Persian Language Understanding in Task-oriented Dialogue System for Online Shopping

Zeinab Borhanifard
School of Electrical and Computer Engineering
University of Tehran
Tehran, Iran
borhanifardz@ut.ac.ir

Hossein Basafa
School of Electrical and Computer Engineering
University of Tehran
Tehran, Iran
hbasafa@ut.ac.ir

Seyedeh Zahra Razavi
Department of Computer Science
University of Rochester
Rochester, USA
srazavi@cs.rochester.edu

Heshaam Faili
School of Electrical and Computer Engineering
University of Tehran
Tehran, Iran
hfaili@ut.ac.ir

***Abstract*—Natural language understanding is a critical module in task-oriented dialogue systems. Recently, state-of-the-art approaches use deep learning methods and transformers to improve the performance of dialogue systems. In this work, we propose a natural language understanding model with a specific-shopping named entity recognizer using a joint learning-based BERT transformer for task-oriented dialogue systems in the Persian Language. Since there is no published available dataset for Persian online shopping dialogue systems, to tackle the lack of data, we propose two methods for generating training data: fully-simulated and semi-simulated method. We created a simulated dataset with a hybrid of rule-based and template-based generation methods and a semi-simulated dataset where the language generation part is done by a human to increase the quality of the dataset. Our experiments with the natural language understanding module show that a combination of the datasets can improve results. These dataset generation methods can apply in other domains for low-resource languages in task-oriented dialogue systems too to solve the cold start problem of datasets.**

***Index Terms*—task-oriented dialogue systems, natural language understanding, online shopping**

I. Introduction

Dialogue systems play a significant role in business and social life today. People communicate with chatbots through text or audio using natural language. Chatbots have shown tremendous potential in different areas such as entertainment, game, customer service, and everyday tasks. Popular personal assistants, such as Amazon Alexa, Apple Siri, Google Home, and Microsoft Cortana are designed to help users get routine tasks done easier. Although these systems can handle simple tasks such as reporting weather, playing music, setting alarm, etc., they are limited in conducting complex tasks such as a complete shopping experience or open-domain dialogue. In fact, handling multi-turn context-aware dialogue is one of the fundamental challenges in natural language processing and intelligent information retrieval, artificial intelligence in general [1].

Dialogue systems can be broadly divided into two categories: chatbots that are designed to handle free-form conversations, and task-oriented systems that aim to assist users to achieve specific tasks or goals (such as shopping in a market, finding products, booking a restaurant, buying tickets, etc.) in a conversation [1], [2]. Task-oriented dialogue systems, which are the focus of this paper, have been widely studied and applied in some domains such as booking restaurants and buying tickets [3]. However, other areas such as online shopping remained less studied until recent years [4], [5]. With the rapid growth of e-commerce, shopping assistant bots attracted huge attention as they could provide pleasant shopping experiences to users and benefits both customers and sellers.

In this paper, we present a Persian online shopping dialogue system for purchasing items through spoken conversation. To pursue this goal we needed to have a Natural Language Understanding (NLU) module trained by an adequate number of labeled shopping dialogues. However, there is no published available dataset in Persian online shopping. To tackle the lack of data, we propose methods for collecting simulated datasets. To generate the dataset, we crawled the Digikala website for more than 500 items in nine product categories along with their product features. Using these items, we designed a simulated user that tries to order some items through a conversation with a simulated agent. Applying this method, we generated 3000 simulated dialogues. We then produced a semi-simulated dialogue dataset, where the simulated dialogue acts, were manually translated to more diverse and well-structured utterances. We trained the NLU module using both datasets and showed that the combination of the two datasets can increase the NLU performance in a great manner.

The collected data from both simulation methods are used to train the NLU module. We use BERT [6] to create a named entity recognition(NER) component in online shopping and fine-tune it using ParsBERT [7].

The contributions of this paper are summarized as follows:

- To the best of our knowledge, this online shopping dialogue system is the first Persian dialogue system that has been developed in the context of online retail.
- We generate an annotated Persian dialogue dataset, which includes complete natural dialogues on ordering items from an online store.
- We propose a high-performance natural language understanding module based on neural network methods, using both simulated and semi-simulated datasets. We learned that adding a small number of more natural training data would significantly increase the system's accuracy.
- We present a shopping-specific named entity recognizer by fine-tuning the BERT method.

II. Related works

Deep learning models are now widely used in the NLU module. These deep learning methods are used to handle NLU sub-tasks including domain detection, intent identification, and slot filling. Among previous works, we can mention [8] in which LSTM and RNN neural networks have been used for the classification of the user's utterances. Also, due to the special nature of texts on social networks such as Twitter which are very short and brief, [9] uses a combination of CNN and RNN to extract the maximum knowledge from the utterances. In general, applying deep learning methods in the NLU component showed higher performance comparing to previous rule-based methods, generation methods, and discriminative methods [1]. Recently, BERT has been fine-tuned in NLU and showed very good performances in many applications [10], [11]. In this work, we use BERT for the Persian language.

Previous works in task-based dialogue systems mainly focused on movie reservation, ticket booking, taxi ordering, hotel reservations, etc., while there are few works related to shopping domain [1], [3].

[4] presented a task-oriented dialogue system for e-commerce with a method for leveraging existing data sources. They collected customers' search queries and relevant forums questions and proposed a method to utilize them in dialogue to face the cold-start problem. They also defined proposed specific methods mostly based on deep learning techniques to implement an e-commerce chatbot. [12] has also introduced two multimodal neural models in the encode-attend-decode paradigm and demonstrate their performance on two subtasks, namely text response generation and best image response selection. [13] defined the real-world problem of query tracking e-commerce conversational search, in which the goal is to update the internal query after each round of interaction. The authors proposed a self attention-based neural network to handle the task from a machine comprehension perspective. Furthermore, they built a novel e-commerce query tracking dataset from an operational e-commerce search engine.

Some other works used retrieval methods for dialogue response generation. For instance, [14] focuses on retrieval-based response matching for multi-turn conversations. They formulate existing utterances into context using a proposed deep utterance aggregation model to form a fine-grained context representation. More specifically, self-matching attention is first introduced to route the core information in each utterance. Then the model matches a response with each refined utterance and the final matching score is obtained after an attentive turns aggregation.

[5] proposed a deep multi-task sequence labeling model to tackle slot filling in e-commerce. The proposed method used a multi-task model with cascade and residual connections, which jointly learns segment tagging, named entity tagging, and slot filling. The method achieves results comparable to the state-of-the-art models.

We also studied relevant works in Persian dialogue systems. In [15] a Persian dialog system is designed and focused on dialogue management module and uses a voice spoken dataset for training. [16] proposed a ticket reservation dialog system that focused on the dialogue management module. None of them used deep learning methods though. As we had no baseline available published in the Persian dialogue system in online shopping, we have proposed a new model based on BERT to understand natural language in the dialogue system.

III. Approach

We proposed two dialogue generation methods that can substitute for the existing data gathering methods. Then, we applied a state-of-the-art pre-trained language model based on a BERT [6] model pre-trained for Persian called ParsBERT [7] and fine-tuned on our data that results in a high-quality model demonstrating significant performance in our task.

A. Dataset generation

We have crawled DigiKala, a well-known online web-based shopping store in Iran. We extracted information for more than 500 items from nine selected supermarket product categories along with the items' attributes. We transformed the raw crawled data into a structured format. We then created 3600 annotated dialogues based on these products.

In order to automatically generate a simulated dataset, there are at least three approaches: rule-based, template-based, and hybrid methods [17]. Rule-based methods use predefined rules and utterances so they are flexible and programmable, though sometimes lacking fluency. Moreover, generating a dataset using this method is expensive both in terms of time and money. Template-based methods utilize templates that should be filled with data attributes. The method usually generates not only a more natural and fluent dataset but also a much more diverse one. However, it is still hard to contain the complexity of human language models. For example, two phrases "Please send me three packs of potato chips" and "Send me three packs of potato chips, please" are the same in meaning, but this method is not capable of handling such information as the templates are constant in utterances. Using hybrid methods, we try to attain the benefits of both methods by providing flexible, natural, and diverse utterances. This is done by providing sub-utterances and combine them to construct longer completed

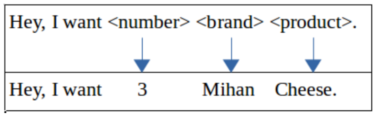

Fig. 1. Template filling example (translated)

TABLE I
DOMAINS AND SLOTS IN PROPOSED DIALOGUE SYSTEM

Categories	Slots
Cheese	Brand, Type, Size, Shape, Fat, Weight
Cream	Brand, Flavor, Size, Fat, Weight
Honey	Brand, Type, Size, Weight
Jam	Brand, Type, Size, Weight
Butter	Brand, Flavor, Size, Fat, Weight
Cereal	Brand, Flavor, Size, Filling, Weight
Halva	Brand, Type, Size, Weight
PeanutButter	Brand, Flavor, Size, Type, Weight
CocoaSpread	Brand, Size, Type, Weight

sentences. The method presented here is based on a hybrid approach.

1) Simulated dataset: We simulated an annotated dialogue dataset utilizing a hybrid approach. Utterances in dialogue systems are encoded in semantic frames, then the dialogue management module is responsible to map user acts to system ones, and a natural system utterance is generated using the decided system act. So, we need to focus on the dialogue management module and consider semantic frames combinations that take place in natural dialogues. To generate such dialogues, we only need to suggest product attributes as slots and utterance intents. As we consider product categories as domains, each product is followed by its category as an utterance domain. We generate such utterance acts in a sequence to form dialogue acts. Then we have assigned a random template to each utterance act and fill the template utterance by the act slot values. Note that each template needs to be filled with some slot parameters like brand, size, shape, etc., so the job is done by filling the parameters placeholders; for example a sentence ”Hey, I want (number) (product)” is considered as a user utterance template. When the act is *inform(product=cheese, number=3)*, the filled utterance would be ”Hey, I want 3 cheese” (In contrast to English, using singular form after a number is grammatically correct). This example is shown in figure 1. Categories and slots details are also shown in Table **??**, we used them as parameters in dialogue acts.

The organization of the utterances in a dialogue follow predefined rules, and the templates are considered independent. Finally, we have selected random products for each dialogue and generated annotated dialogues as mentioned above for about 3000 samples. A generated dialogue sample and its corresponding labels are presented in Table II. We use 133 utterance templates in total categorized in different intent-slot pairs. The average number of templates for each pair is about three. We also divide utterances into customers' and sellers' utterances, since some utterances can only be used by one group of speakers in dialogue. For example, ”Thank you for your purchase” is an utterance that is only used by salespersons. Basic intents include:

- request: It indicates the utterance looking for a specific attribute. The attributes can be the product brand, type, shape, or anything else.
- inform: This refers to any useful information that may either contain the product attribute values or product domain.
- greet: This refers to show semi-sentences using at the beginning of dialogues to greet and attract users' attention to contact.
- bye: same as above, but using for farewell.
- confirm: It indicates the final approval for the decided product(s). The customer may accept or reject the proposed product. If the product is rejected, the intent would be ”negation” without any parameter.
- thank: It shows appreciation to the seller or customer.

We produce the utterance acts by changing the parameters which result in different utterances for the randomly selected products.

2) Semi-simulated dataset: The semi-simulated method was designed to improve the data quality compared to the original simulated data. Similar to the simulated structure, we use a rule-based framework to generate the main dialogue flow. In order to increase the clarity, system “request” acts (asking about slot values from users) are always followed by “select“ acts, presenting all the available options. Moreover, we increase the diversity and naturalness of the utterances by generating them manually. Each user dialogue act is translated to one of many available natural language templates. We also provide complete and clear sentences instead of short ambiguous ones we used in the fully simulated method. A short sample dialogue can be seen in Figure 2. As can be seen, the generated dialogues are less like natural human-human conversations, but more like a human-machine dialogue which is optimized for reducing ambiguity for the user and contain diverse user language. As we will see in the next section, adding a small number of semi-simulated dialogues to the training data can significantly improve the model performance.

B. Online shopping dialogue system

A task-oriented dialogue system usually consists of four main components: natural language understanding (NLU), dialogue state tracking (DST), dialogue policymaking, and natural language generation (NLG), as shown in Figure 3. The natural language understanding module receives a user's utterance as input and interprets it into a semantic frame which in turn gets handed over to a DST module. DST is responsible for updating the current dialogue state and maintaining all needed information from the conversation history. Based on the current state, the dialogue policy module decides what the next dialogue action should be. Finally, the NLG module translates the dialogue act into a natural language response expressed by the system [1], [3].

The natural language understanding module is a vital component in dialogue systems. This module is responsible for

TABLE II
SIMULATED DIALOGUE EXAMPLE

Utterances (original)	Utterances (translated)	semantic frame
+ سلام علیکم	+Hello	greet()
-سلام	-Hi	greet()
+پنیر داری؟	+Do you have any cheese?	inform(product=cheese)
-دومینو داریم. همینو بذارم؟	-We have Domino. Do you want it?	request(brand=?), select(brand=Domino)
+بله	+Yes please.	inform(brand=Domino)
-نوع صبحانه باشه؟	-Do you want breakfast type?	request(type=?)
+بله	+Yes.	inform(type=breakfast)
-چه شکلی باشه؟	-What kind of shape do you like?	request(shape=?)
+قالبی لطفا	+Solid please.	inform(shape=solid)
-کم‌چرب و پرچرب داریم. کدومشو میخوای؟	-We have middle-fat and full-fat. which one?	request(fat=?), select(fat=middle-fat, fat=full-fat)
+فرقی نداره.	+I don't care.	inform(fat=don't care)
-بزرگ باشه یا متوسط؟	-Do you want big or medium one?	request(size=?), select(size=big, size=medium)
+متوسط بهتره.	+Medium is better.	inform(size=medium)
-چند تا بدم؟	-How many I should take?	request(number=?)
+یکی لطفا	+one please.	inform(number=1)
-آیا کالای پنیر با شناسه ۱۵۵۷۷ را تایید می‌کنید؟	-Do you accept one cheese by id 15577?	confirm(product=cheese, number=one, id=15577)
+بله	+Yes.	confirm()
-سفارش دیگری دارید؟	-Any other order?	reqmore()
+ ممنونم	+No, thanks.	thank()
-خدانگهدار	-Ok, bye.	bye()

extracting the meaning from the input utterance [2]. In a task-oriented context, natural language understanding parses input utterances into predefined semantic slots. To achieve this goal, an NLU system typically consists of three subtasks including domain detection, intent identification, and slot filling [18].

We propose a joint framework for intent learning, domain identification, and slot filling. We use a BERT pre-trained language model as one of the state-of-the-art methods in natural language understanding. In fact, BERT has shown the highest accuracy in named entity recognition in the Persian language [19]. Our online shopping dialogue system is in Persian and used ParsBERT for fine-tuning specific online shopping named entity recognition.

In order to train the model, we feed the neural network with the following components: 1- input user utterance as a sequence of words, 2- IOB tags of all words, 3- the intent of the sentence, 4- the domain of sentence, and 5- a context array containing the last three dialogue turns. An [CLS] as the first token and an [SEP] as the last token is added into each utterance. The popular IOB (in-out-begin) [2] format is used for annotating the slot tags, as shown in Figure4.

We developed the dialogue state management module based on a rule-based method. This module takes as input a semantic frame and outputs the state of the dialogue, including the product features, the user has already talked about. Then the dialogue policy module takes in the current state, searches online e-commerce databases for the domain, intents, and slot tags and values that are determined the state, and generates the appropriate system act, using finite state automata. The natural language generation module translated the system acts into natural language utterances, using a template-based method along with the options employed in the semi-simulated data generation in section III.

IV. EXPERIMENTS

We trained the proposed NLU model both on the simulated dataset and on a combination of the simulated and semi-simulated datasets. For fine-tuning the proposed NLU model, we used ParsBert. Hyperparameters are fine-tuned on the development set. The maximum length of the input sequence is 40 and the batch size is 128. Adam [20] is used for optimization with an initial learning rate of 1e-4 and a dropout probability of 0.1.

We generated 1000, 2000, and 3000 dialogues using the simulated data generation method and trained the proposed NLU method on each of the datasets, the results are shown in table III. As can be seen, increasing the number of simulated conversations does not necessarily increase precision. This can be justified by the fact that this dataset consists of short user and system utterances, so adding a huge number of dialogues have little effect on raising the word embedding quality. We select a dataset with 3000 dialogue.

We then added a dataset of 100 dialogues generated by the semi-simulated method, which as mentioned contained longer, more informative, and more well-structured utterances. As the volume of the dataset is low, we insert six copies of data (600 dialogues) into the 3000 dialogues that came from the simulated dataset and shuffle all dialogues. The resulting dataset has 3600 conversations with 117 intent labels and 262 slot tags in nine product domains and a greeting domain. We used 70 percent of the dialogues for training,10 percent for validation, and 20 percent for testing. As it is seen in Table IV, training the proposed NLU model on the merged dataset results in 5.4 percents increased in F1 measure

V. CONCLUSION

We developed an online shopping dialogue system to assist users in shopping and improve their experience. This system aids users in searching, selecting, and buying products

Utterances (translated in English and original in persian)	domain	semantic frame
user: Hi! کاربر: سلام	general	greet()
system: Hello, welcome to our shopping assistant system! How are you? سیستم: سلام! به سامانه خرید ما خوش آمدید. چطور می‌تونم کمکتون کنم؟	general	greet()
user: I'm looking for Kalleh cheese. کاربر: پنیر کاله میخواستم	cheese	inform(brand=Kalleh)
system: Sure. Do you have any specific size of cheese in mind? We have small, average and large. سیستم: حتما. اندازه خاصی مدنظرتون هست؟ کوچک، متوسط و بزرگ داریم.	cheese	request(size=?),select(size=[small,average,large])
user: Well. . . I would go for small. کاربر:کوچیک باشه لطفا	cheese	inform(size=small)
system: What about the fat level? We have these options: average and full fat سیستم: چربیش چطور؟ این گزینه‌ها رو داریم: متوسط و پرچرب	cheese	request(fat=?),select(fat=[average,full-fat])
Average is fine. کاربر: متوسط خوبه	cheese	inform(fat=average)
Found it! How many of that do you need? سیستم: محصول مورد نظر شما رو یافتم! چند بسته لازم دارید؟	cheese	inform(count=٢)
!user: Two boxes please کاربر: دو بسته لطفا	cheese	
system: Alright! Do you confirm to add this product to your shopping bag? Two boxes of small average fat Kalleh cheese with product id: ١٢٣٤٥٦ سیستم: بسیار خوب. آیا تایید می‌کنید که این محصول به سبد خرید شما اضافه بشه؟ دو بسته کوچک پنیر کاله با چربی متوسط به شماره محصول ۱۲۳۴۵۶	order	confirm(count=٢,id=١٢٣٤٥٦)
user: Yes, please! کاربر: بله لطفا	order	confirm
system: Great! is there anything else I can help you with? سیستم: به روی چشم. چیز دیگه‌ای هم نیاز دارید کمکتون کنم؟	general	reqmore()
user: No, thank you! کاربر: نه ممنون	general	Thanks()
Then have a good day! general bye() سیستم: بسیار عالی. روز خوبی داشته باشید.	general	bye()

Fig. 2. Semi-simulated dialogue example

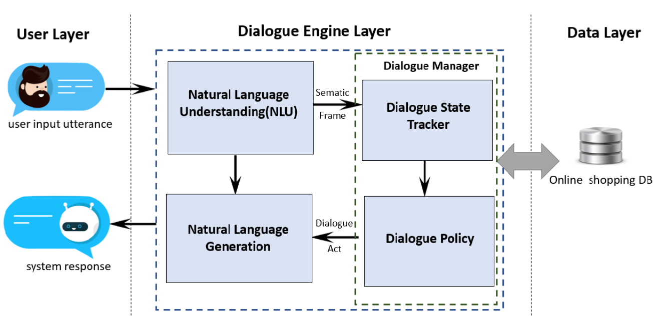

Fig. 3. modular task-oriented dialogue systems [1]–[3]

W	I	want	two	Kalleh	full	fat	white	cheese.
S	O	O	B-number	B-brand	B-fat	I-fat	B-type	O
I	cheese-inform							

Fig. 4. IOB-tagging example

TABLE III
COMPARISON OF RESULTS IN SIMULATED DATASET

Dialogue number of train size	precision	recall	F1
1000	*80/2*	*40/3*	*53/64*
2000	*83/36*	*42/10*	*55/95*
3000	*86/71*	*52/5*	*65/20*

TABLE IV
COMPARISON OF RESULTS IN SIMULATED AND MERGED DATASET

Method	precision	recall	F1
NLU method on simulated dataset	*86/71*	*52/5*	*65/20*
NLU method on merged dataset	*94/47*	*56/38*	*70/62*

from large online retailers in a conversational manner. We proposed a joint learning natural language understanding with a shopping-specific named entity recognizer based on BERT. This system is in Persian that is a low-resource language with no published available dataset in online shopping. We proposed two dataset generation methods including simulated dataset and semi-simulated dataset generation. Our suggested dataset generation methods can apply in many domains in a task-oriented dialogue system for solving cold start problem in low-resource language.

REFERENCES

[1] J. Gao, M. Galley, and L. Li, *Neural Approaches to Conversational AI: Question Answering, Task-oriented Dialogues and Social Chatbots*. Now Foundations and Trends, 2019.

[2] D. Jurafsky and J. H. Martin, *Speech and Language Processing An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 3th edition,draft version, 2019.

[3] H. Chen, X. Liu, D. Yin, and J. Tang, "A survey on dialogue systems: Recent advances and new frontiers," *Acm Sigkdd Explorations Newsletter*, vol. 19, no. 2, pp. 25–35, 2017.

[4] Z. Yan, N. Duan, P. Chen, M. Zhou, J. Zhou, and Z. Li, "Building task-oriented dialogue systems for online shopping," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, 2017, pp. 4618–4625.

[5] Y. Gong, X. Luo, Y. Zhu, W. Ou, Z. Li, M. Zhu, K. Q. Zhu, L. Duan, and X. Chen, "Deep cascade multi-task learning for slot filling in online shopping assistant," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, 2019, pp. 6465–6472.

[6] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *NAACL-HLT (1)*, 2019.

[7] M. Farahani, M. Gharachorloo, M. Farahani, and M. Manthouri, "Parsbert: Transformer-based model for persian language understanding," *arXiv preprint arXiv:2005.12515*, 2020.

[8] S. Ravuri and A. Stolcke, "Recurrent neural network and lstm models for lexical utterance classification," in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[9] J. Y. Lee and F. Dernoncourt, "Sequential short-text classification with recurrent and convolutional neural networks," *arXiv preprint arXiv:1603.03827*, 2016.

[10] Q. Chen, Z. Zhuo, and W. Wang, "Bert for joint intent classification and slot filling," *arXiv preprint arXiv:1902.10909*, 2019.

[11] X. Jiao, Y. Yin, L. Shang, X. Jiang, X. Chen, L. Li, F. Wang, and Q. Liu, "Tinybert: Distilling bert for natural language understanding," *arXiv preprint arXiv:1909.10351*, 2019.

[12] A. Saha, M. Khapra, and K. Sankaranarayanan, "Towards building large scale multimodal domain-aware conversation systems," *arXiv preprint arXiv:1704.00200*, 2017.

[13] Y. Yang, Y. Gong, and X. Chen, "Query tracking for e-commerce conversational search: A machine comprehension perspective," in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 2018, pp. 1755–1758.

[14] Z. Zhang, J. Li, P. Zhu, H. Zhao, and G. Liu, "Modeling multi-turn conversation with deep utterance aggregation," in *Proceedings of the 27th International Conference on Computational Linguistics*, 2018, pp. 3740–3752.

[15] H. Mahmoudi and M. M. Homayounpour, "A persian spoken dialogue system using pomdps," in *2015 The International Symposium on Artificial Intelligence and Signal Processing (AISP)*. IEEE, 2015, pp. 217–221.

[16] M. Habibi, H. Sameti, and H. Setareh, "On-line learning of a persian spoken dialogue system using real training data," in *10th International Conference on Information Science, Signal Processing and their Applications (ISSPA 2010)*. IEEE, 2010, pp. 133–136.

[17] A. H. Oh and A. I. Rudnicky, "Stochastic language generation for spoken dialogue systems," in *ANLP-NAACL 2000 Workshop: Conversational Systems*, 2000. [Online]. Available: https://www.aclweb.org/anthology/W00-0306

[18] R. De Mori, F. Bechet, D. Hakkani-Tur, M. McTear, G. Riccardi, and G. Tur, "Spoken language understanding," *IEEE Signal Processing Magazine*, vol. 25, no. 3, pp. 50–58, 2008.

[19] N. Taghizadeh, Z. Borhanifard, M. GolestaniPour, and H. Faili, "Nsurl-2019 task 7: Named entity recognition (ner) in farsi," *arXiv preprint arXiv:2003.09029*, 2020.

[20] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.