# Modeling Dyadic Conversations for Personality Inference

Qiang Liu[1]*, Fuzheng Zhang[2], Xing Xie[2], Shu Wu[1], Liang Wang[1]
[1]Institute of Automation, Chinese Academy of Sciences
[2]Microsoft Research
{qiang.liu, shu.wu, wangliang}@nlpr.ia.ac.cn
{fuzzhang, xing.xie}@microsoft.com

## ABSTRACT

Nowadays, automatical personality inference is drawing extensive attention from both academia and industry. Conventional methods are mainly based on user generated contents, e.g., profiles, likes, and texts of an individual, on social media, which are actually not very reliable. In contrast, dyadic conversations between individuals can not only capture how one expresses oneself, but also reflect how one reacts to different situations. Rich contextual information in dyadic conversation can explain an individual's response during his or her conversation. In this paper, we propose a novel model for learning unsupervised Personal Conversational Embeddings (PCE) based on dyadic conversations between individuals. We adjust the formulation of each layer of a conventional Gated Recurrent Unit (GRU) with sequence to sequence learning and personal styles of both sides of the conversation. Based on the learned PCE, we can infer the personality of each individual. We conduct experiments on two datasets: the Movie Script dataset and the XiaoIce datatset, which are collected from conversations between characters in movie scripts and volunteer users' conversation records on the XiaoIce chatbot respectively. We find that modeling dyadic conversations between individuals can significantly improve personality inference accuracy. Experimental results on two datasets illustrate the successful performance of our proposed method.

## Keywords

Personality inference, dyadic conversations, personal conversational embeddings

## 1. INTRODUCTION

Research in psychology has suggested that behaviors and preferences of individuals can be explained to a great extent by underlying psychological constructs: personality traits [9, 28]. Understanding an individual's personality can give us hints to make predictions about his or her preferences across different contexts and environments, which can be applied to enhance recommender systems [21] and advertisement targeting [6]. Some research shows that personality can significant affect users' choices of social relationships [1], products [20], entertainment [5] and websites [19]. The most widely-used measurement of personality is the Big Five model, which contains five personality traits: Extraversion, Agreeableness, Conscientiousness, Neuroticism and Openness [17]. Detailed explanations of these five personality traits can be found in Table 1.

Traditional methods for personality inference require individuals to answer hundreds of questions formulated based on psychological research [17]. This requires the collaboration of individuals and is very time consuming, which makes such methods hard to be applied to a large number of users on the web. Recently, extensive research has shown that user generated contents on the web can be utilized to automatically infer personality [9]. Some works [20, 39] make prediction of user personality based on their likes on Facebook, and show this technology to be practical. Some works [29, 33] infer personality according to textual contents posted on social media, and show language usage is a key factor for inferring personality. There are also some works [24] that models pictures on social media for personality inference.

However, user generated contents are not reliable enough for personality inference. Because user generated contents can only capture how you express yourself, which is usually in a way that you want others know about you. For example, users tend to show their good and friendly sides on social media. In contrast, conversation records between individuals can not only capture how you express yourself, but also reflect how you react to different situations and individuals. This is usually in a more natural way that you can not help but to say something in certain situations. Someone showing the friendly side on social media may be unfriendly in the daily life and conversations. And someone that is not active on social media may also prefer to talk a lot with his or her friends and parents. Contextual information in conversation records can provide reasons and explanations for an individual's response. For example, a good tempered person would be angry when someone says something rude to him or her. But this does not make this person bad tempered. Meanwhile, there are a growing number of applications associated with dyadic conversation data, such as online chatbots, e.g., Microsoft XiaoIce chatbot[1], replying and commenting on

*This work was done while the author was a research intern at Microsoft Research.



*WOODSTOCK '97 El Paso, Texas USA*



DOI: 10.475/123_4

[1]http://www.msxiaoice.com/

**Table 1: Explanations of the Big Five personality model.**

| Personality | Explanation |
|---|---|
| Extraversion | A tendency to seek stimulation in the company of others. |
| Agreeableness | A tendency to be compassionate and cooperative towards others. |
| Conscientiousness | A tendency to act in an organized or spontaneous way. |
| Neuroticism | A tendency to have sensitive emotions to the environment. |
| Openness | A tendency to be open to experiencing a variety of activities. |

social media, e.g., Twitter[2] and Weibo[3], and email replying [18]. Accordingly, it is vital to model dyadic conversations between individuals for personality inference. On the other hand, these applications can benefit from personality inference in modeling users and promoting their services.

In this work, we learn unsupervised **Personal Conversational Embeddings (PCE)** based on dyadic conversations between individuals and utilize the learned PCE of each individual for personality inference. Recently, Recurrent Neural Networks (RNN) have been widely applied and achieved successful performance in modeling sentences. Considering the vanishing or exploding gradients problem [3] in conventional RNN architecture, and the high computational cost in the Long Short-Term Memory (LSTM) unit [15], we select the Gated Recurrent Unit (GRU) [7] for modeling sentences in dyadic conversations. Based on GRU, to model personality of individuals, we need to take the following three important contextual factors in a dyadic conversation into consideration: "what you are replying to", "who you are" and "who you are talking to". "What you are replying to" is the previous message of a replying sentence, indicating the cause of the corresponding response. "Who you are" and "who you are talking to" reflects personal styles of both sides of the conversation, indicating how an individual tends to talk and be replied to. Thus, we propose an augmented GRU model, in which PCE can be learned, via sequence to sequence learning and adjusting model formulation with personal styles. With the learned PCE of each individual, we can make personality inferences based on two-layer fully-connected neural networks.

To the best of our knowledge, we are the first to predict user personality in a conversational scenario. The main contributions of this work are listed as follows:

- We address the challenge of personality inference with dyadic conversations. This can utilize various contextual information in conversation records for better inferring personality of individuals.
- We propose an augmented GRU model for learning unsupervised personal conversational embeddings based on dyadic conversation. Besides modeling sentences with GRU, our method can model several important contextual factors in dyadic conversation.

[2]https://twitter.com/
[3]http://weibo.com

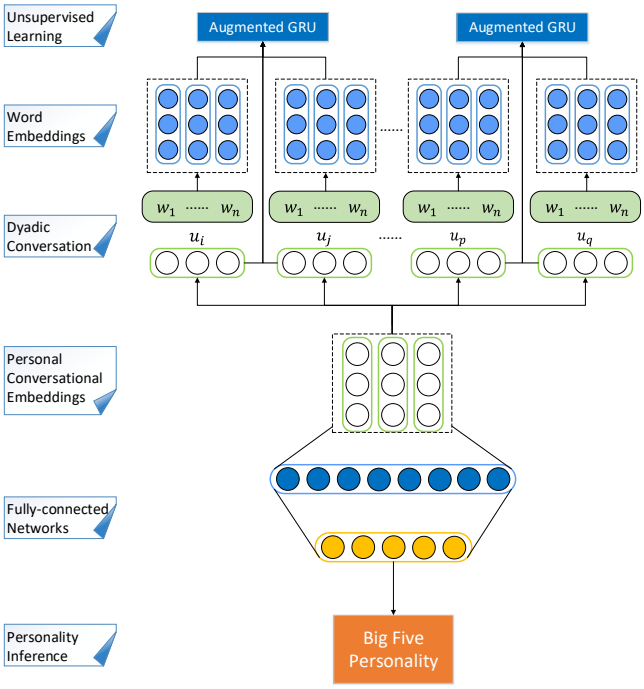



**Figure 1: Framework of learning Personal Conversational Embeddings (PCE) and inferring the big five personality based on dyadic conversations between individuals.**

- We conduct experiments on two datasets, namely the Movie Script dataset and the XiaoIce dataset. We find modeling dyadic conversations can significantly improve the accuracy of personality inference, and experimental results show the strong performance of our method.

The rest of the paper is organized as follows. In section 2, we review some related works on personality inference and recurrent neural networks. Section 3 details our method for learning personal conversational embeddings based on dyadic conversation. In section 4, we report and analyze experimental results. Section 5 concludes our work and discusses future research.

## 2. RELATED WORK

In this section, we briefly review some related work on personality inference and recurrent neural networks.

### 2.1 Personality Inference

Nowadays, social media provides a great amount of personal information on users, e.g., social relationships, tweets, blogs and uploaded photos. The personality of an individual can be extracted from these user generated contents [13], and extensive research has been done on automatic personality inference [9]. Facebook[4] has collected data on users and started a project called myPersonality[5], which has attracted significant attention from researchers [2, 10, 14, 20,

[4]http://www.facebook.com
[5]http://mypersonality.org

33, 39]. Different types of features have been investigated on Facebook data, such as user profiles [14], using patterns [2], likes [20, 39] and textual contents [10, 33]. Bachrach et al. [2] analyze the correlation between Facebook using patterns and users' personality. Wu et al. [39] apply matrix factorization on user-item matrix for personality inference, and conclude that computer-based automatic personality inference is more accurate than those made by humans. Schwartz et al. [33] infer users' personality, gender, and age by applying topic model on user posted contents. Data on Twitter has also been utilized for analyzing personality [12, 16, 24, 30]. Hughes et al. [16] investigate the difference between Facebook usage and Twitter usage for analyzing personality. Quercia et al. [30] predict personality based user profiles, such as followings, followers, and listed counts. Liu et al. [24] predict personality via extracting features from users' uploaded photos. Moreover, avatars and emojis of users have also been investigated for personality inference [38].

## 2.2 Recurrent Neural Networks

Recently, RNN has been widely applied for various sentence modeling tasks. Back-Propagation Through Time (BPTT) [32] is usually used for optimization of RNN models. However, in practice, basic RNN structure will face the vanishing or exploding gradients problem when learning long-term temporal dependency [3]. To overcome this drawback, researchers extend RNN with memory units, and propose some advanced structures, such as LSTM [15] and GRU [7, 8]. According to existing research [7], LSTM and GRU share similar performances on various tasks, but GRU has much lower computational cost.

Recently, RNN, including LSTM and GRU, has been widely applied for generating conversation and response [23, 34, 35, 36, 37]. Basically, this is a sequence-to-sequence problem. Vinyals et al. [37] build the first neural conversational model based on RNN. Shang et al. [35] incorporate an encoder-decoder framework in GRU for generating responses. Serban et al. [34] propose a hierarchical model for conversation machines. Sordoni et al. [36] utilize previous paragraphs as contexts, and propose a context-sensitive response generating model. Li et al. [23] adjust LSTM formulation with personal information, such as address and birthday, to better generate personalized conversational response.

Variety of contextual information has also been incorporated in RNN models for different specific tasks. Ghosh et al. [11] apply topic model and treat surrounding texts of a specific sentence as contexts. Then each layer of LSTM is adjusted with contexts, and the Contextual LSTM (CLSTM) model is proposed. CLSTM achieves state-of-the-art performance in next sentence selection and sentence topic prediction. Visual features have also been incorporated in LSTM for multimodal applications, such as image caption [27] and visual question answering [31]. Moreover, RNN has been jointly modeled with behavioral contexts for user modeling, and achieves state-of-the-art performances in recommender systems [25, 26].

# 3. LEARNING PERSONAL EMBEDDINGS

In this section, we present our method for learning unsupervised personal conversational embeddings based on dyadic conversations. First, we formulate the notations of this paper. Then, we introduce conventional RNN and GRU structures. Finally, we discuss how to involve several contextual factors, i.e., "who you are", "what you are replying to" and "who you are talking to", in dyadic conversation into an augmented GRU model.

## 3.1 Notations

In this work, we have a set of individuals denoted as $U$=$\{u_1$,$u_2$,...$\}$, and conversation records between them. Each pair of dyadic conversation between two individuals consists of a message (the sentence before the corresponding response) and a response (the replying sentence) denoted as $M$=$\{BOS$,$w_1$,$w_2$,...,$EOS\}$ and $R$=$\{BOS$,$w_1$,$w_2$,...,$EOS\}$ respectively. $BOS$ and $EOS$ are begin and end of the word sequence respectively, and $w_t$ means one word in a sentence (message or response). Sentences in the conversation are associated with personal identification of both sides of the conversation. Given conversation records of individuals, we plan to infer each individual's big five personality: extraversion, agreeableness, conscientiousness, neuroticism, and openness.

## 3.2 Recurrent Neural Networks

The architecture of RNN is a recurrent structure with multiple hidden layers. At each time step, we can predict the output unit given the hidden layer, and then feed the new output back into the next hidden status. It has been successfully applied in a variety of applications [11, 25, 27, 37]. The formulation of each hidden layer in RNN is:

$$\mathbf{h}_t = tanh(\mathbf{w}_t\mathbf{M} + \mathbf{h}_{t-1}\mathbf{N}) \ , \tag{1}$$

where $\mathbf{w}_t \in \mathbb{R}^d$ is the word embedding, $\mathbf{h}_t \in \mathbb{R}^d$ is the hidden state of RNN, and $\mathbf{M} \in \mathbb{R}^{d\times d}$ and $\mathbf{N} \in \mathbb{R}^{d\times d}$ are transition matrices in RNN.

## 3.3 Gated Recurrent Units

Due to the vanishing or exploding gradients problem [3], it is hard for conventional RNN to learn long-term dependency in sequences. Accordingly, memory units are proposed and applied, e.g., LSTM [15] and GRU [7, 8]. LSTM and GRU share similar performances in various tasks, but GRU is much faster [7]. So, GRU is becoming a better method for learning embeddings in sentences. The formulation of each layer of GRU is:

$$\begin{aligned} \mathbf{z}_t &= \sigma\left(\mathbf{w}_t\mathbf{M}_z + \mathbf{h}_{t-1}\mathbf{N}_z\right) \\ \mathbf{r}_t &= \sigma\left(\mathbf{w}_t\mathbf{M}_r + \mathbf{h}_{t-1}\mathbf{N}_r\right) \\ \tilde{\mathbf{h}}_t &= tanh\left(\mathbf{w}_t\mathbf{M}_h + \left(\mathbf{h}_{t-1}\cdot\mathbf{r}_t\right)\mathbf{N}_h\right) \\ \mathbf{h}_t &= (1-\mathbf{z}_t)\cdot\mathbf{h}_{t-1} + \mathbf{z}_t\cdot\tilde{\mathbf{h}}_t \end{aligned} \ , \tag{2}$$

where $\mathbf{w}_t \in \mathbb{R}^d$ is the embedding of a word in a sentence, and $\mathbf{h}_t \in \mathbb{R}^d$ is the hidden state of GRU, $\mathbf{M}_*$ and $\mathbf{N}_*$ are all $d\times d$ dimensional matrices. $\mathbf{z}_t$ is a reset gate, determining how to combine the new input with the previous memory. $\mathbf{r}_t$ is an update gate, defining how much of the previous memory is cascaded into the current state. $\tilde{\mathbf{h}}_t$ denotes the candidate activation of the hidden state $\mathbf{h}_t$.

## 3.4 Personal Speaking Embeddings

Though GRU has succeeded in modeling sentences, it can not be directly applied for personality analysis. GRU is not able to involve personal styles of the current speaker into consideration. However, personal style, including word preferences and language usage patterns, is vital for personality inference. Research in conversation machine has done some investigation on modeling personal information [23].

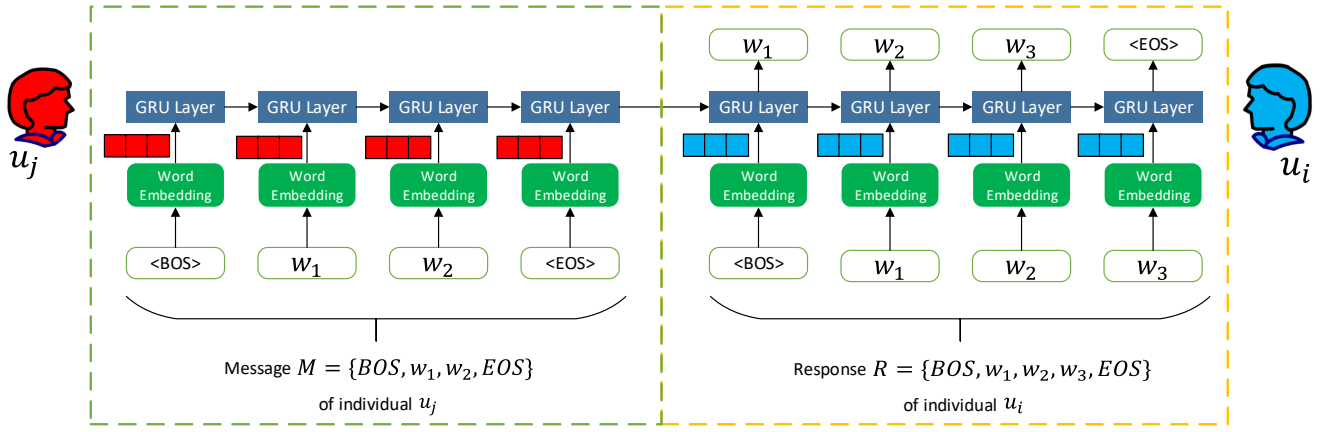



**Figure 2: The sequence to sequence learning process of dyadic conversation between $u_i$ and $u_j$.**

Thus, we incorporate personal styles with the conventional GRU model, and learn **Personal Speaking Embeddings (PSE)** based on sentences generated by individuals. The learning procedure is unsupervised along with the structure of GRU via predicting the next word, which makes our method flexible and does not require much annotated data. The PSE models "who you are", capturing the speaker's language usage preferences reflected in sentences in conversations. Mathematically, incorporating the personal style of the current speaking individual $u_i$, the formulation of each layer of GRU becomes:

$$\begin{aligned}
\mathbf{z}_t &= \sigma\left(\mathbf{w}_t\mathbf{M}_z + \mathbf{h}_{t-1}\mathbf{N}_z + \mathbf{u}_i\mathbf{P}_z\right) \\
\mathbf{r}_t &= \sigma\left(\mathbf{w}_t\mathbf{M}_r + \mathbf{h}_{t-1}\mathbf{N}_r + \mathbf{u}_i\mathbf{P}_r\right) \\
\tilde{\mathbf{h}}_t &= tanh\left(\mathbf{w}_t\mathbf{M}_h + \left(\mathbf{h}_{t-1}\cdot\mathbf{r}_t\right)\mathbf{N}_h + \mathbf{u}_i\mathbf{P}_h\right) \\
\mathbf{h}_t &= \left(1-\mathbf{z}_t\right)\cdot\mathbf{h}_{t-1} + \mathbf{z}_t\cdot\tilde{\mathbf{h}}_t
\end{aligned}, \quad (3)$$

where $\mathbf{u}_i \in \mathbb{R}^d$ is the embedding of current speaking individual, which can be learned unsupervisedly in the GRU structure. Formulation of the reset gate, the update gate, and the hidden state candidate activation are adjusted with personal style via the calculation $\mathbf{u}_i\mathbf{P}_z$, $\mathbf{u}_i\mathbf{P}_r$ and $\mathbf{u}_i\mathbf{P}_h$ respectively, where $\mathbf{P}_*$ are all $d \times d$ dimensional matrices. The learned embedding $\mathbf{u}_i$ captures how the individual prefers to speak and express himself or herself. This process can compress all the sentences in the conversation generated by an individual into a fixed dimensional latent representation. With the compressed representation of each individual, personality analysis and inference can be performed.

## 3.5 Personal Replying Embeddings

There is another important contextual factor in dyadic conversations: the message before the response, i.e., the sentence being replied to. This refers to "what you are replying to", capturing the cause and reason of the replying sentence. It is vital for analyzing personality. For example, if someone says something impolite to you, you may also say some rude words. But this does not mean you are a rude person. Thus, we incorporate the message before the response in a conversation into our model via sequence to sequence learning, and learn **Personal Replying Embeddings (PRE)**. The sequence to sequence learning is a widely-used structure for modeling message and response in question answering and conversation machine [23, 34, 35, 36, 37].

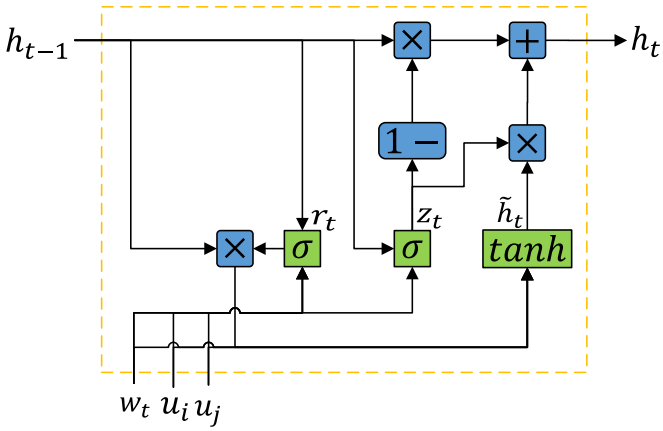



**Figure 3: Structure of the augmented GRU model for learning Personal Conversational Embeddings (PCE) incorporating the personal styles of both sides of the dyadic conversation.**

In our model, sentences are arranged in the order of conversations, and each pair of message $M$ and response $R$ are feeded into the model. Figure 2 illustrates the sequence to sequence learning process of the dyadic conversation between individual $u_i$ and individual $u_j$. The model is usually learned with an encoder-decoder method. For message $M$=$\{BOS$,$w_1$, $w_2$,...,$EOS\}$ generated by individual $u_j$, the sentence is encoded into a sentence embedding with the GRU model and fed to the following decoder as an input vector. Then, for response $R$=$\{BOS$,$w_1$,$w_2$,...,$EOS\}$ generated by individual $u_i$, the sentence is modeled and decoded according to the formulation in Equation 3. The learned PRE takes causes of replying sentences into consideration, reducing noise in data and being more flexible. To the best of our knowledge, no existing personality inference methods can model messages and responses in conversational data.

## 3.6 Personal Conversational Embeddings

Furthermore, the personal style of the individual being

Table 2: Details of the Movie Script dataset and the XiaoIce dataset.

| dataset | #individuals | #sentences | language | description |
|---|---|---|---|---|
| Movie Script | 880 | 180k | English | conversations between characters in movie scripts |
| XiaoIce | 660 | 850k | Chinese | conversations between users and XiaoIce on the chatbot |

replied to is also useful for analyzing personality. It tells us "who you are talking to", and captures how an individual tends to be replied to. For example, a friendly person usually receives polite treatment, while a rude person usually receives unfriendly words. Accordingly, following the sequence to sequence learning in PRE, we incorporate the personal style of the individual being replied to in our model, and learn **Personal Conversational Embeddings (PCE)** based on all the contextual information in dyadic conversation. For the conversation between $u_i$ and $u_j$, the formulation in Equation 3 can be adjusted as:

$$\begin{aligned}\mathbf{z}_t &= \sigma\left(\mathbf{w}_t\mathbf{M}_z + \mathbf{h}_{t-1}\mathbf{N}_z + \mathbf{u}_i\mathbf{P}_z + \mathbf{u}_j\mathbf{Q}_z\right)\\ \mathbf{r}_t &= \sigma\left(\mathbf{w}_t\mathbf{M}_r + \mathbf{h}_{t-1}\mathbf{N}_r + \mathbf{u}_i\mathbf{P}_r + \mathbf{u}_j\mathbf{Q}_r\right)\\ \tilde{\mathbf{h}}_t &= tanh\left(\mathbf{w}_t\mathbf{M}_h + \left(\mathbf{h}_{t-1}\cdot\mathbf{r}_t\right)\mathbf{N}_h + \mathbf{u}_i\mathbf{P}_h + \mathbf{u}_j\mathbf{Q}_h\right)\\ \mathbf{h}_t &= \left(1-\mathbf{z}_t\right)\cdot\mathbf{h}_{t-1} + \mathbf{z}_t\cdot\tilde{\mathbf{h}}_t\end{aligned} \quad , \quad (4)$$

where $\mathbf{u}_j \in \mathbb{R}^d$ is the embedding of the individual being replied to. To be noted that, embeddings of both sides of dyadic conversations share same parameters. Formulation of the reset gate, the update gate and the hidden state candidate activation can be adjusted via the calculation $\mathbf{u}_j\mathbf{Q}_z$, $\mathbf{u}_j\mathbf{Q}_r$ and $\mathbf{u}_j\mathbf{Q}_h$ respectively, where $\mathbf{Q}_*$ are all $d \times d$ dimensional matrices. The learned PCE captures how an individual tends to speak and be replied to, as well as reasons and causes of responses in conversation. It can compress language usage and all the contextual information in dyadic conversations into a latent personal embedding, which presents perspectives for user modeling and personality analysis.

## 3.7 Learning and Prediction

Our GRU models, including word embeddings and personal embeddings, can be learned via predicting the next word in the sentence with a cross-entropy loss. We train all models by employing the derivative of the loss with respect to all parameters through the BPTT [32] algorithm. We iterate over all the samples in the conversational data in each epoch until the convergence is achieved or a maximum epoch number is met. We empirically set the vocabulary size as 5000, the learning rate as 0.01, the maximum epoch number as 100, and select the dimensionality of hidden embeddings in $[25, 50, 75, 100]$.

After personal embeddings, including PSE, PRE and PCE, are well learned with the GRU model, prediction of personality inference can be made. Here, we apply two-layer fully-connected neural networks. In the first layer, we map embeddings from the conversational space to the personality space. For embedding $\mathbf{u}_i$, the formulation can be:

$$\mathbf{s}_i = \sigma\left(\mathbf{u}_i\mathbf{W}_1 + \mathbf{c}_1\right) \ , \quad (5)$$

where $\mathbf{s}_i \in \mathbb{R}^{d'}$, $\mathbf{W}_1 \in \mathbb{R}^{d\times d'}$ and $\mathbf{c}_1 \in \mathbb{R}^{d'}$. $\mathbf{s}_i$ is the joint representation of five traits of the big five personality in a common space. For simplification, we select dimensionality as $d' = d$ in this paper. Then, in the second layer, we can make predictions on the big five personality traits:

$$\mathbf{y}_i = f\left(\mathbf{s}_i\mathbf{W}_2 + \mathbf{c}_2\right) \ , \quad (6)$$

where $\mathbf{y}_i \in \mathbb{R}^5$, $\mathbf{W}_2 \in \mathbb{R}^{d'\times 5}$, $\mathbf{c}_2 \in \mathbb{R}^5$, and $f(x)$ is chosen as a *sigmoid* function $f(x) = \exp\left(1/1+e^{-x}\right)$. $\mathbf{y}_i$ is the predicted values of the five personality traits: extraversion, agreeableness, conscientiousness, neuroticism, and openness. The prediction model can be trained with the widely-used back-propagation algorithm.

# 4. EXPERIMENTS

In this section, we introduce our experiments. First, we introduce our experimental datasets, i.e., the Movie Script dataset and the XiaoIce dataset, and several compared methods. Then, we give comparison among different methods. We also investigate the dimensionality impact and the consistency of personal embeddings. Finally, we apply the learned personal conversational embeddings in individual retrieval and illustrate some results on the Movie Script dataset.

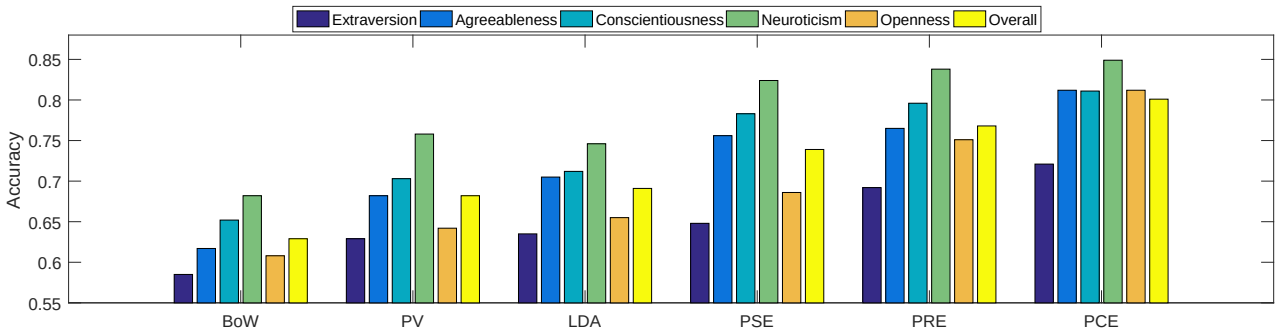
## 4.1 Data

To investigate the effectiveness of personal conversational embeddings, we collect two datasets: the Movie Script dataset and the XiaoIce dataset. Our experiments are conducted on these two datasets.

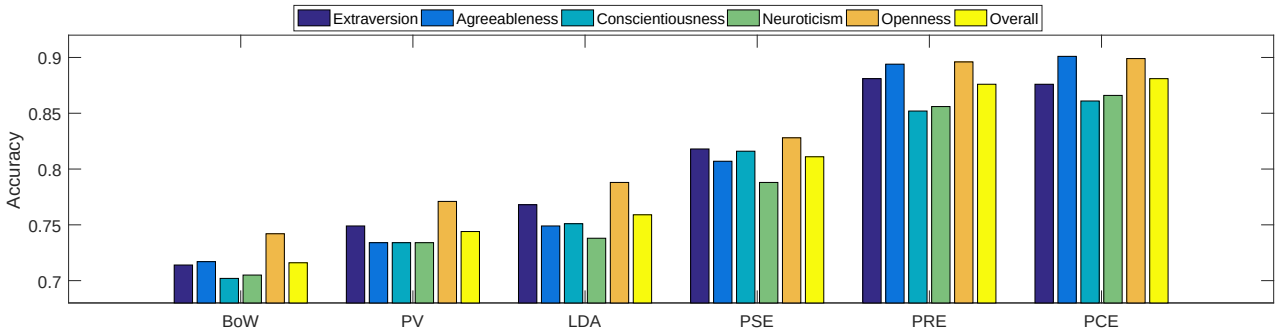
**The Movie Script dataset** is collected via crawling conversations between characters in movie scripts from Internet Movie Script Database (IMSDB)[6]. IMSDB is a website containing thousands of English movie scripts all over the world. Among them, we select 200 movies according to their popularity. The popularity is determined based on the number of ratings on IMDB[7] of each movie, where the threshold is set as 100k. Data from movie scripts has been widely used for constructing datasets for question answering tasks [23]. Moreover, as we know, the personality of characters is important in scripts and novels. In a good script or novel, the personality of a character can cause his or her actions and promote the development of the plot. Thus, it is reasonable to utilize movie scripts for our experiments. The dataset contains a total of about 180k sentences. To avoid data sparsity, we select about 880 characters with at least 50 sentences in scripts. We manually annotate these characters with the big five personality traits for training and evaluation. We label each personality trait in a binary value.

**The XiaoIce dataset** is collected from the XiaoIce chatbot, one of the most popular chatbots in the world created by Microsoft Research. We recruit hundreds of XiaoIce users in China as volunteers for answering a questionnaire. The questionnaire contains 100 question based psychological standards. According to answers towards the questionnaire, their big five personality can be judged and annotated. Then, their conversation records on the XiaoIce chatbot

[6]http://www.imsdb.com/
[7]http://www.imdb.com/




(a) Performance on the Movie Script dataset.




(b) Performance on the XiaoIce dataset.

**Figure 4: Performance comparison on inferring the big five personality traits evaluated by accuracy. The dimensionality of embeddings is set to be $d = 50$. The larger the value, the better the performance.**

are collected with their own permission. In this dataset, removing some users without enough data, we totally have about 660 users and 850k sentences. Moreover, personality inference on XiaoIce helps to promote services and user satisfaction on the chatbot.

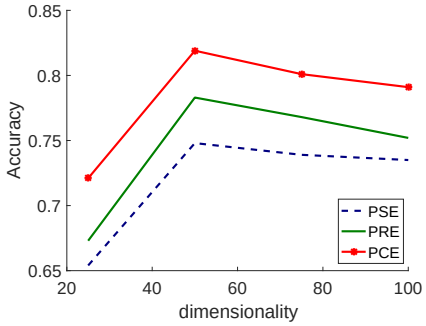
Details of the Movie Script dataset and the XiaoIce dataset are illustrated in Table 2. For both datasets, we randomly select 80% of individuals for training, and the remaining 20% for testing.

## 4.2 Compared Methods

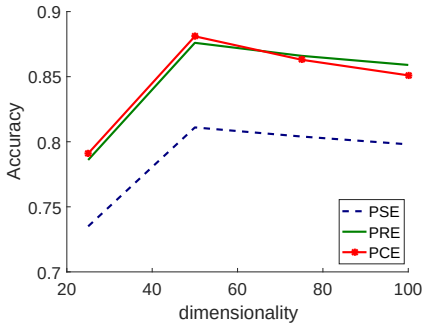
In our experiments, the following methods are implemented and compared on above two datasets:

- **Bag of Words (BoW)**. BoW is a simple baseline method in our experiments. BoW features are extracted on all the sentences an individual has said in the dataset. This method does not utilize the contextual information in conversation. Due to the limitation of annotated data, we select the 300 most significant words according to Pearson correlation coefficient[8]. Based on the 300-dimensional feature, softmax is performed for personality inference.
- **Paragraph Vector (PV)** [22]. PV is a state-of-the-art method to learn unsupervised embeddings for paragraphs based on word co-occurrence. We treat all contents generated by an individual as a paragraph, on which PV features are extracted. This method can compress all the contents of an individual into a fixed-dimensional feature. Then, softmax is performed for personality inference. Obviously, this is a method that lacks modeling of the conversational information.
- **Latent Dirichlet allocation (LDA)** [4, 33]. LDA is a widely-used topic model. It has been applied in personality inference with user generated textual contents, and achieved the state-of-the-art performance. To implement LDA, all contents generated by an individual are treated as a paragraph. This method is neither able to model the conversational information.
- **Personal Speaking Embeddings (PSE)**. Based on an augmented GRU model, PSE learns personal embeddings based on conversational data. This is a baseline among our three proposed methods.
- **Personal Replying Embeddings (PRE)**. Extended from PSE, PRE incorporates an important contextual factor in dyadic conversation: the message before the corresponding response. This can explain the cause of the replying sentence and reduce noise in the data.
- **Personal Conversational Embeddings (PCE)**. As the most advanced one among our methods, PCE models all contextual information in dyadic conversation: the message before the response, as well as personal styles of both sides of dyadic conversations.

[8]https://en.wikipedia.org/wiki/Pearson_product-moment_correlation_coefficient




(a) Performances on the Movie Script dataset.




(b) Performance on the XiaoIce dataset.

**Figure 5: Performance of personality inference with personal embeddings with varying dimensionality $d = [25, 50, 75, 100]$. Performances are evaluated by the overall accuracy of inferring the big five personality traits.**

## 4.3 Performance Comparison

To investigate the effectiveness of our proposed methods incorporating contextual information in dyadic conversation, we illustrate performance comparison with personality $d = 50$ on the Movie Script dataset and the XiaoIce dataset in Figure 4. Results are evaluated by accuracies of the big five personality traits, i.e., extraversion, agreeableness, conscientiousness, neuroticism, and openness, as well as the overall accuracy of the big five personality.

According to results on the Movie Script dataset in Figure 4(a), PV outperforms BoW, indicating that a compressed embedding is better than directly analyzing bag of words for inferring personality. And LDA has very similar performances with PV. Constructed based on the GRU model, PSE can further improve the performance of PV and LDA, showing the advantage of RNN models for learning personal embeddings. Furthermore, modeling the message before the corresponding response, PRE outperforms PSE with a relatively large advantage. PCE, which can incorporate the personal styles of both side of dyadic conversations, has better performance comparing with PRE, and achieves best accuracies on all five personality traits. PCE relatively improves the overall accuracy by 27.3%, 17.4%, 16.0%, 8.4% and 4.3% comparing with BoW, PV, LDA, PSE and PRE respectively. Meanwhile, we observe that, neuroticism has the highest inference accuracy among five personality traits. This may indicate that, whether an individual is sensitive, angry and depressed is distinctive in dramatic plots.

**Table 3: Evaluation of the consistency of personal embeddings via comparing two parts of the conversation of the same individual. Results are evaluated by recall and RMSE.**

| | Movie Script | | XiaoIce | |
|---|---|---|---|---|
| | recall | RMSE | recall | RMSE |
| BoW | 0.198 | —— | 0.189 | —— |
| PV | 0.256 | —— | 0.242 | —— |
| LDA | 0.248 | —— | 0.261 | —— |
| PSE | 0.307 | 0.067 | 0.288 | 0.052 |
| PRE | 0.419 | 0.043 | **0.372** | 0.038 |
| PCE | **0.440** | **0.035** | 0.368 | **0.027** |

Results on the XiaoIce dataset in Figure 4(b) show some similar observations among BoW, PV, LDA, PSE and PRE as those on the Movie Script dataset. Relatively, PRE improves the overall accuracy by 22.3%, 17.7%, 16.1% and 8.0% comparing with BoW, PV, LDA and PSE respectively. However, PCE and PRE has nearly the same performance on the five personality traits. This is because, every user is talking to XiaoIce on the chatbot, which makes individuals being replied to the same one in the whole dataset. This causes the personal styles of the individual being replied to in conversations to be useless. Meanwhile, there are higher overall accuracies on the XiaoIce dataset than those on the Movie Script dataset. Moreover, different from results in Figure 4(a), neuroticism and conscientiousness have lowest accuracies among five personality traits. This may indicate that, users do not tend to show their organized or sensitive sides when casually chatting with a virtual machine.

These significant improvements shown in figures indicate the superiority of our method brought by modeling the rich contextual information in dyadic conversations with an augmented GRU model.

## 4.4 Impact of Dimensionality

To investigate the impact of dimensionality on learning personal embeddings based on dyadic conversation, and select the best parameter for personality inference, we illustrate the performance of PSE, PRE and PCE with varying dimensionality $d = [25, 50, 75, 100]$ on the Movie Script dataset and the XiaoIce dataset in Figure 5. Results are evaluated by the overall accuracy of inferring five traits in the big five personality.

From results in Figure 5, we can obtain some observations. First, we can observe similar performance comparison among PSE, PRE and PCE as in the previous subsection. Second, the best dimensionality of learning personal embeddings on both datasets is clearly $d = 50$, which is used in the rest of our experiments. Third, performances are stable with a large range of dimensionality in $d = [50, 75, 100]$, indicating the flexibility of our methods. Forth, PSE, PRE and PCE share very similar tendency along with dimensionality. But methods incorporating more contextual information tend to be more easy to overfit the data when dimensionality is high, where the performance of PCE is even worse than that of PRE when $d > 50$ on the XiaoIce dataset.

## 4.5 Consistency of Embeddings

Despite having good performance in personality inference, the embedding of an individual should consistently represent

**Table 4: Some results of individual retrieval with personal conversational embeddings on the Movie Script dataset. We illustrate three most similar individuals of each query in the table.**

| query individual | most similar individuals | query individual | most similar individuals |
|---|---|---|---|
| *Marlin* in "Finding Nemo" | *Carl* in "Up"<br>*Gru* in "Despicable Me"<br>*Po's dad* in "Kung Fu Panda" | *Red* in "Shawshank Redemption" | *Master Oogway* in "Kung Fu Panda"<br>*Lincoln* in "Lincoln"<br>*Chuck* in "Cast Away" |
| *Dory* in "Finding Nemo" | *Po* in "Kung Fu Panda"<br>*Olif* in "Frozen"<br>*Agnes* in "Despicable Me" | *Tyler* in "Fight Club" | *Jordan* in "The Wolf of Wall Street"<br>*Holmes* in "Sherlock Holmes"<br>*Calvin* in "Django Unchained" |
| *Michael* in "The Godfather" | *Michael* in "The Godfather 2"<br>*Vito* in "The Godfather 2"<br>*Michael* in "The Godfather 3" | *Fletcher* in "Liar Liar" | *Bruce* in "Bruce Almighty"<br>*Ace* in "Ace Ventura: Pet Detective"<br>*Stanley* in "The Mask" |

the individual, and should not change very much in different time periods and situations. Accordingly, we should conduct experiments to evaluate the consistency of personal embeddings. Thus, we divide data from an individual in the dataset into two equal parts in chronological order. For two part of an individual's conversational data, embeddings are learned and features are extracted separately. To avoid data sparsity, we only select individuals with at least 100 sentences. Experimental results on the Movie Script and the XiaoIce dataset are shown in Table 3.

First, we use the first part of an individual to retrieve for the second part of the same individual based on Euclidean distance. With better embeddings, more correct pairs should be matched. The evaluation metrics used here is recall. Recall is calculated as the ratio of the number of correctly found pairs and the total number of positive samples. The larger the recall value, the better the consistency. From results in Table 3, we can observe that PRE and PCE can significantly outperform other methods on both datasets. On the Movie Script dataset, the recall value of PCE is slightly better than that of PRE. While on the XiaoIce dataset, the recall value of PCE becomes a little worse. This is because, all users are talking to the same individual, i.e., XiaoIce, making embeddings less distinctive.

Second, we evaluate the difference between embeddings of two parts of the same individual. We use Root Mean Square Error (RMSE)[9] to evaluate embedding differences. The smaller the RMSE value, the better the consistency. Because BoW features are not real value embeddings, PV and LDA are not trained with GRU, they have different ranges of values comparing with PSE, PRE and PCE. It is not meaningful to compute their embedding differences to compare with our proposed personal embeddings. Accordingly, we ignore the RMSE evaluation of consistency of BoW, PV and LDA. Results in Table 3 show the lowest RMSE values of PCE on both datasets with relatively large advantages. These results indicate the good consistency of personal conversational embeddings for representing individuals.

## 4.6 Individual Retrieval

In this subsection, we investigate a case study on individual retrieval on the Movie Script dataset to demonstrate advantages of personal conversational embeddings. We use some famous characters in movies as queries, and retrieve for other similar characters in the dataset based on Euclidean distance with personal conversational embeddings learned from their conversations. We illustrate some representative results in Table 4. For each query, three most similar characters are shown.

From the illustration, we can observe some interesting phenomena and suitable matchings. In the animated film "Finding Nemo", we have two characters, *Marlin* and *Dory*. *Marlin* is matched with *Carl*, *Gru*, and Po's dad, while *Dory* is matched with *Po*, *Olif*, and *Agnes*. Former ones are conservative fathers, and latter ones are childish characters. *Red* in "Shawshank Redemption" is matched with *Master Oogway*, *Lincoln*, and *Chuck*, who are all wise elders. *Tyler* in "Fight Club" is matched with *Jordan*, *Holmes*, and *Calvin*, who are all insane, insolent, and arrogant individuals. *Michael* in "The Godfather" is matched with himself in other chapters of the movie series, as well as his father at a young age. Due to the consistent nonsensical style of Jim Carrey's comedies, several roles he played, i.e., *Fletcher*, *Bruce*, *Ace* and *Stanley*, are matched. Obviously, matched characters share very similar personalities. This indicates that, personal conversational embeddings are able to well capture the characteristics of an individual.

# 5. CONCLUSIONS AND FUTURE WORK

In this paper, we propose a novel method for personality inference. With an augmented GRU model, we learn unsupervised personal conversational embeddings based on dyadic conversations between individuals. We incorporate previous messages of corresponding responses via sequence to sequence learning. The formulation of each layer of networks is adjusted with personal styles of both sides of the dyadic conversation. With the learned personal conversational embeddings, personality inference can be performed. According to experiments conducted on two datasets, i.e., the Movie Script dataset and the XiaoIce dataset, accuracy of personality inference can be significantly improved by modeling the rich contextual information in conversation records. Experimental results show the successful performance of our proposed method.

In the future, we will further investigate the following directions. First, social relationship is an important factor for personality analysis. So, it is reasonable to investigating the relationship between individuals in the conversation. Second, we can incorporate more contextual information, such as location, time, recent topics or even social environment. The main challenge for doing this is finding available data.

[9]https://www.kaggle.com/wiki/RootMeanSquaredError

# 6. REFERENCES


[1] J. B. Asendorpf and S. Wilpers. Personality effects on social relationships. *Journal of personality and social psychology*, 74(6):1531, 1998.

[2] Y. Bachrach, M. Kosinski, T. Graepel, P. Kohli, and D. Stillwell. Personality and patterns of facebook usage. In *Annual ACM Web Science Conference*, pages 24–32, 2012.

[3] Y. Bengio, P. Simard, and P. Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on Neural Networks*, 5(2):157–166, 1994.

[4] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

[5] I. Cantador, I. Fernández-Tobías, and A. Bellogín. Relating personality types with user preferences in multiple entertainment domains. In *CEUR Workshop Proceedings*, 2013.

[6] J. Chen, E. Haber, R. Kang, G. Hsieh, and J. Mahmud. Making use of derived personality: The case of social media ad targeting. In *ICWSM*, pages 51–60, 2015.

[7] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.

[8] J. Chung, C. Gülçehre, K. Cho, and Y. Bengio. Gated feedback recurrent neural networks. In *ICML*, pages 2067–2075, 2015.

[9] G. Farnadi, G. Sitaraman, S. Sushmita, F. Celli, M. Kosinski, D. Stillwell, S. Davalos, M.-F. Moens, and M. De Cock. Computational personality recognition in social media. In *User Modeling and User-Adapted Interaction*, pages 1–34, 2016.

[10] G. Farnadi, S. Zoghbi, M.-F. Moens, and M. De Cock. Recognising personality traits using facebook status updates. In *ICWSM*, 2013.

[11] S. Ghosh, O. Vinyals, B. Strope, S. Roy, T. Dean, and L. Heck. Contextual lstm (clstm) models for large scale nlp tasks. *arXiv preprint arXiv:1602.06291*, 2016.

[12] J. Golbeck, C. Robles, M. Edmondson, and K. Turner. Predicting personality from twitter. In *Social Computing and Networking*, pages 149–156, 2011.

[13] J. Golbeck, C. Robles, and K. Turner. Predicting personality with social media. In *International Conference on Human Factors in Computing Systems*, pages 253–262, 2011.

[14] G. Hagger-Johnson, V. Egan, and D. Stillwell. Are social networking profiles reliable indicators of sensational interests? *Journal of Research in Personality*, 45(1):71–76, 2011.

[15] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[16] D. J. Hughes, M. Rowe, M. Batey, and A. Lee. A tale of two sites: Twitter vs. facebook and the personality predictors of social media usage. *Computers in Human Behavior*, 28(2):561–569, 2012.

[17] O. P. John and S. Srivastava. The big five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research*, 2(1999):102–138, 1999.

[18] F. Kooti, L. M. Aiello, M. Grbovic, K. Lerman, and A. Mantrach. Evolution of conversations in the age of email overload. In *WWW*, pages 603–613, 2015.

[19] M. Kosinski, Y. Bachrach, P. Kohli, D. Stillwell, and T. Graepel. Manifestations of user personality in website choice and behaviour on online social networks. *Machine learning*, 95(3):357–380, 2014.

[20] M. Kosinski, D. Stillwell, and T. Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15):5802–5805, 2013.

[21] R. Lambiotte and M. Kosinski. Tracking the digital footprints of personality. *Proceedings of the IEEE*, 102(12):1934–1939, 2014.

[22] Q. V. Le and T. Mikolov. Distributed representations of sentences and documents. In *ICML*, pages 1188–1196, 2014.

[23] J. Li, M. Galley, C. Brockett, J. Gao, and B. Dolan. A persona-based neural conversation model. In *ACL*, pages 994–1003, 2016.

[24] L. Liu, D. Preotiuc-Pietro, Z. R. Samani, M. E. Moghaddam, and L. Ungar. Analyzing personality through social media profile picture choice. In *ICWSM*, pages 211–220, 2016.

[25] Q. Liu, S. Wu, D. Wang, Z. Li, and L. Wang. Context-aware sequential recommendation. In *ICDM*, pages 1053–1058, 2016.

[26] Q. Liu, S. Wu, L. Wang, and T. Tan. Predicting the next location: A recurrent model with spatial and temporal contexts. In *AAAI*, pages 194–200, 2016.

[27] J. Mao, W. Xu, Y. Yang, J. Wang, Z. Huang, and A. Yuille. Deep captioning with multimodal recurrent neural networks (m-rnn). In *ICLR*, 2014.

[28] D. J. Ozer and V. Benet-Martinez. Personality and the prediction of consequential outcomes. *Annual Review of Psychology*, 57:401–421, 2006.

[29] G. Park, H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, M. Kosinski, D. J. Stillwell, L. H. Ungar, and M. E. Seligman. Automatic personality assessment through social media language. *Journal of personality and social psychology*, 108(6):934, 2015.

[30] D. Quercia, M. Kosinski, D. Stillwell, and J. Crowcroft. Our twitter profiles, our selves: Predicting personality with twitter. In *Social Computing and Networking*, pages 180–185, 2011.

[31] M. Ren, R. Kiros, and R. Zemel. Exploring models and data for image question answering. In *NIPS*, pages 2953–2961, 2015.

[32] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *Cognitive modeling*, 5:3, 1988.

[33] H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, L. Dziurzynski, S. M. Ramones, M. Agrawal, A. Shah, M. Kosinski, D. Stillwell, M. E. Seligman, et al. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one*, 8(9):1–16, 2013.

[34] I. V. Serban, A. Sordoni, Y. Bengio, A. Courville, and J. Pineau. Building end-to-end dialogue systems using generative hierarchical neural network models. In *AAAI*, pages 3776–3784, 2016.

[35] L. Shang, Z. Lu, and H. Li. Neural responding machine for short-text conversation. In *ACL*, pages 1577–1586, 2015.

[36] A. Sordoni, M. Galley, M. Auli, C. Brockett, Y. Ji, M. Mitchell, J.-Y. Nie, J. Gao, and B. Dolan. A neural network approach to context-sensitive generation of conversational responses. In *NAACL*, pages 196–205, 2015.

[37] O. Vinyals and Q. Le. A neural conversational model. *arXiv preprint arXiv:1506.05869*, 2015.

[38] H. Wei, F. Zhang, J. Yuan, C. Cao, H. Fu, X. Xie, and Y. Rui. Beyond the words: Predicting user personality from heterogeneous information. In *WSDM*, pages 305–314, 2017.

[39] Y. Wu, M. Kosinski, and D. Stillwell. Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, 112(4):1036–1040, 2015.