JMLR: Workshop and Conference Proceedings 21:220–223, 2012 The 11th ICGI

Learning Interpretations Using Sequence Classification

Menno van Zaanen MVZAANEN@UVT.NL
TiCC, Tilburg University, Tilburg, The Netherlands

Janneke van de Loo JANNEKE.VANDELOO@UA.AC.BE
CLiPS, University of Antwerp, Antwerp, Belgium

Editors: Jeffrey Heinz, Colin de la Higuera and Tim Oates

Abstract

In this paper we present a system that assigns interpretations, in the form of shallow semantic frame descriptions, to natural language sentences. The system searches for relevant patterns, consisting of words from the sentences, to identify the correct semantic frame and associated slot values. For each of these choices, a separate classifier is trained. Each classifier learns the boundaries between different languages, which each correspond to a particular class. The different classifiers each have their own viewpoint on the data depending on which aspect needs to be identified.

Keywords: Sequence classification, language boundary identification, shallow semantics

1. Introduction

There are many speech recognition applications which involve the understanding of spoken utterances, such as dialog systems and command & control applications. This implies that utterances have to be converted to a form of meaning representation. This representation can range from a simple, shallow representation to complex hierarchical representations. This paper presents a method for the derivation of shallow semantic representations from natural language sentences, inspired by grammatical inference methods which aim to find the boundaries between different languages.

The method is applied to utterances spoken in the context of a command & control card game. The dataset has been collected in the context of the ALADIN project[1], which aims at developing adaptive, self-learning vocal interfaces for people with physical impairments.

2. Experiment

In the experiment described here, we assume to have access to clean natural language sentences (for instance provided by a perfect automatic speech recognition module) together with an interpretation of each sentence that describes the (shallow) semantics contained in the sentence. The task is then to assign correct interpretations to unseen sentences.

1. Adaptation and Learning for Assistive Domestic Vocal INterfaces (`http://www.esat.kuleuven.be/psi/spraak/projects/ALADIN`).

Table 1: Example sentences with their *frame*, *from_suit*, and *from_value* values.

Sentence	*frame*	*from_suit*	*from_value*
Move the two of spades to the three of hearts	*movecard*	*spades*	*two*
Move the five of hearts to the top	*movecard*	*hearts*	*five*
New card	*dealcard*	*empty*	*empty*

2.1. Dataset

The dataset used for the experiment consists of 2,020 utterances (with an average length of 4.3 words) spoken by people playing a voice controlled card game: Patience. This is a relatively simple game in the sense that only a limited set of operations are possible. Players can move a (stack of) card(s) from one stack to another or ask for a new card. Details about the dataset can be found in van de Loo et al. (2012). Here we will briefly describe the general structure of the data.

The utterances were recorded in a setup in which participants were asked to play Patience using spoken commands, which they could make up themselves. The commands were manually transcribed and each command was annotated with a shallow semantic frame description, consisting of a command frame with associated slots (attributes) and values. Slots for which the value was not specified in the command received the value *empty*.

The main semantic category of a sentence is denoted by the frame type, which may be either *dealcard* or *movecard*. The *dealcard* frame does not have any slots; it describes the command of asking for a new card. The *movecard* frame has ten slots. These can be divided into two groups, denoting the source and the target (indicated by *from* and *to* respectively) of the card movement. A card may be identified by its *suit* and *value*. Alternatively, a position on the playing field can be identified: the *hand*, with the new cards; one of the seven *column*s in the middle; or one of the four *foundation* stacks at the top, where all cards should end up at the end of the game. Cards are not taken from the foundation, so this slot is not relevant as a source. The *to_columnempty* and *to_foundationempty* slots indicate that the target is an empty column or foundation. The number of possible values for each of the slots is shown in table 2.

2.2. System description

The identification of the frame type and each of the slot values for a sentence is performed using sequence classifiers (van Zaanen and Gaustad, 2010). For each parameter (either the frame type or any of the slot values) a new sequence classifier is trained. This means that there is one frame classifier and ten slot classifiers. Similarly to the work in van Zaanen et al. (2011), multiple viewpoints on the same data are used. Each classifier uses a different division of sequences. This is visualized in table 1, where the sentences are in different "languages" (or classes) depending on the task (*frame*, *from_suit*, *from_value*, ...).

The final, full interpretation of a sentence is computed by having the sequence classifiers providing values for the frame type and each of the slots and combining the outcomes into a complete interpretation. The evaluation (described in section 2.3) is performed by comparing the full, learned interpretations with those of the gold standard.

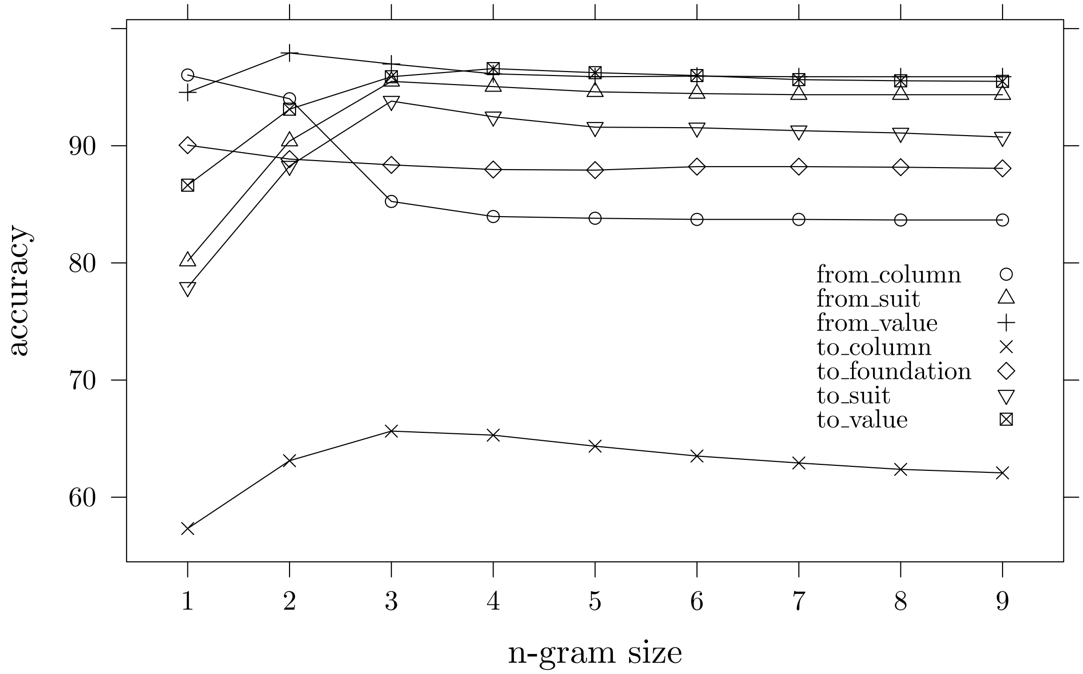

Figure 1: Accuracy results of the different sequence classifiers, apart from *frame*, *from_hand*, *to_columnempty*, and *to_foundationempty*, which have accuracies very close to 100% for $n > 1$.

A sequence classifier aims to identify boundaries between languages. The training data consists of a collection of bags of sequences. Each bag of sequences is a sample from one of the languages. The boundaries between the languages are identified by considering each n-gram (with a predefined length n) and its corresponding *tf*idf* (van Rijsbergen, 1979). The *tf* component measures how frequently the pattern occurs in the language and the *idf* component measures in how many languages the pattern occurs. Thus, *tf*idf* provides a score describing the discriminative power of the pattern.

A new sequence is classified by searching for occurrences of each pattern and, for each occurrence, summing up the *tf*idf* scores. The language (or class) with the highest score is selected. According to the patterns, this is the best fitting language for the sequence.

2.3. Empirical results

Figure 1 shows the results of the different sequence classifiers. The x-axis shows the size of the pattern (n) and the y-axis displays the accuracy (percentage of correctly classified sentences). The accuracies of the different classifiers level off quickly with increasing pattern sizes (n) due to the short average sentence length. A similar image is seen when plotting the classifiers using the range of n-grams (not shown here).

Table 3 shows the accuracy results of the complete interpretations, which are created by combining the results of the separate classifiers. All results significantly ($p < .001$) outperform the majority baseline. The best performance is achieved with tri-grams, which

Table 2: Number of classes for each classifier.

frame	2	
	from	*to*
suit	7	7
value	14	14
hand	2	–
column	3	7
columnempty	–	2
foundation	–	5
foundationempty	–	2

Table 3: Accuracies of full interpretations.

n	accuracy	n	accuracy
baseline	53.12	baseline	53.12
1	62.18		
2	68.81	1-2	67.48
3	71.83	1-3	70.20
4	70.54	1-4	70.45
5	69.95	1-5	70.50
6	69.85	1-6	70.40
7	69.65	1-7	70.45
8	69.46	1-8	70.35
9	69.31	1-9	70.28

is in line with previous findings (van Zaanen et al., 2011), but the differences in performance between the classifiers are not significant, apart from the difference between the unigram accuracy and the accuracies with longer ($n > 1$) n-grams ($p < .05$).

3. Conclusions

In this paper, we describe an extension of the sequence classifier (van Zaanen and Gaustad, 2010) enabling the identification of semantic interpretations. The interpretations (in the form of frame structures) denote shallow semantics of natural language sentences. In this experiment, interpretations describe actions that correspond to commands in a game of Patience. Interpretations are built by combining the output of eleven separately trained classifiers. Each classifier is trained with a different viewpoint on the same set of sequences.

References

J. van de Loo, G. De Pauw, J. Gemmeke, P. Karsmakers, B. Van Den Broeck, W. Daelemans, and H. Van hamme. Towards shallow grammar induction for an adaptive assistive vocal interface: a concept tagging approach. In *Proceedings of the workshop on Natural Language Processing for Improving Textual Accessibility (NLP4ITA)*, Istanbul, Turkey, 2012. European Language Resources Association (ELRA).

C. J. van Rijsbergen. *Information Retrieval*. University of Glasgow, Glasgow, UK, 2nd edition, 1979. Printout.

M. van Zaanen, T. Gaustad, and J. Feijen. Influence of size on pattern-based sequence classification. In P. van der Putten, C. Veenman, J. Vanschoren, M. Israel, and H. Blockeel, editors, *Proceedings of the 20th Belgian-Dutch Conference on Machine Learning; The Hague, The Netherlands*, pages 53–60, 2011.

Menno van Zaanen and Tanja Gaustad. Grammatical inference as class discrimination. In José Sempere and Pedro García, editors, *Grammatical Inference: Theoretical Results and Applications; Valencia, Spain*, number 6339 in Lecture Notes in AI, pages 245–257, Berlin Heidelberg, Germany, 2010. Springer-Verlag.