandro Sordani, Michel Galley, Michael Auli, Chris Brockett, Yangfeng Ji, Margaret Mitchell, Jian-Yun Nie, Jianfeng Gao, and Bill Dolan. 2015. A neural network approach to context-sensitive generation of conversational responses. In *Proceedings of NAACL-HLT*.
- Oriol Vinyals and Quoc Le. 2015. A neural conversational model. In *Proceedings of the ICML Deep Learning Workshop*. pages 1–7.
- Jianshu Weng, Ee-Peng Lim, Jing Jiang, and Qi He. 2010. Twitterrank: finding topic-sensitive influential twitterers. In *Proceedings of the third ACM international conference on Web search and data mining*. ACM, pages 261–270.
- Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alexander J Smola, and Eduard H Hovy. 2016. Hierarchical attention networks for document classification. In *HLT-NAACL*. pages 1480–1489.
- Fattane Zarrinkalam, Hossein Fani, Ebrahim Bagheri, Mohsen Kahani, and Weichang Du. 2015. Semantics-enabled user interest detection from twitter. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2015 IEEE/WIC/ACM International Conference on*. IEEE, volume 1, pages 469–476.
- Chen Zhang and Joyce Y Chai. 2009. What do we know about conversation participants: Experiments on conversation entailment. In *Proceedings of the SIGDIAL 2009 Conference*. pages 206–215.
- Chen Zhang and Joyce Y Chai. 2010. Towards conversation entailment: An empirical investigation. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*. pages 756–766.
- Tiancheng Zhao, Ran Zhao, and Maxine Eskenazi. 2017. Learning discourse-level diversity for neural dialog models using conditional variational autoencoders. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, pages 654–664.