Pattern Recognition Letters 42 (2014) 35–39

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Two-phase reanalysis model for understanding user intention ☆

Sangwoo Kang [a,*], Jungyun Seo [b]

[a] Department of Computer Science, Sogang University, South Korea
[b] Department of Computer Science and Interdisciplinary Program of Integrated Biotechnology, Sogang University, South Korea

ARTICLE INFO

Article history:
Received 22 July 2013
Available online 4 January 2014

Keywords:
Natural language processing
Dialogue system
Machine learning
User intention analysis

ABSTRACT

This paper proposes a two-phase reanalysis model for understanding user intention in utterances, by considering the correlative characteristics between the three attributes relating to user intention. The proposed model comprises two phases. In the first phase, each attribute is analyzed in the optimized sequence. The results of the analysis are then used as features that undergo reanalysis in the second phase, with the assumption that the relationship between the attributes is correlative. The experiments conducted showed that the proposed model improves user intention analysis over the baseline model, with an error reduction rate in Speech Act, Concept Sequence, and Arguments of 0.64%, 14.78%, and 5.84%, respectively.

1. Introduction

Literature on dialogue systems for end-user applications has seen a rapid growth in recent years [12]. The aim of a dialogue system is to perform tasks based on natural language processing, and thus has the potential to significantly improve interaction between users and machines.

A conventional dialogue system consists of the following components: natural language understanding, dialogue management, and response generation. Natural language understanding is intended to interpret user intention as accurately as possible, by analyzing input sentences, which is required for dialogue management to generate adequate responses to users. User intention is translated into three attributes, namely of Speech Act (SA), Concept Sequence (CS), and a set of Arguments (ARGs) [13,6]. SA describes an utterance that attempts to affect the addressee, while CS consists of a set of concepts representing domain-dependent actions, and ARGs denote essential details of CS in pairs comprising data type and corresponding value.

The majority of existing literature on user intention analysis focuses on a statistics-based machine-learning approach. Some studies have regarded SA, CS, and ARGs as entities that act independently of one another, while other research has viewed these attributes as having a hierarchical relationship, and has subsequently analyzed them in a sequential order [17]. However, the possibility of a correlative relationship between the three attributes had not been explored.

This paper proposes a two-phase reanalysis model for the attributes of user intention, in which the tentative results of the first analysis phase serve as features to undergo analysis in the second phase, with the view that the three attributes are correlated. The experimental results showed that the proposed model significantly enhances performance when compared to baseline models that sequentially analyze three attributes.

The remainder of this paper is organized as follows. Section 2 reviews existing work concerning the analysis of the user intention. Section 3 explains the architecture and methodology of the proposed model. Section 4 describes our evaluation of the proposed model. Finally, Section 5 concludes the paper.

2. Related work

Machine-learning methods have the ability to automatically identify classifiers for attributes of user intention, using annotated corpora with semantic tags. Such a method can also efficiently handle ambiguities by exploiting probabilities, which is why many recent analysis models are based on statistics.

Lee and Seo [14] initially proposed a model for Korean SA analysis by utilizing sentential probabilities as well as context probabilities. Choi [2] constructed an integrated model that analyzes SA and discourse simultaneously, based on a Maximum Entropy Model (MEM) as a machine-learning method. They address data sparseness by using discourse structure information as the feature function of MEM. Kim and Seo [7] developed a two-step SA analysis model by means of neural networks: in the first step, an initial pool

☆ This paper has been recommended for acceptance by D. Dembele.

* Corresponding author. Address: #908 R-Hall, Sogang Univ., Sinsudong, Mapo-ku, 121-742 Seoul, South Korea. Tel.: +82 2 706 8954; fax: +82 2 706 895.

E-mail address: swkang@sogang.ac.kr (S. Kang).

http://dx.doi.org/10.1016/j.patrec.2013.12.015
0167-8655/© 2014 Published by Elsevier B.V.

of useful features is established, and in the second step, a neural network without a hidden layer is trained on some of the more useful features selected from the first step. Eun [3] used a Support Vector Machine (SVM) classifier for his analysis model, which extracts features of vocabulary, Part-of-Speech (PoS), phase type, and context from SA. Lee and Seo [15] constructed a Hidden Markov Model (HMM), which produces sentences according to transitioning states based on the intention of the speaker. For this HMM, the decision tree data structure was adopted to address data sparseness in the calculation of transition probabilities and observation probabilities. Lee et al. [13] described user intention as being identifiable in SA in the Interchanged Format and translated user intention into a set of SA, CS, and ARGs in order to analyze SA and CS by means of an MEM. Seon et al. [17] attempted to predict identifiable user intention in the current utterance by means of a conditional MEM. Linguistic features were exploited to predict CS, which in turn were used as features for SA prediction.

In recent literature, joint modeling approaches have been presented for SLU. A discriminative log-linear joint model was proposed for semantic role labeling, which incorporates more global features [18]. Tur studied transfer learning for the SLU problem. He showed that multi-task learning can be a useful tool for the SLU model of a single domain with a source task that is allowed to access a large amount of labeled data [19]. A unified model that combines sequence labeling and sequence classification was presented for the joint prediction of a dialog act and named entity [4]. This paper only considered particular cases in a set of correlated components. In particular, the NE recognition problem requires that a label be assigned to a phrase rather than a word. The standard method for dealing with this segmentation problem is to use *BIO* encoding [16], in which the slot labels, identical in meaning to ARG labels, are drawn from a set of classes constructed by supplementing each label as *X–B*, *X–I*, or *O*. Here, *X–B* means "begin a phrase of slot *X*," *X–I* means "continue a phrase of slot *X*," and *O* means "not in a phrase."

Some of the existing literature views SA, CS, and ARGs as independent attributes of user intention; thus, each is determined in isolation. In other cases, it is assumed that one attribute is subordinate to another, and therefore the attributes are determined in sequence. It is thus clear that the majority of the existing studies do not regard SA, CS, and ARGs as having correlative characteristics. Based on the assumption that there is a correlative relationship between attributes, this paper proposes a two-phase reanalysis model that uses the analysis results of the first phase as features of the other attributes in the second analysis phase.

3. Two-phase analysis model for user intention

3.1. Architecture of the proposed model

In this paper, SA, CS, and ARGs are determined in terms of their correlative relationship in order to more accurately analyze the user intention in utterances. Fig. 1 outlines the proposed model, in which the user intention is tentatively analyzed in the first phase, with the results used to refine the analysis in the second phase. In the proposed model, a three-step cascading process is performed for the three attributes in each phase. In each step, the corresponding attribute analyzer uses the results from the preceding analyzers as features. We determined the optimal analysis sequence via experiments.

The proposed model makes use of a reanalysis phase that refines each of the three attributes by reflecting the tentative analysis results obtained from the first phase. In each of the phases, the analyzers use clue words and PoS in the given utterance as its basis feature, which is denoted by *U*. $\boldsymbol{F_P}$ represents the results of the analyzers used during the first phase, while $\boldsymbol{F_{P^*}}$ indicates the most up-to-date results of the previous analyzers in the both phases (detailed in Fig. 2). $\boldsymbol{F_{P^*}}$ is determined by the optimal analysis sequence for maximizing the performance of each attribute.

The next section describes in detail each of the analyzers for SA, CS, and ARGs in establishing the optimal analysis sequence.

3.2. Analysis model for attributes of user intention

The attributes of user intention are analyzed by SA, CS, and ARGs. In the proposed model, the following ten SAs are defined: GREETING, EXPRESS, ASK_IF (YES/NO-QUESTION), ASK_REF (WH-QUESTION), RESPONSE, REQUEST, ASK_CONFIRM, CONFIRM, INFORM, and ACCEPT. SA is determined by an SVM classifier [20,8], and a one-against-one approach is used to extend this binary classifier to be able to accommodate ten categories [1]. According to the one-against-one-approach, the number of necessary SVM classifiers is $k(k-1)2$, where k represents the number of categories. If an utterance is found to be characteristic of multiple SAs, a vote is conducted to determine a single speech act to which the utterance corresponds. The SA classifier in the proposed model makes use of word/PoS pair tags as the sentential features. The sentential features in an utterance are extracted from the lexical information of clue words and the sequence of PoS tags, and these features provide very effective information for analyzing user intention. And the CS of the preceding system utterance as the context feature.

It is not possible to establish the detailed intention identifiable in an utterance using only SA, which is domain independent; hence, CS defines the detailed intention that is identifiable in a given utterance. Therefore, a user intention is appended with tag of CS [5]. As shown in Table 1, the model defines CS tags for the schedule management domain using the following structure: two tables, four operators, and eight fields [13]. The CS of Timetable-Retrieval-Content indicates that the user utterance "Tell me the weekend's schedule" is interpreted as consisting of the intention corresponding to content retrieval from the timetable for the upcoming weekend.

CS is analyzed according to the same method used for SA analysis. SVM classifiers are used to determine the CS of the current utterance by means of the sentential features (word/PoS pair), using the CS of the preceding system utterance as the context feature.

Expressions referring to place, time, date, and persons offer clues as to the topic of the utterance, and such information is represented in pairs consisting of the data type and its value, which are generally referred to as argument sets. In this paper, the following eight major arguments in the schedule management domain are identified and their values extracted from utterances: agent, content, date, day of week, field, person, place, and time. The analysis of the arguments is converted to a classification after boundary markers are added to each of the eight argument types. For instance, the argument set for a person can be supplemented with boundary markers indicating beginning, in, and out, denoted by *B*, *I*, and *O* tags, respectively. In the model proposed in this paper, Conditional Random Fields (CRFs) are used to analyze arguments [11]. CRFs are undirected graph models for maximizing conditional probabilities, and are suitable for sequential labeling [10,9].

Section 4 describes the optimal analysis sequence of attributes that is empirically determined during the first and second phases.

4. Experiment and result

To evaluate the performance of the proposed model, a corpus was constructed using the *Wizard-of-Oz* method in the schedule management domain. In the *Wizard-of-Oz* framework [21], two experimental participants, the *User* and the *Wizard*, communicate