

# Personal Attribute Prediction from Conversations

Yinan Liu, Hu Chen, Wei Shen\*

TKLNDST, College of Computer Science, Nankai University, Tianjin, 300350, China  
{liuyn,2120210473}@mail.nankai.edu.cn,shenwei@nankai.edu.cn

## ABSTRACT

Personal knowledge bases (PKBs) are critical to many applications, such as Web-based chatbots and personalized recommendation. Conversations containing rich personal knowledge can be regarded as a main source to populate the PKB. Given a user, a user attribute, and user utterances from a conversational system, we aim to predict the personal attribute value for the user, which is helpful for the enrichment of PKBs. However, there are three issues existing in previous studies: (1) manually labeled utterances are required for model training; (2) personal attribute knowledge embedded in both utterances and external resources is underutilized; (3) the performance on predicting some difficult personal attributes is unsatisfactory. In this paper, we propose a framework DSCGN based on the pre-trained language model with a noise-robust loss function to predict personal attributes from conversations without requiring any labeled utterances. We yield two categories of supervision, i.e., document-level supervision via a distant supervision strategy and contextualized word-level supervision via a label guessing method, by mining the personal attribute knowledge embedded in both unlabeled utterances and external resources to fine-tune the language model. Extensive experiments over two real-world data sets (i.e., a profession data set and a hobby data set) show our framework obtains the best performance compared with all the twelve baselines in terms of nDCG and MRR.

## CCS CONCEPTS

• **Computing methodologies** → Information extraction.

## KEYWORDS

Conversations; Personal Attribute Knowledge; Language Model

### ACM Reference Format:

Yinan Liu, Hu Chen, Wei Shen\*. 2022. Personal Attribute Prediction from Conversations. In *Companion Proceedings of the Web Conference 2022 (WWW '22 Companion)*, April 25–29, 2022, Virtual Event, Lyon, France. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3487553.3524248>

## 1 INTRODUCTION

Some recent studies [1, 18] propose to construct the personal knowledge base (PKB) for individuals, which contains a resource of structured information about entities personally related to its user, their

\*Corresponding author

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).  
*WWW '22 Companion*, April 25–29, 2022, Virtual Event, Lyon, France

© 2022 Association for Computing Machinery.

ACM ISBN 978-1-4503-9130-6/22/04...\$15.00

<https://doi.org/10.1145/3487553.3524248>

attributes, and the relations between them. The advantages of the PKB are twofold. On one hand, a PKB can be merged with large general KBs (e.g., Freebase, DBpedia, and YAGO), so that the personal life events are connected with world knowledge. On the other hand, the PKB can provide personal background knowledge for many downstream applications such as Web-based chatbots, personalized recommendation, and personalized search.

Conversational systems are becoming very important as many personal assistants emerge (e.g., Siri and Cortana). Understanding user utterances plays an essential role in holding meaningful conversations with users. In reality, the PKB is beneficial for understanding user utterances better. Meantime, conversation data can also be regarded as a main source to populate a PKB by drawing personal attribute knowledge (e.g., hobbies, professions, and visited cities) from the users' conversations in many platforms (e.g., social media and dialogues). To enrich the PKB using personal knowledge from conversations, predicting personal attributes from conversations is an important task that needs to be solved urgently.

Given a subject (i.e., a user of a conversational system), a predicate (i.e., a user attribute), and user utterances, our task is to predict the ranking of the object (i.e., attribute value) for the given subject-predicate combination according to user utterances. For instance, we could rank the student and doctor professions high when the user mentions the words *undergrad*, *school*, *study*, *medicine*, etc. Subsequently, the generated subject-predicate-object triple could be used to populate the PKB.

Recently, several supervised models [15, 16] have been proposed to predict personal attributes from conversations. The first work [15] leveraged neural networks based on utterance embeddings and attention mechanisms to solve this task. Tiginova et al. [16] achieved better performance by designing a reinforce policy to integrate utterance keyword extraction and Web document retrieval. However, there exist three issues in previous works: (1) the construction of training data via manually annotating utterances is required and difficult; (2) personal attribute knowledge embedded in both utterances and external resources (e.g., Wikipedia) is underutilized; (3) the performance on predicting some difficult personal attributes (e.g., profession and hobby) with too many candidate values is unsatisfactory.

To deal with these issues, we propose a novel end-to-end language model based framework DSCGN without requiring any labeled utterances, which derives two categories of supervision, i.e., document-level supervision via a Distant Supervision strategy and Contextualized word-level supervision via a label Guessing method by mining personal attribute knowledge embedded in unlabeled utterances and external resources. DSCGN performs confidence based pre-trained language model (PLM) fine-tuning via a Noise-robust loss function incorporating these two categories of supervision. Our contributions can be summarized as follows:

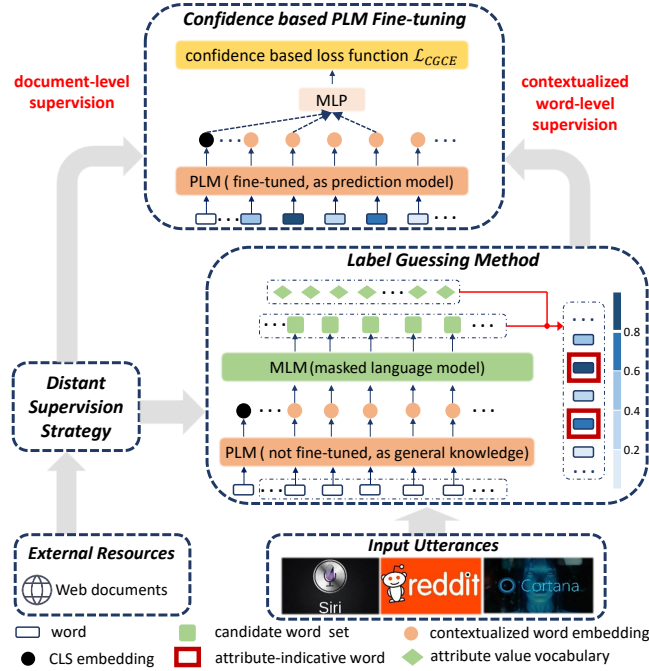


Figure 1: Overview of our framework.

- We propose the first language model based framework without requiring any labeled utterances to predict personal attributes from conversations.
- We propose a noise-robust loss function that can incorporate document-level supervision and contextualized word-level supervision to fine-tune the language model.
- Extensive experiments over two real-world data sets (i.e., a profession data set and a hobby data set), which are derived from Reddit show our framework obtains the best performance compared with all the twelve baselines in terms of nDCG and MRR.

## 2 THE FRAMEWORK: DSCGN

The overview of our framework is shown in Figure 1. We first introduce the distant supervision strategy and the label guessing method. Finally, we describe confidence based PLM fine-tuning.

### 2.1 Distant Supervision Strategy

To tackle the labeled utterances shortage issue, we leverage a distant supervision strategy to annotate external resources automatically to obtain document-level supervision. Intuitively, Wikipedia is beneficial for predicting personal attributes from conversations as a distant knowledge resource, since the semantics of personal attribute values (i.e., teacher, driver, and dancer) can be interpreted by their corresponding Wikipedia pages (Wiki-pages). Additionally, we found that some pages in Wikipedia about concepts related to the attribute values (e.g., tools used for a profession and the profession’s specializations) are also helpful for our task. Accordingly, we could mine the attribute knowledge embedded in Wikipedia to help predict personal attributes from conversations. To derive high-quality labeled Wikipedia data, for each personal attribute value, we select its corresponding Wiki-page by string matching based on the surface forms of the attribute value (e.g., Dentist) and the Wiki-page title (e.g., Wiki: Dentist). To obtain more labeled Wikipedia

data, we construct Wiki-category by identifying at least one relevant category for each attribute value and including all leaf pages under the category like [16]. In this way, the document-level supervision  $S_{doc} = \{(d_i, A_{d_i})\}$  could be derived automatically, where  $d_i$  denotes a Web document from external resources (e.g., the Wiki-page or a page in Wiki-category) and  $A_{d_i}$  denotes the corresponding attribute value of  $d_i$ .

### 2.2 Label Guessing Method

When provided an attribute value, humans can understand its semantics based on general knowledge by associating it with some attribute-indicative words. To create contextualized word-level attribute value supervision (contextualized word-level supervision) automatically, inspired by [9], we propose a label guessing method to find attribute-indicative words via the PLM pre-trained with the masked language model (MLM) task on large-scale text corpora.

First, for the  $i$ -th occurrence  $w_i$  of the word  $w$  in unlabeled utterances and external resources (e.g., Wiki-page and Wiki-category) provided by the distant supervision strategy, we feed its contextualized embedding  $h_{w_i}$  produced by PLM encoder to the MLM head, which will output a probability distribution over the entire vocabulary  $V$ , indicating the likelihood of each word  $v \in V$  appearing at this position.

$$p(v|h_{w_i}) = \text{Softmax}(W_2 \phi(W_1 h_{w_i} + b)),$$

where  $\phi(\cdot)$  is the activation function;  $W_1 \in \mathbb{R}^{h \times h}$ ,  $W_2 \in \mathbb{R}^{|V| \times h}$ , and  $b \in \mathbb{R}^h$  are learnable parameters pre-trained with the MLM objective of PLM. Intuitively, words that are interchangeable most of the time are likely to have similar meanings. We retain top- $K$  (sorted by  $p(v|h_{w_i})$ ) similar words as the candidate word set  $C_{w_i}$  for the word occurrence  $w_i$ . Thus, we can generate  $K$  valid replacement words for each occurrence of the attribute value. Subsequently, for an attribute value  $a$ , we form the attribute value vocabulary  $V_a$  using the top- $M$  words ranked by frequency in all candidate word sets of its occurrences ( $C_{a_1}, C_{a_2}, \dots, C_{a_n}$ ), discarding stopwords and words that appear in many attribute value vocabularies. Then, we define the semantic similarity  $Sim_{w_2a}(w_i, a)$  between word  $w_i$  and attribute value  $a$  as follows.

$$Sim_{w_2a}(w_i, a) = \frac{|C_{w_i} \cap V_a|}{|V_a|}$$

If this similarity is larger than a threshold  $\beta$ ,  $w_i$  is regarded as an attribute-indicative word with regard to the attribute value  $a$ . Thereby, from unlabeled utterances and external resources, we obtain contextualized word-level supervision  $S_{word} = \{(w_i, A_{w_i})\}$ , where  $w_i$  denotes an attribute-indicative word and  $A_{w_i}$  denotes the corresponding attribute value of  $w_i$ .

### 2.3 Confidence based PLM Fine-tuning

A PLM can be fine-tuned by leveraging task-specific training data so that it can achieve a better result when performing that task. In this paper, to leverage the personal attribute knowledge embedded in both unlabeled utterances and external resources for predicting personal attributes from conversations, we try to fine-tune the PLM based on the document-level supervision  $S_{doc}$  yielded by the distant supervision strategy (Section 2.1) and contextualized word-level supervision  $S_{word}$  derived by the label guessing method (Section 2.2).

Specifically, considering the balance between model convergence and noise-robustness, and in order to give more attention to the

true attribute-indicative words and relieve the effect of the mislabeled words, we propose a confidence based loss function  $\mathcal{L}_{CGCE}$  based on the generalized cross entropy (GCE) loss [19] by incorporating the confidence based weight function  $\xi(\cdot)$  which is calculated by the semantic similarity  $Sim_{w2a}(\cdot)$  as follows:

$$\mathcal{L}_{CGCE} = - \sum_{t \in S_{doc} \cup S_{word}} \xi(t) \frac{1 - \text{Softmax}(W_{A_t} h_t + b_{A_t})^q}{q}$$

$$\xi(t) = \begin{cases} Sim_{w2a}(t, A_t) & \text{if } t \text{ refers to an attribute-indicative word} \\ & \text{from } S_{word}, \\ 1 & \text{if } t \text{ refers to a Web document from } S_{doc}. \end{cases}$$

where  $W_{A_t} \in \mathbb{R}^{K \times h}$  and  $b_{A_t} \in \mathbb{R}^K$  are learnable parameters of the linear layer ( $K$  is the number of attribute values);  $0 < q < 1$  is a hyperparameter; if  $t$  refers to an attribute-indicative word from  $S_{word}$ ,  $h_t$  denotes its contextualized embedding, otherwise if  $t$  refers to a Web document from  $S_{doc}$ ,  $h_t$  denotes its CLS embedding.

After fine-tuning, given a user and a user attribute, we combine all utterances of a user as a sentence and give this sentence as the input of the fine-tuned PLM to obtain the probabilities of different attribute values of this attribute.

### 3 EXPERIMENTAL STUDY

#### 3.1 Experimental Setting

**Data Sets.** Two real-world data sets (i.e., a profession data set and a hobby data set) are used in this experiment. They are annotated and publicly provided by [16], and extracted from publicly-available Reddit submissions and comments between 2006 and 2018. They both consist of about 6000 users, with a maximum of 500 and an average of 23 users per attribute value. The number of attribute values for profession (hobby) is 71 (149) respectively. Specially, for each Reddit user, all posts containing explicit personal assertions used for labeling have been removed.

**Evaluation Measures.** We adopt the same evaluation metrics MRR (Mean Reciprocal Rank) and nDCG (normalized Discounted Cumulative Gain) as previous works [15, 16]. The nDCG score is calculated by averaging all users. Considering that a user may have many attribute values, we follow the assumption that there is only one correct attribute value for each user. Thus, MRR can be calculated independently for each attribute value before averaging.

**Baseline Models.** No-keyword + BM25 [16] utilizes a user’s full utterances and a Web document from  $S_{doc}$  w.r.t. an attribute value as the input of BM25 [12]. BM25 is a strong unsupervised retrieval model to calculate the similarity between two documents. RAKE (TextRank) + BM25 (KNRM) [16] uses keywords extracted from a user’s full utterances by RAKE (TextRank) and a Web document from  $S_{doc}$  w.r.t. an attribute value as the input of BM25 (KNRM). RAKE [13] and TextRank [10] are two SOTA unsupervised keyword extraction methods. KNRM [17] is an efficient neural retrieval model, which considers semantic similarity via term embeddings. BERT IR [16] splits both user utterances and a Web document from  $S_{doc}$  into 256-token chunks respectively to fit the input size of BERT, and predicts the relevance between them based on a binary cross-entropy loss.  $HAM_{avg}$  ( $HAM_{CNN}$ ,  $HAM_{CNN-attn}$ , and  $HAM_{2attn}$ ) [15] can predict the score of different attribute values by training the neural network based on  $S_{doc}$ .  $HAM_{CNN-attn}$  and

**Table 1: Performance on predicting personal attributes from conversations using the external resource (Wiki-page or Wiki-category).**

Labeled utterances	Method	Wiki-page				Wiki-category			
		profession		hobby		profession		hobby	
		MRR	nDCG	MRR	nDCG	MRR	nDCG	MRR	nDCG
no	No-keyword + BM25	.15	.32	.16	.42	.17	.37	.13	.35
	RAKE + BM25	.16	.33	.17	.42	.19	.39	.14	.37
	RAKE + KNRM	.16	.33	.12	.32	.13	.34	.12	.31
	TextRank + BM25	.21	.39	.21	.46	.26	.45	.20	.42
	TextRank + KNRM	.21	.38	.15	.36	.18	.36	.16	.36
	$HAM_{avg}$	.06	.07	.06	.05	.06	.07	.03	.02
	$HAM_{2attn}$	.06	.07	.04	.05	.06	.07	.06	.07
	$HAM_{CNN}$	.20	.18	.22	.14	.27	.34	.17	.27
	$HAM_{CNN-attn}$	.21	.28	.13	.10	.25	.31	.16	.25
	DSCGN	.43	.57	.29	.50	.44	.60	.29	.49
yes	BERT IR	.30	.45	.22	.43	.28	.44	.18	.42
	$CHARM_{BM25}$	.29	.46	.24	.47	.28	.47	.21	.43
	$CHARM_{KNRM}$	.27	.44	.22	.44	.35	.55	.27	.49

$HAM_{2attn}$  both utilize attention mechanisms within and across documents from  $S_{doc}$ .  $HAM_{avg}$  and  $HAM_{2attn}$  both adopt two stacked fully connected layers, while  $HAM_{CNN}$  and  $HAM_{CNN-attn}$  both adopt text classification CNN.  $CHARM_{BM25}$  ( $CHARM_{KNRM}$ ) [16] uses keywords extracted from a user’s full utterances by BERT and a Web document from  $S_{doc}$  as the input of BM25 (KNRM).

**Setting Details.** We utilize the BERT base-uncased model as the PLM. The training batch size and the learning rate are set to 4 and 1e-5 respectively. The parameters  $K$ ,  $M$ , and  $\beta$  of the label guessing method are set to 50, 100, and 0.4 respectively. The hyperparameter  $q$  is set to 0.4. It is noted that the adopted lists of personal attribute values and the result for each baseline except  $HAM_{avg}$ ,  $HAM_{2attn}$ ,  $HAM_{CNN}$ , and  $HAM_{CNN-attn}$ , are taken from [16] directly. To represent utterance embedding, the baselines [15] use wordvec embeddings pre-trained on the Google News corpus via MindSpore Framework<sup>1</sup>. Source code used in this paper is publicly available<sup>2</sup>.

#### 3.2 Experimental Results

Besides external resources, the baselines BERT IR,  $CHARM_{BM25}$ , and  $CHARM_{KNRM}$  require extra labeled utterances for model training, and perform ten fold cross-validation on the profession (hobby) data set under the zero-shot setting in which the attribute values in the training and test data are disjoint. Our method DSCGN and other baselines do not require labeled utterances as input and perform on the profession (hobby) data set directly, which is a more difficult setting, since it pushes “zero-shot” to the extreme – no labeled utterances for any attribute values are given.

**Effectiveness Study.** From the results in Table 1, it can be seen that whichever external resource (i.e., Wiki-page or Wiki-category) is used, our proposed method DSCGN obtains the best performance compared with the twelve baselines on both data sets in terms of MRR and nDCG. To be specific, there are nine baselines that just leverage the external resources as distant supervision like us. Nevertheless, our method significantly outperforms them, which shows that our method can leverage the personal attribute knowledge embedded in external resources better. In addition to the external resources, the baselines  $CHARM_{BM25}$ ,  $CHARM_{KNRM}$ , and BERT IR utilize extra labeled utterances for model training, which puts DSCGN at a disadvantage. However, in spite of this disadvantage, we found that our framework still promotes by at least about 9 (5) percentages compared with the best baseline in terms of MRR

<sup>1</sup><https://www.mindspore.cn/en>

<sup>2</sup><https://github.com/CodingPerson/DSCGN>

**Table 2: Ablation study.**

Ablations	profession		hobby	
	Wiki-category		Wiki-page	
	MRR	nDCG	MRR	nDCG
DSCGN	<b>0.44</b>	<b>0.60</b>	<b>0.29</b>	<b>0.50</b>
DSCGN w/o confidence based weight function $\xi(\cdot)$	0.42	0.58	0.27	0.48
DSCGN w/o label guessing method	0.38	0.54	0.27	0.48
DSCGN w/o distant supervision strategy	0.06	0.23	0.04	0.20

(nDCG) over the profession data set, and obtains the best performance in terms of MRR (nDCG) over the hobby data set, indicating the superiority of our proposed framework for the task of predicting personal attributes from conversations.

**Ablation Study.** To show the different parts of our framework (i.e., confidence based weight function  $\xi(\cdot)$ , label guessing method, and distant supervision strategy) are crucial for our task, we conduct an ablation study by removing each part of the proposed framework DSCGN and testing the performance of the remaining parts over the profession (hobby) data set using Wiki-category (Wiki-page) as the external resource. From the experiment results shown in Table 2, compared with these truncated versions of DSCGN, the full version of DSCGN obtains the best performance, which demonstrates each part of our framework is helpful for our task. To be specific, removing the label guessing method or the distant supervision strategy, the performance of DSCGN degrades a lot, indicating the personal attribute knowledge embedded in utterances and external resources are complementary.

## 4 RELATED WORK

**Personal Attribute Prediction from Conversations.** Previous works [4, 5] relied on the condition that user attributes are explicitly mentioned in the user utterances, so these methods are not suitable for the profession (hobby) data set without personal assertions used in this paper. Recently, HAM [15] proposed some models based on neural network architectures (e.g., stacked fully connected layers and CNNs), to predict a score of different attribute values (e.g., different professions) for a given subject-predicate pair by using average methods or attention mechanisms within and across utterances. CHARM [16] first leveraged the pre-trained LM BERT to extract keywords related to the attribute from user utterances, and then adopted some SOTA ranking models based on information retrieval techniques (i.e., RAKE and KNRM) to match these keywords against documents that indicate possible values of the attribute. Specially, they proposed a reinforce policy gradient method to train the process of keywords extraction.

**Personal Attribute Prediction from Social Media.** In recent years, numerous methods have been proposed to predict attributes (e.g., age [3, 7], gender [3, 8], and occupational class [11]) of users from social media (e.g., Twitter and Facebook). The vast majority of these works rely on rich meta-data of social media such as hashtag, user profile description, and social network structure. This makes them unsuitable for our task, as conversations do not have these rich meta-data. In addition, there are only three works [2, 3, 11] that predict user latent attributes based solely on user-generated text. However, from the results provided by [15], we found that these three methods perform not well on the task of predicting personal attributes from conversations, so we did not select them as our baselines in this paper. Additionally, some recent works explored the task of named entity location prediction from Twitter [6, 14], but their methods are not applicable for personal attributes.

## 5 CONCLUSION

To tackle three issues existing in previous studies (i.e., labeled utterances shortage, personal attribute knowledge underutilization, and unsatisfactory performance), we propose a flexible PLM based framework DSCGN without requiring any labeled utterances. Two categories of supervision (i.e., document-level supervision and contextualized word-level supervision) are utilized by DSCGN to fine-tune the language model via a noise-robust loss function. To demonstrate the effectiveness of DSCGN, we conduct extensive experiments over two real-world data sets, which show whichever external resource (i.e., Wiki-page or Wiki-category) is used, our framework obtains the best performance compared with all the twelve baselines in terms of nDCG and MRR.

## 6 ACKNOWLEDGMENTS

This work was supported in part by National Natural Science Foundation of China (No. U1936206), Natural Science Foundation of Tianjin (No. 19JCQNJC00100), YESS by CAST (No. 2019QNRC001), and CAAI-Huawei MindSpore Open Fund.

## REFERENCES

- [1] Krisztian Balog and Tom Kenter. 2019. Personal knowledge graphs: A research agenda. In *SIGIR*. 217–220.
- [2] Angelo Basile, Gareth Dwyer, Maria Medvedeva, Josine Rawee, Hessel Haagsma, and Malvina Nissim. 2017. N-GrAM: New Groningen Author-profiling Model—Notebook for PAN. In *CEUR Workshop Proceedings*, Vol. 1866.
- [3] Roy Khristopher Bayot and Teresa Gonçalves. 2017. Age and gender classification of tweets using convolutional neural networks. In *International Workshop on Machine Learning, Optimization, and Big Data*. Springer, 337–348.
- [4] Hongyan Jing, Nanda Kambhatla, and Salim Roukos. 2007. Extracting social networks and biographical facts from conversational speech transcripts. In *ACL*. 1040–1047.
- [5] Xiang Li, Gokhan Tur, Dilek Hakkani-Tür, and Qi Li. 2014. Personal knowledge graph population from user utterances in conversational understanding. In *2014 IEEE Spoken Language Technology Workshop (SLT)*. IEEE, 224–229.
- [6] Yinan Liu, Wei Shen, Zonghai Yao, Jianyong Wang, Zhenglu Yang, and Xiaojie Yuan. 2021. Named entity location prediction combining twitter and web. *IEEE Transactions on Knowledge and Data Engineering* 33, 11 (2021), 3618–3633.
- [7] Yaguang Liu and Lisa Singh. 2021. Age Inference Using A Hierarchical Attention Neural Network. In *CIKM*. 3273–3277.
- [8] Sunghwan Mac Kim, Qionghai Xu, Lizhen Qu, Stephen Wan, and Cécile Paris. 2017. Demographic inference on twitter using recursive neural networks. In *ACL*. 471–477.
- [9] Yu Meng, Yunyi Zhang, Jiaxin Huang, Chenyan Xiong, Heng Ji, Chao Zhang, and Jiawei Han. 2020. Text Classification Using Label Names Only: A Language Model Self-Training Approach. In *EMNLP*. 9006–9017.
- [10] Rada Mihalcea and Paul Tarau. 2004. TextRank: Bringing order into text. In *EMNLP*. 404–411.
- [11] Daniel Preoțiuc-Pietro, Vasileios Lamos, and Nikolaos Aletras. 2015. An analysis of the user occupational class through Twitter content. In *ACL*. 1754–1764.
- [12] Stephen Robertson and Hugo Zaragoza. 2009. The Probabilistic Relevance Framework: BM25 and Beyond. *Foundations and Trends in Information Retrieval* 3, 4 (2009), 333–389.
- [13] Stuart Rose, Dave Engel, Nick Cramer, and Wendy Cowley. 2010. Automatic keyword extraction from individual documents. *Text mining: applications and theory* 1 (2010), 1–20.
- [14] Wei Shen, Yinan Liu, and Jianyong Wang. 2018. Predicting named entity location using Twitter. In *ICDE*. 161–172.
- [15] Anna Tiginova, Andrew Yates, Paramita Mirza, and Gerhard Weikum. 2019. Listening between the lines: Learning personal attributes from conversations. In *WWW*. 1818–1828.
- [16] Anna Tiginova, Andrew Yates, Paramita Mirza, and Gerhard Weikum. 2020. CHARM: Inferring Personal Attributes from Conversations. In *EMNLP*. 5391–5404.
- [17] Chenyan Xiong, Zhuyun Dai, Jamie Callan, Zhiyuan Liu, and Russell Power. 2017. End-to-end neural ad-hoc ranking with kernel pooling. In *SIGIR*. 55–64.
- [18] An-Zi Yen, Hen-Hsen Huang, and Hsin-Hsi Chen. 2019. Personal Knowledge Base Construction from Text-based Lifelogs. In *SIGIR*. 185–194.
- [19] Zhilu Zhang and Mert R Sabuncu. 2018. Generalized cross entropy loss for training deep neural networks with noisy labels. In *NIPS*.