A Feature-Enriched Method for User Intent Classification by Leveraging Semantic Tag Expansion

Wenxiu Xie[1], Dongfa Gao[1(✉)], Ruoyao Ding[1], and Tianyong Hao[2(✉)]

[1] School of Information Science and Technology, Guangdong University of Foreign Studies, Guangzhou, China
vasiliky@outlook.com, gaodf@gdufs.edu.cn, ruoyaoding@outlook.com

[2] School of Computer Science, South China Normal University, Guangzhou, China
haoty@126.com

Abstract. User intent identification and classification has become a vital topic of query understanding in human-computer dialogue applications. The identification of users' intent is especially crucial for assisting system to understand users' queries so as to classify the queries accurately to improve users' satisfaction. Since the posted queries are usually short and lack of context, conventional methods heavily relying on query n-grams or other common features are not sufficient enough. This paper proposes a compact yet effective user intention classification method named as ST-UIC based on a constructed semantic tag repository. The method proposes to use a combination of four kinds of features including characters, non-key-noun part-of-speech tags, target words, and semantic tags. The experiments are based on a widely applied dataset provided by the First Evaluation of Chinese Human-Computer Dialogue Technology. The result shows that the method achieved a F1 score of 0.945, exceeding a list of baseline methods and demonstrating its effectiveness in user intent classification.

Keywords: User intent · Classification · Target Words · Semantic Tag

1 Introduction

The research of human-computer dialogue has become a hot topic in both academia and industry in recent year [1]. A spoken dialogue system enables users to access information over the Internet using spoken languages as the medium of interaction [2]. As a vital component of spoken dialog systems, Spoken Language Understanding (SLU) aims to identify the domain and intent of users as expressed in natural language speech automatically. SLU is widely available and commonly used in a variety of application areas, such as mobile phone-based personal assistants like Siri [3, 4]. SLU typically involves three steps: domain classification, user intent determination and semantic tagging [5]. In such process, classifying user intents of input queries into a specific domain is the first and vital step of query semantic analysis in SLU [1, 3, 6, 7].

M. Zhang et al. (Eds.): NLPCC 2018, LNAI 11109, pp. 224–234, 2018.
https://doi.org/10.1007/978-3-319-99501-4_19

Since user intents are usually domain specific, predicting the labels of intents and domains can be treated as a single classification problem.

Detecting and classifying user intents is a growing research area not only in SLU but also in the research of search engines [8–11]. Besides, understanding the intent behind a user's query may help prune out irrelevant information and personalize answers, thus improving user satisfaction [8, 9, 12]. For both SLU and Web search engines, the most challenge of user intent detection and classification is how to understand the short, lacking of context, and noise-contained queries posted by users [5]. According to the research in [13], 93.15% of user queries for search engines are no more than 3 words and the average length is 1.85 words only. Furthermore, spoken sentences of queries in SLU usually do not follow formal language grammar and exhibit self-corrections, hesitations, repetitions and other irregular phenomena [14]. Therefore, how to detect user intent accurately and expand the semantic features of a query to assist user intent classification is an essential step.

According to the existing research [15, 16], the target words of a query can potentially represent the user intent in a SLU process. To that end, this paper presents a Semantic Tag-empowered User Intent Classification method named as ST-UIC for user intent detection and classification. In addition, ST-UIC integrates 4 kinds of features: character, non-key-noun part-of-speech (POS) tags, target words, and semantic tags, named as CKTS. A strategy is proposed to identify and extract query target words. Moreover, a sematic tag repository is automatically constructed for feature expansion purpose. Based on a publically available dataset, the F1 score on testing dataset reaches 94.5% and outperforms five baseline methods, demonstrating the effectiveness of the proposed method in user intent classification.

2 Related Work

The user intent classification task defined by Zhang et al. [1] is that, given an input message, classify the user intent into a specific domain category. With respect to that user intent is semantic constraint on the sought-after answers, how to effectively identify user intents to prune out irrelevant information that may mislead the classification process is a crucial problem. As a vital task of spoken language understanding, there has been existing representative research with respect to user intent classification (also known as domain classification) in last decades. One pioneering work was AT&T's "How May I Help You?" [17], where the users' fluently spoken utterances were classified into a number of predefined intent categories. Later, a variety of practical goal-oriented spoken dialog systems [5, 6, 18] were built and the research of SLU ranges from determining phrases via grammars, extracting predefined named entities, detecting users' intents for classification, to combinations of users' intent and named entities [18]. With the increasing use of web search, user search query logs were applied as a valuable source of unlabeled information for user intent classification [10, 19, 20]. Hakkani-Tür et al. [19] exploited search queries from search engine query logs to improve query intent classification in SLU. They assumed that clicked URL categories could be assigned as the domain labels of user queries. For example, label "hotels" was assigned to the user query "Holiday Inn and Suites" when the user clicked the URL "http://www.hotels.com". Celikyilmaz et al. [20]

utilized unlabeled queries collected from internet search engine click logs for SLU intent detection. Those ideas of utilizing query click logs for improving query intent classification were similar to the proposed semantic tags expansion in the paper. The difference was that the domain labels were sentence-level expansion of the query and the proposed method was word-level.

Besides, most of the previous research in SLU applied manually generated features, such as grammar information and predefined phrases. Gupta et al. [18] presented a SLU system which utilized a statistical classifier for intent determination and a rule-based fixed grammars for named entity extraction. Hernández et al. [21] proposed a simple model for user intent classification which leveraging only the text including in the query. The feature they extracted from the query text, including entity names, query length, transactional terms, interrogative terms and stop words, was verified as simple yet effective. In Ganti et al. [22] research, co-occurrence between the query keyword and the tags that associated with the retrieved search results were leveraged as tag ratio features for avoiding the sparse feature spaces issue in query intent classification task.

Deep learning [3, 7, 23, 24] recently achieved good performance on user intent classification task in SLU, due to their ability to learn compact and discriminative features. Experiments presented that proposed methods improved performances over baseline techniques on a large context-sensitive SLU dataset. Tur et al. [3] presented an application of deep convex networks (DCNs) for semantic utterance classification. The results showed that the proposed DCN-based method was effective on a domain classification task for spoken language understanding. Deng et al. [23] also applied deep learning techniques kernel version DCN (K-DCN) on a intent classification task of SLU, which yielded a good classification performance. Although the study of deep learning has already led to impressive theoretical results, several problems lie ahead such as requiring large training datasets, time-consuming for parameter tuning, and involving a difficult optimization [25].

3 The User Intent Classification Method

A Semantic Tag-empowered User Intent Classification (ST-UIC) method is proposed to identify user intent and expand semantic features for user intent classification. The ST-UIC method contains five major steps: preprocessing, target word extraction, non-key-noun POS expansion, semantic tag expansion, and intent classification. The framework of the method is shown as Fig. 1.

The preprocessing step includes query segmentation and transformation. All numerics in Chinese character are also converted into Arabic format. Then, a dependency relation-based strategy is proposed for target word extraction. After preprocessing, four proposed features are expanded for classification, i.e. character, non-key-noun part-of-speech (non-key-noun POS) tags, target words, and semantic tags, named as CKTS. Both Character and target word features are extracted for maintaining contextual information and representing the user intent of original queries. For the queries that none target words are identified, key noun words are extracted instead. Then, key noun words are further expanded with semantic tags to enrich the semantic information of the query. The non-key-noun POS feature is proposed as a supplementary strategy to enrich the

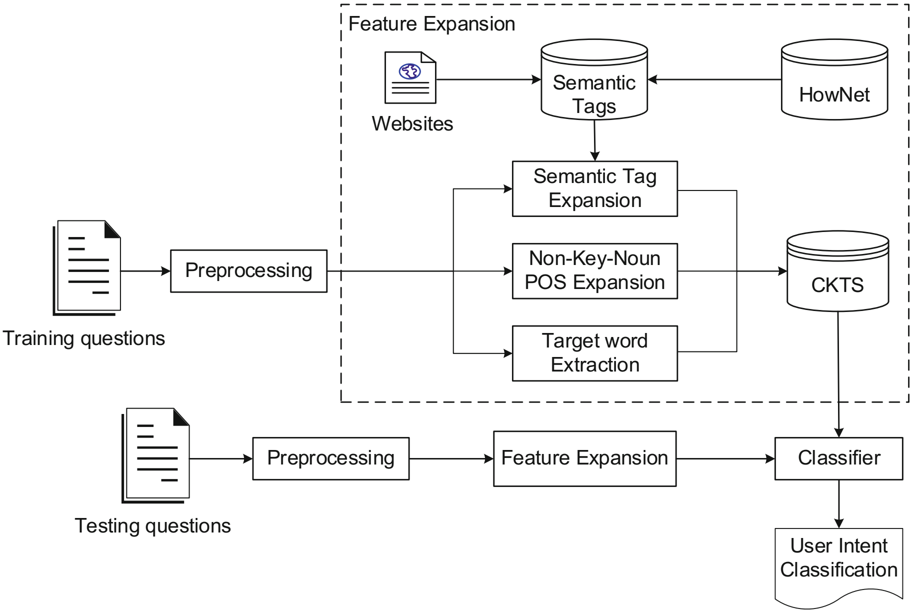

Fig. 1. The framework of the proposed ST-UIC method for user intent classification.

syntactic features of non-key-noun words. Finally, the expanded features are sent to a trained classifier to obtain user intent categories.

3.1 Target Word Extraction

Based on an existing research [15], the target words of a query substantially represent the intent of corresponding asker. For instance, the target word of the query "帮我链接到新浪网" ('*help me access to the Sina homepage*') is "新浪网" ('*the Sina homepage*'). This target word represents user intent to some extent. Through expanding the semantic tags "<网站\门户网站>"('*website*') of the target word "新浪网", the query can be directly linked to the intent category "website". As the queries come from a human-computer chit-chat and task-oriented dialogue, the queries are always short and verb-centered. For example, the same query contains a verb "链接"('*access*') and an object "新浪网", and these two words can well represent the intent of the query.

By leveraging Chinese Language Technology Platform (LTP) for dependency relation analysis, we design a strategy for extracting target words from queries. After analyzing the dependency relations of queries, we found that the target word of the verb-center query frequently contained in a "*VOB*" relation. We thus further represented the dependency relations of a query as "$word_1$←*pos: relation: pos*→$word_2$". For all the dependency relations of a query, we extract the $word_2$ from the relation that is "*VOB*". Therefore, "新浪网"is extracted as target word of this query.

In addition to the typical relations, there are also some other cases that are not verb-centered, e.g., the query “颈椎病用什么药物治疗?” (‘*What is the medication treatment of cervical spondylosis?*’). Therefore, a key noun word extraction strategy is further proposed to deal with the non-verb-center queries for further feature expansion. We applied jieba, a Chinese text segmentation tool, for keyword noun words extraction. We use jieba TextRank keyword extraction function with a constraint that extract words with which POSs are included. After analyzing the queries, the POSs we set are {N(noun), NS(toponym), NT(organization/group name), NZ(other proper noun), NL (noun phrase)}. For the same example, the key noun words extracted of the query are “颈椎病”(‘*cervical spondylosis*’) and “药物”(‘*medication*’). Then, the extracted key noun words set KN_w will be expanded with semantic or syntactic features.

3.2 Semantic Tag Expansion

The input queries are always short and lack of context. We thus propose to expand the semantic information of queries. HowNet, a bilingual general knowledge base that describing relations between concepts as well as relations between concept attributes, is utilized for key noun words’ semantic expansion in this process. Each word in a HowNet taxonomy has an upper concepts in word’s definition. For instance, the upper concepts of word “糖尿病”(‘*diabetes*’) in query “糖尿病注意什么?”(‘*what does diabetes need to pay attention to*?’) is “disease|疾病”. Therefore, after upper concepts expansion, the query can directly link to the category “health”. We further reconstruct the HowNet corpus into a repository only containing words which POS is “Noun” and the corresponding upper concepts as semantic tags. Yet, HowNet is not large enough to covering all word concepts especially new words, such as “新浪网”and “颈椎病”. Therefore, a new semantic tag repository is constructed. Words and corresponding semantic tags are extracted from the websites such as Hudong Wikipedia. For instance, word “新浪网”and corresponding semantic tags “〈中国网站\互联网〉”(‘*Internet*’) are extracted from the websites. All the words and semantic tags are structured as a tuple, i.e. (word,<semantic tags>). The newly constructed repository contains 255,824 words and their associated semantic tags. For all key noun words extracted from the process, ST-UIC match words in the repository to retrieve matched semantic tags as features.

$$f(w) = \begin{cases} semantic\,tag, & if\ w\ in\ KN_w \\ POS\,tag, & otherwise \end{cases} \tag{1}$$

Yet, not all queries are containing target words or key noun words. For example, queries of chit-chat category such as “你在干啥呢” (‘*what are you doing*’) may not contain nouns. For this circumstance, the non-key-noun POS feature is proposed as a supplementary strategy to enrich the syntactic features. For those that are not key noun words, the corresponding POS tags are expanded as features. Therefore, the feature expansion function $f(w)$ is as Eq. (1), here w is the word in key noun word set KN_w extracted by previous process.

4 Evaluation and Results

A standard dataset provided by iFLYTEK Corporation and the First Evaluation of Chinese Human-Computer Dialogue Technology (ECDT) [1] is used. All the data are mapped to a taxonomy containing 2 coarse grained categories and 31 fine grained categories. In this paper, the original train and develop data are used as **Training dataset A**, containing 3068 labeled queries. The original test data is used as **Testing dataset A**, containing 667 labeled queries.

To evaluate the stability of the proposed method, five experiments are conducted. The evaluation measures are the four widely used statistical classification measures: Accuracy, Precision, Recall and F1 score (F1).

The first experiment is to test the effectiveness of the proposed feature and to find out the optimal combination of features for the user intent classification. The proposed feature are Target Words (TWs), Semantic Tags (STs) and non-Key-noun POS tags (nK-POS). In the contrast, the commonly used feature Characters (C), Segment words (SWs) and Part-of-Speech tags (POS) are adopted as baseline features. Training on the Training dataset A with the same typical Logistic Regression (LR) classifier, the performance on the Testing dataset A are calculated and presented in Table 1. From the results, Comparing to use C or SWs alone, the F1 score of using C + SWs increased from 0.864 and 0.904 to 0.918. However, when adding SWs to the best result #11, the performances on all metrics decrease. The results of #1 and #2 indicate that the character feature is more useful and contains more semantic information than the segment word feature. As shown in the result #10, #11, #12 and #13, #14, #15, the proposed non-key-noun POS produce better performance than the commonly used POS. And comparing the result #7, #8, #9 with #13, #14, #15, adding the POS feature, the performance turn to be decreased. The proposed combination features C + nK-POS + TWs + STs (CKTS)

Table 1. The performance comparison of using different features on user intent classification.

#	Features	Accuracy	Precision	Recall	F1
1	SWs	0.855	0.932	0.826	0.864
2	C	0.888	0.934	0.887	0.904
3	C + SWs	0.910	0.945	0.900	0.918
4	C + nK-POS	0.901	0.936	0.893	0.911
5	SWs + nK-POS	0.870	0.932	0.846	0.878
6	C + SWs + nK-POS	0.904	0.939	0.895	0.913
7	SWs + TWs + STs	0.901	0.942	0.877	0.901
8	C + TWs + STs	0.907	0.946	0.904	0.920
9	C + SWs + TWs + STs	0.930	0.961	0.917	0.936
10	SWs + nK-POS + TWs + STs	0.897	0.937	0.880	0.902
11	C + nK-POS + TWs + STs	**0.936**	**0.962**	**0.932**	**0.945**
12	C + SWs + nK-POS + TWs + STs	0.931	0.958	0.921	0.937
13	SWs + POS + TWs + STs	0.892	0.931	0.865	0.890
14	C + POS + TWs + STs	0.919	0.956	0.903	0.926
15	C + SWs + POS + TWs + STs	0.918	0.955	0.895	0.919

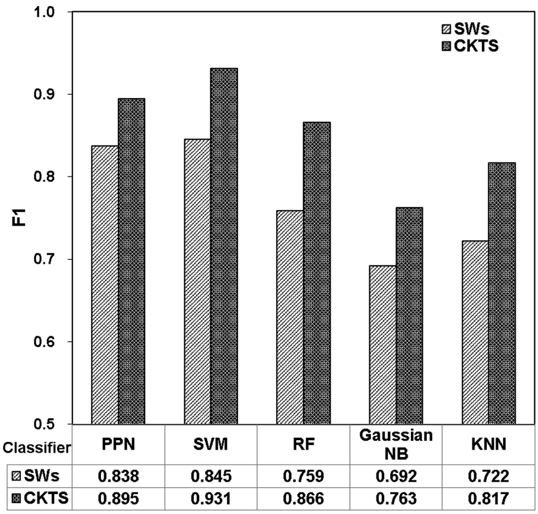

Fig. 2. The performance comparison 5 different classifiers on fine-grained classification using SWs and CKTS.

outperform the other types of feature combinations, demonstrating the effectiveness of proposed feature on user intent classification.

To better evaluate the robustness of proposed feature combination, both SWs and CKTS are applied to 5 commonly used classifiers including Support Vector Machine (SVM), Perceptron (PPN), Random Forest (RF), Gaussian Naive Bayes (GaussianNB) and *k*-Nearest Neighbor (KNN). The experiment results on fine-grained classification are shown in Fig. 2, where the proposed CKTS contributes every classifier than SWs, demonstrating the usefulness of the proposed features on user intent classification task.

In the third experiment, the stability of the proposed method is tested with different sizes of training data. The training dataset A is randomly divided into 5 training subsets containing 600, 1200, 1800, 2400, and 3000 queries respectively. The results are measured in accuracy, precision, recall and F1. As illustrated in Fig. 3, our method receives a stable performance on all evaluation metrics. Moreover, when only 600 queries are used, which are less than testing queries (667), the F1 of user intent classification still achieves 0.854, only 0.01 less than the baseline 0.864 in Table 2. This also verifies the effectiveness of proposed method.

To verify the strength and weakness of proposed method on different categories, we conduct the fourth experiment. The results, as shown in Table 2, presented that the proposed method achieves a precision of 1.000 on 20 of 31 fine-grained categories. Moreover, for the categories such as “bus”, “calc”, “contacts” etc., the proposed method gain the precision and F1 of 1.000 as well. For the categories “app”, “music”, “epg”, and “radio”, the proposed method obtain lower recall of 0.667, 0.864, 0.806, and 0.875.

The last experiment presents the comparisons of our method with existing user intent classification methods as baselines. The baselines are top five methods won the challenge in the shared task organized by the official ECDT website[1]. Using the same training and testing datasets, we compare the performance on all the fine-grained

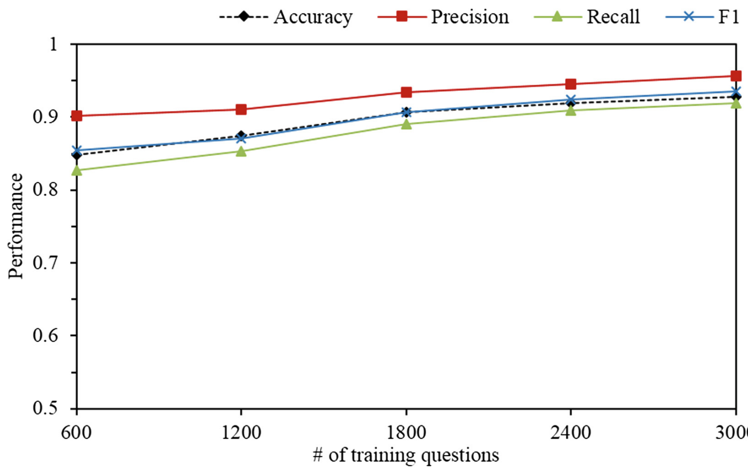

Fig. 3. The performance of our method with the increasing size of training datasets.

Table 2. The performance on fine grained categories of proposed method

Fine-grained categories	Precision	Recall	F1	Fine-grained categories	Precision	Recall	F1
App	0.857	0.667	0.750	Music	0.905	0.864	0.884
Bus	1.000	1.000	1.000	News	1.000	1.000	1.000
Calc	1.000	1.000	1.000	Novel	1.000	0.875	0.933
Chat	0.847	0.980	0.909	Poetry	1.000	0.882	0.938
Cinemas	0.778	0.875	0.824	Radio	1.000	0.875	0.933
Contacts	1.000	1.000	1.000	Riddle	1.000	1.000	1.000
Cookbook	0.978	0.989	0.983	Schedule	1.000	0.900	0.947
Datetime	1.000	0.833	0.909	Stock	1.000	0.875	0.933
Email	1.000	1.000	1.000	Telephone	0.952	0.952	0.952
Epg	0.967	0.806	0.879	Train	1.000	1.000	1.000
Flight	1.000	0.952	0.976	Translation	1.000	1.000	1.000
Health	1.000	0.944	0.971	TVchannel	0.885	0.958	0.920
Lottery	1.000	1.000	1.000	Video	0.750	0.934	0.832
Map	0.917	0.957	0.936	Weather	1.000	1.000	1.000
Match	1.000	1.000	1.000	Website	1.000	0.778	0.875
Message	1.000	1.000	1.000				

1 http://ir.hit.edu.cn/SMP2017-ECDT-RANK.

Table 3. The performance comparison of the proposed method with 5 baselines.

Methods	Developed by	F1
LSTM + domain dictionary	Spoken Dialogue System Lab, SCAU (SIGSDS)	0.941
CNN + Ensemble Learning	DeepBrain Corporation	0.929
CNN + rules	Institute of automation, Chinese Academy of Sciences	0.926
Lib-SVM + unigram + bigram + trigram + 4-gram	School of Computer & Information Technology, Shanxi University	0.912
CNN + domain dictionary	Boyan Information Technology Company Limited	0.899
LR + CKTS	ST-UIC	**0.945**

categories in F1. From the result, as presented in Table 3, our method achieves a best F1 of 0.945. As reported on ECDT website, SIGSDS and Whisper apply a deep learning method (Long Short-Term Memory, LSTM, and Convolutional Neural Network, CNN) with manually constructed domain dictionaries. Our method obtains a comparable performance by adopting a traditional Logistic Regression with feature expansion.

5 Conclusions

Aiming for user intent identification and classification, this paper proposed a method called ST-UIC based on dependency relation analysis for target word extraction. Moreover, a semantic tag repository, containing 255,824 words and corresponding semantic tags, was automatically constructed for feature expansion. Using a publicly available dataset, five experiments were conducted for evaluating the effectiveness of ST-UIC method through a comparison with five baseline methods. The results presented that ST-UIC achieved the best performance in the comparison, demonstrating its effectiveness for user intent classification tasks.

Acknowledgements. This work was supported by National Natural Science Foundation of China (No.61772146) and Innovative School Project in Higher Education of Guangdong Province (No.YQ2015062). Guangzhou Science Technology and Innovation Commission (No. 201803010063).

References

1. Zhang, W.-N., Chen, Z., Che, W., Hu, G., Liu, T.: The First Evaluation of Chinese Human-Computer Dialogue Technology. arXiv preprint arXiv:1709.10217 (2017)
2. Zue, V., Seneff, S.: Spoken dialogue systems. Synth. Lect. Hum. Lang. Technol. **2**, 1–151 (2009)

3. Tur, G., Deng, L., Hakkani-Tür, D., He, X.: Towards deeper understanding: deep convex networks for semantic utterance classification. In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 5045–5048 (2012)
4. Zhang, J., Yang, T.Z., Hazen, T.J.: Large-scale word representation features for improved spoken language understanding. In: International Conference on Acoustics, Speech and Signal Processing, pp. 5306–5310 (2015)
5. Liu, J., Pasupat, P., Wang, Y., Cyphers, S., Glass, J.: Query understanding enhanced by hierarchical parsing structures. In: IEEE Workshop on Automatic Speech Recognition and Understanding, pp. 72–77 (2013)
6. Liu, B., Lane, I.: Multi-Domain Adversarial Learning for Slot Filling in Spoken Language Understanding. arXiv preprint arXiv:1711.11310. pp. 1–6 (2017)
7. Xu, P., Sarikaya, R.: Contextual domain classification in spoken language understanding systems using recurrent neural network. In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 3–7 (2014)
8. Ashkan, A., Clarke, C.L.A., Agichtein, E., Guo, Q.: Classifying and characterizing query intent. In: Boughanem, M., Berrut, C., Mothe, J., Soule-Dupuy, C. (eds.) ECIR 2009. LNCS, vol. 5478, pp. 578–586. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-00958-7_53
9. Hu, J., Wang, G., Lochovsky, F., Sun, J., Chen, Z.: Understanding user's query intent with Wikipedia. In: International Conference on World Wide Web, pp. 471–480. ACM (2009)
10. Park, K., Jee, H., Lee, T., Jung, S., Lim, H.: Automatic extraction of user's search intention from web search logs. Multimed. Tools Appl. **61**, 145–162 (2012)
11. Jansen, B.J., Booth, D.L., Spink, A.: Determining the user intent of web search engine queries. In: International Conference on World Wide Web, pp. 1149–1150. ACM (2007)
12. Zhang, S., Wang, B.: A survey of web search query intention classification. J. Chin. Inf. Process. **22**(4), 75–82 (2008)
13. Yu, H., Liu, Y., Zhang, M., Ru, L., Ma, S.: Research in search engine user behavior based on log analysis. J. Chin. Inf. Process. **21**, 109–114 (2007)
14. De Mori, R., Béchet, F., Hakkani-Tür, D., McTear, M., Riccardi, G., Tur, G.: Spoken language understanding: a survey. In: Automatic Speech Recognition and Understanding Workshop, pp. 365–376 (2007)
15. Hao, T., Xie, W., Wu, Q., Weng, H., Qu, Y.: Leveraging question target word features through semantic relation expansion for answer type classification. Knowl. Based Syst. **133**, 43–52 (2017)
16. Hao, T., Xie, W., Xu, F.: A wordnet expansion-based approach for question targets identification and classification. In: Sun, M., Liu, Z., Zhang, M., Liu, Y. (eds.) CCL 2015. LNCS (LNAI), vol. 9427, pp. 333–344. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-25816-4_27
17. Gorin, A., Riccardi, G., Wright, J.: How may I help you? Speech Commun. **23**, 113–127 (1997)
18. Gupta, N., Tur, G., Hakkani-Tur, D., Bangalore, S., Riccardi, G., Gilbert, M.: The AT&T spoken language understanding system. IEEE Trans. Audio Speech Lang. Process. **14**(1), 213–222 (2006)
19. Hakkani-Tür, D., Heck, L., Tur, G.: Exploiting query click logs for utterance domain detection in spoken language understanding. In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 5636–5639 (2011)
20. Celikyilmaz, A., Hakkani-Tür, D., Tur, G.: Leveraging web query logs to learn user intent via bayesian discrete latent variable model. In: International Conference on Machine Learning (2011)

21. Hernández, I., Gupta, D., Rosso, P., Rocha, M.: A simple model for classifying web queries by user intent. In: Spanish Conference Information Retrieval, pp. 235–240 (2012)
22. Ganti, V., König, A.C., Li, X.: Precomputing search features for fast and accurate query classification. In: ACM International Conference on Web Search and Data Mining, pp. 61–70 (2010)
23. Deng, L., Tur, G., He, X., Hakkani-Tur, D.: Use of kernel deep convex networks and end-to-end learning for spoken language understanding. In: IEEE Workshop on Spoken Language Technology, pp. 210–215 (2012)
24. Shi, Y., Yao, K., Chen, H., Pan, Y.-C.Y., Hwang, M.-Y., Peng, B.: Contextual spoken language understanding using recurrent neural networks. In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 5271–5275 (2015)
25. Bengio, Y.: Deep learning of representations: looking forward. In: Dediu, A.-H., Martín-Vide, C., Mitkov, R., Truthe, B. (eds.) SLSP 2013. LNCS (LNAI), vol. 7978, pp. 1–37. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39593-2_1