A Chunking Parser for Semantic Interpretation of Spoken Route Directions in Human-Robot Dialogue

Raveesh Meena, Gabriel Skantze, Joakim Gustafson
Department of Speech, Music and Hearing
KTH, Stockholm
raveesh@csc.kth.se, gabriel@speech.kth.se, jocke@speech.kth.se

Abstract

We present a novel application of the *chunking parser* for data-driven semantic interpretation of spoken route directions into *route graphs* that are useful for robot navigation. Various sets of features and machine learning algorithms were explored. The results indicate that our approach is robust to speech recognition errors, and could be easily used in other languages using simple features.

1. Introduction

It is desirable to endow urban robots with spoken dialogue capabilities so that they can seek route directions from passersby to navigate their way in unknown surroundings. To understand freely spoken route directions a robot's dialogue system would require a spoken language understanding (SLU) component that (i) is *robust* in handling automatic speech recognition (ASR) errors, (ii) learns *generalization* to deal with unseen concepts in free speech, and (iii) preserve the highly *structured relations* among various concepts in a route direction instruction. Existing approaches to SLU in dialogue systems do not cater to one or the other of these requirements.

Grammar based parsers for semantic interpretations of text, besides requiring envisaging all the combinatory rules, are not robust in handling ASR errors. While *keyword spotting* based approaches to SLU has been useful in form-filling task domains, the semantics of route directions is highly structured and cannot be treated as simple "bag of words or concepts". For example, route instruction "*after the church turn left*" contains not just the concepts AFTER, CHURCH, TURN and LEFT but also the structural relations among them: that the action of turning to left has to be taken only after the church.

Earlier in Johansson et al. (2011), we presented a data-driven approach to semantic interpretation of manual transcriptions of route instructions, given in Swedish, into *conceptual route graphs* (CRGs) that represent the semantics of human route descriptions. In Meena et al. (2012) we applied this approach for semantic interpretation of spoken route directions given in English. The results indicate that our approach is robust in handling ASR errors. Here we present an overview of the work reported in Meena et al. (2012).

2. Chunking parser for semantic interpretation

Our approach (Johansson et al., 2011) to automatically interpret manual transcriptions of route instructions into CRGs (Müller et al., 2000), is a novel application of Abney's *chunking parser* (Abney, 1991). In a CRG, e.g. Figure 1, the nodes represent the semantic concepts and the edges their attributes. The concepts, their attributes and argument types are defined in the type hierarchy of the domain model using the specification in the JINDIGO dialogue framework (Skantze, 2010).

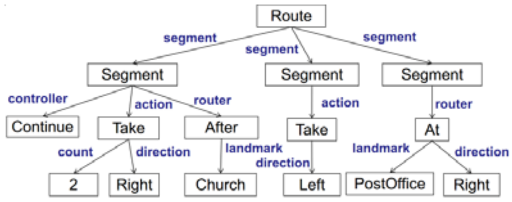

Figure 1: The conceptual route graph for the route instruction "*go straight and take the second right after the church then eh take a left the post is on the right hand side.*"

We apply the *Chunker* stage of the chunking parser for finding base concepts in a given sequence of words. For example, route instruction "*turn left after eh the church*" could be chunked as the following:

[**ACTION** *turn*] [**DIRECTION** *left*] [**ROUTER** *after*] [**FP** *eh*] [**LANDMARK** *the church*]

To turn chunking into a classification problem, we followed the common practice of assigning two labels for each type of chunk: one with prefix B- for the first word in the chunk and one with prefix I- for the remaining words. The *Attacher* then takes a base concept (a chunk) as input and does two things: First, it may assign a more specific concept class (like CHURCH). To allow it to generalize, the Attacher also assigns all ancestor classes, based on the domain model (i.e. BUILDING for CHURCH; this, however, is not shown in the example). The second task for the Attacher is to assign attributes, e.g. *direction*, and assign them values, e.g. →, which means that the interpreter should look for a matching argument in the right context. While the Chunker in our approach is a *single-label classifier* the Attacher is a *multi-label classifier* where none, one or several labels may be assigned to a chunk. The Chunker output from above could be modified by the Attacher as the following:

[**TAKE** (*direction*: →) *turn*] [**RIGHT** *right*] [**AFTER** (*landmark*: →) *after*] [**DM** *eh*] [**CHURCH** *the church*]

As a final step, heuristic rules were used to group the CONTOLLER, ROUTER and ACTION chunks into *route segments*, which collectively form a CRG. To measure the performance of the Chunking parser we used the notion of Concept Error Rate (CER), which in our approach is the weighted sum of the edits required in the reference CRG key to obtain the resulting CRG.

3. Chunking parser for SLU

In Meena et al. (2012) we made three extensions to this approach. First, we introduced another chunk learner – the *Segmenter* – to automatically learn *route segments* in a sequence of chunks. The Chunker output shown earlier could be segmented as the following:

[SEGMENT [ACTION *turn*] [DIRECTION *right*] [ROUTER *after*] [FP *eh*] [LANDMARK *the church*] **]**

The Attacher performs the same tasks as earlier, except that it now looks for attachments only within the route segment. Secondly, we verified whether our approach could be applied for semantic interpretation of route directions given in another language. Third, we evaluated the performance of the Chunking parser on ASR results. Towards this, we used the IBL corpora of route instructions (Kyriacou et al., 2005). It contains audio recordings and manual transcriptions of 144 spoken route instructions given in English. As a first step, we evaluated the Chunking parser's performance on manual transcriptions and obtained the baseline for comparing its relative performance on ASR results. 30 route instructions were manually annotated and used as the cross-validation set. Next, we trained an off-the-shelf ASR system with the remaining 113 route instructions. The best recognized hypothesis for each instruction (in the cross-validation set) obtained through this trained ASR was used for validating the Chunking parser's performance.

We tested the Chunking parser's performance on Naïve Bayes (NB) and Linear Threshold Unit (LTU) algorithms. Two types of LTU: Sparse Perceptron (SP) and Sparse Averaged Perceptron (SAP) were tested. Due to space constraints we present, in the first column of Table 1, only those feature (combinations) for which the learners obtained the best results. Interested readers are referred to Meena et al. (2012) for complete details.

4. Results

The performance scores of a *keyword spotting* based method for *Chunking parser* were used as the baseline for drawing comparisons. A baseline CER of 50.83 was obtained for the Chunker using the NB learner. For the Segmenter a CER of 77.22 was obtained using the NB learner, and for the Attacher a CER of 75.07 was achieved with the SAP learner.

In general LTUs performed better than NB algorithms. The best performance for the Chunker is obtained using the SAP learner in conjunction with the additive features shown in Table 1-A. The Segmenter performed best using the SAP (cf. Table 1-B). The Attacher performed best using the SAP learner (cf. column 2, Table 1-C); however, due to poor placement of route segment boundaries by the Segmenter, it could not always attach concepts to their valid argument(s). To obtain an estimate of Attacher's performance independent of the Segmenter we compared only the sub-graphs in all the *route segments* of a CRG with their counterpart in the reference CRG (cf. column 3, Table 1-C).

A: Chunker performances with additive features.			
Features	**CER_{NB}**	**CER_{SP}**	**CER_{SAP}**
Word instance	50.83	46.15	45.17
+Word window	17.31	21.33	20.82
+Previous tags	18.16	10.86	**10.64**
B: Segmenter performances with the best training feature.			
Features	**CER_{NB}**	**CER_{SP}**	**CER_{SAP}**
Chunk label window	31.67	**25.83**	28.89
C: Attacher performances with the best training feature.			
Features	**CER_{SP}**	**CER_{SAP}**	**$rgCER_{SAP}$**
Bag of words (BW)	29.42	**29.11**	**19.99**

Table 1: *Performances of the Chunking parse components.*

Figure 2 illustrates the Chunking parser performances (CER) w.r.t to ASR performances in recognizing the spoken route instructions. The dashed-red line is Chunker parser's baseline performance on the manual transcriptions. The CER curves suggest that Chunking parser's performance follows WER.

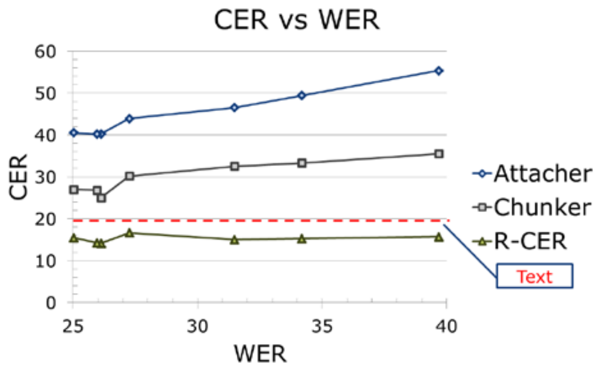

Figure 2*: Chunking parser's CER w.r.t ASR's WER (R-CER is relative-CER = CER minus WER)*

The rather steady *relative* CER: R-CER (the relative gain in CER due to ASR errors) in Figure 2 highlights the robustness of our approach in dealing with errors in speech recognition. In addition to this, Chunking parser's performance on transcribed route instructions given in Swedish (CER 25.60 (Johansson et al., 2011)) and in English (CER of 19.99 vs. baseline CER of 77.70) are encouraging figures that indicate that our approach can be easily used in other languages using simple features.

5. References

Abney, S. (1991). Parsing by chunks. In Berwick, R. C., Abney, S. P., & Tenny, C. (Eds.), *Principle-Based Parsing: Computation and Psycholinguistics* (pp. 257-278). Dordrecht: Kluwer.

Johansson, M., Skantze, G., & Gustafson, J. (2011). Understanding route directions in human-robot dialogue. In *Proceedings of SemDial* (pp. 19-27). Los Angeles, CA.

Kyriacou, T., Bugmann, G., & Lauria, S. (2005). Vision-based urban navigation procedures for verbally instructed robots. *Robotics and Autonomous Systems*, *51*(1), 69-80.

Meena, R., Skantze, G., & Gustafson, J. (2012). A Data-driven Approach to Understanding Spoken Route Directions in Human-Robot Dialogue. In *Proceedings of Interspeech*. Portland, OR, US.

Müller, R., Röfer, T., Lankenau, A., Musto, A., Stein, K., & Eisenkolb, A. (2000). Coarse qualitative descriptions in robot navigation. In Freksa, C., Brauer, W., Habel, C., & Wender, K-F. (Eds.), *Spatial Cognition II* (pp. 265-276). Springer.

Skantze, G. (2010). *Jindigo: a Java-based Framework for Incremental Dialogue Systems*. Technical Report, KTH, Stockholm, Sweden.