A Stochastic Case Frame Approach for Natural Language Understanding

Wolfgang Minker, Samir Bennacef, Jean-Luc Gauvain

Spoken Language Processing Group
LIMSI-CNRS
91403 Orsay cédex, FRANCE
email: {minker,bennacef,gauvain}@limsi.fr

10.21437/ICSLP.1996-261

ABSTRACT

A stochastically based approach for the semantic analysis component of a natural spoken language system for the ATIS task has been developed. The semantic analyzer of the spoken language system already in use at LIMSI makes use of a rule-based case grammar. In this work, the system of rules for the semantic analysis is replaced with a relatively simple, first order Hidden Markov Model. The performance of the two approaches can be compared because they use identical semantic representations despite their rather different methods for meaning extraction. We use an evaluation methodology that assesses performance at different semantic levels, including the database response comparison used in the ARPA ATIS paradigm.

1. INTRODUCTION

We have been investigating the portability of the understanding component of a natural spoken language system. Stochastic methods are attractive because they can be adapted to new conditions (task, language) if appropriate training corpora are available. Stochastic methods for speech understanding have already investigated in the BBN-HUM [8] and the AT&T-CHRONUS [6] systems.

In this paper we present a strategy for semantic decoding in which a stochastic model replaces a rule-based analysis. The rule-based system was originally developed for L'ATIS [2], a French language system for the Air Travel Information Services (ATIS) task. This component, based on a case grammar formalism [3], offers the advantage that it does not require verifying the correct syntactic structure of a query, but extracts its meaning using syntax as a constraint. In order to investigate language portability, this component has been ported to American English using the ARPA ATIS2 corpus [7]. Both approaches rely on the same case grammar terminology enabling us to compare their performances.

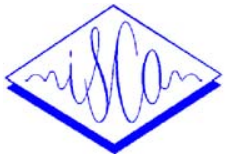
The stochastic model, implemented as a first order Hidden Markov Model (HMM), has been trained on the answerable queries of the ARPA ATIS0 and ATIS2 corpus. Each query was semantically annotated on a word-by-word basis using the case frame based system. These annotations were manually corrected before training the stochastic model. The output of the stochastic decoder is a sequence of semantic expressions which can be directly converted to a semantic frame without supplementary interpretation rules. The strength of this method is that, except for the semantic labeling of the large corpus and the design of a conceptual preprocessing component, the system training is automatic.

We use a multi-level evaluation methodology that assesses performance of the understanding module at different stages, i.e., the semantic representation at various levels of precision including the database response comparison adopted in the ATIS ARPA evaluation paradigm for natural language understanding systems [1]. This allows for a precise error analysis when evaluating the two approaches, so as to determine their relative strengths and weaknesses. Evaluation using the ATIS ARPA reference answers allows for comparison with previously reported results on the same data.

2. RULE-BASED CASE GRAMMAR

Spoken language understanding systems aim to extract the semantic content of a spoken query so as to be able to carry out an appropriate action. Human interaction via voice is of a spontaneous nature with spoken language effects such as false starts, repetitions and requests, which do not necessarily respect the written grammar. It would therefore be improvident to base the semantic extraction on a purely syntactic and sometimes incomplete analysis of the input query. Parsing failures due to ungrammatical syntactic constructs may be reduced, if those portions containing important semantic information could be identified whilst ignoring the non-essential or redundant parts. The robust parsing in CMU's PHOENIX system follows this strategy and applies a case grammar formalism [4].

L'ATIS, a spoken language understanding system for a French version of the ARPA ATIS task has been previously described [2]. Its spoken language understanding component is also based on a case grammar formalism [3] which detects domain-related concepts and instanciates the corresponding semantic structure using a set of constraints. In the request *Je voudrais les vols de Denver à Pittsburgh pour demain s'il vous plaît* (*I would like the flights from Denver to Pittsburgh for tomorrow please*) the concept is **flight** identified by the keyword *vols*, and the constraints are **departure-town** (*Denver*), **arrival-town** (*Pittsburgh*) and **departure-day** (*demain*). From the point of view of the case grammar, the concept corresponds to the casual structure and the constraints correspond to the cases. In L'ATIS, the case grammar is described by a system of rules in a declarative file enumerating the totality of the casual structures and the cases related to the application. The analysis of an input sentence consists of identifying its casual structure and of constructing a semantic representation in the form of a frame. The values of the constraints are instanciated using the case markers. In the example phrase *de Denver à Pittsburgh*, the prepostition *de* designates the value *Denver* to be a **departure-town** and *à* designates *Pittsburgh* to be an **arrival-town**.



4th International Conference on Spoken Language Processing (ICSLP 96)
Philadelphia, PA, USA
October 3-6, 1996

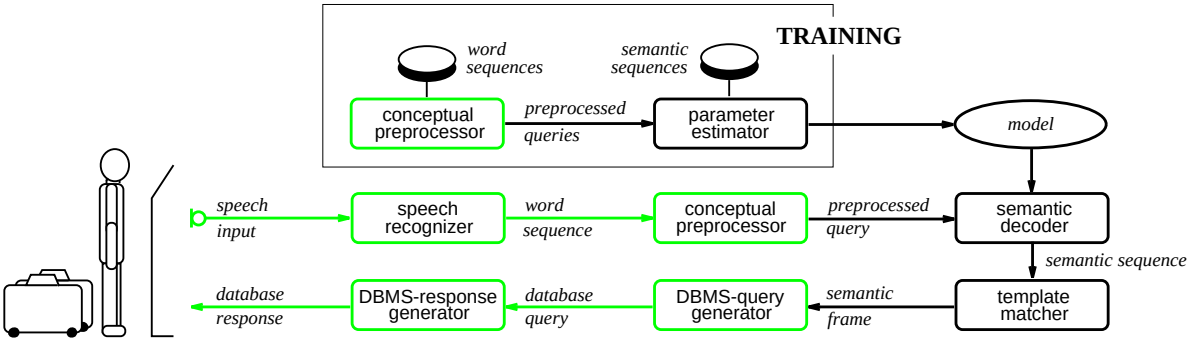


Figure 1: Overview of spoken language understanding system.

In order to investigate language portability, the spoken language understanding component was ported to American English using the type A queries of the ARPA ATIS2 corpus for iterative rule development and testing [9]. The porting process consisted of translating and modifying the case grammar and the rules for response generation. Throughout development, the understanding component was iteratively evaluated in order to monitor the consistency of the changes. With an extended domain coverage, the semantic analyzer contained 13 semantic categories making use of a set of 69 cases - nearly twice as many as in the French system. The case grammar formalism was found to be easily portable to a new language by translation of the system of rules whilst considering some language specificities.

3. STOCHASTIC CASE FRAME ANALYSIS

Figure 1 shows an overview of the understanding system where a stochastic model replaces the system of rules for the semantic analysis. Along with using the same set of symbolic labels used in the case frame approach, the speech recognizer, conceptual preprocessor, DBMS-query and response generator are shared. During training, the parameter estimator estimates the parameters of the stochastic model given preprocessed word sequences (the observations) and the corresponding semantic sequences (the states). The semantic sequences can be derived from the case frame representation by aligning the concepts, case markers and constraining values. The semantic decoder, an ergodic bigram backoff HMM [5], outputs the most likely semantic sequence given the unknown input query. Using the token-value pairs of the preprocessed query, the template matcher reconverts the semantic sequence into a semantic frame for use by the database access and response generation components.

We now discuss the conceptual data preprocessing, the semantic representation, and the model topology applied in the stochastic system.

3.1. Conceptual preprocessor

Stochastically-based approaches require substantial amounts of data for parameter estimation. In the domain of natural spoken language understanding, data annotation is quite difficult and expensive. As a result, the corpora are limited in size which is problematic for maximum likelihood estimators as they do not adequately model events that are rarely observed in the training data. In addition to backoff techniques [5], one possible solution is to preprocess the data using a conceptual analysis. Unification of the input simplifies the work done during semantic analysis, but more importantly reduces the number of parameters and thus the model size. Such preprocessing is relatively easy in a limited task such as ATIS. However this type of analysis is rather domain-dependent and delicate to mainipulate. In order to carry out a systematic and exhaustive conceptual analysis, the preprocessing component in [2] has been extended and refined.

The first step involves query simplification. The input is converted to lower case, numbers converted to digit strings and codes written as single words (`a p slash eighty` ⇒ `ap/80`). Whenever possible, names are replaced by their database codes. The unified ATIS0 and ATIS2 data contain 1,164 distinctive lexical entities. To reduce the model size, the morphologic analysis then

- converts compound phrases into hyphenated compound expressions (*how many* ⇒ `how-many`).
- replaces inflected forms with their corresponding base forms (*cities* ⇒ `city`, *goes* ⇒ `go`).
- groups semantically related words into word classes `(arrive)`, `(capacity)`, `(count)`, `(fare)`,... and assigns non-relevant or out-of-domain words to the classes `(fill)` and `(ood)` respectively.

After morphological analysis, the number of lexical entries is reduced to 737. The conceptual preprocessor transforms the example utterance *Show flight American Airlines fourteen forty three* (atis0 - b600c1sx) into `(fill) (flight) AA 1443`.

3.2. Semantic representation

Figure 2 shows the semantic structures used by the case grammar formalism and the modified structures used in the rule- and stochastically-based systems. The structures have been aligned with the conceptually preprocessed query. Additional local syntactic constraints are introduced between markers (m:case) and constraining case values (v:case) in order to enable the value extraction in the rule-based approach. The case markers may be distinguished as *pre-* or *post-markers*, are *adjacent* or *non-adjacent* to the corresponding values. In the example query in Figure 2, `AA` is a *pre*marker for the flight-number `1443` (m:pre:flight_num).

In the stochastic approach, the notion of locality for case markers is implicitly contained in the semantic sequence, and the initial case grammar formalism is adopted, e.g. (m:fight_num). Within the semantic sequence we define basic *semantic units* corresponding to the concepts (<concept>), case values and case markers. These units

Conceptually preprocessed query			
(fill)	*(flight)*	*AA*	*1443*
Case grammar formalism			
	(<flight>)	(v:airline)(m:flight_num)	(v:flight_num)
Rule-based method			
	(<flight>)	(v:airline)(m:pre:flight_num)	(v:flight_num)
Stochastic approach			
(dummy)	(<flight>)	(v:airline)(m:flight_num)	(v:flight_num)

Figure 2: Semantic representation used by the case grammar formalism and applied in the rule and stochastically based systems for the example query *Show flight American Airlines fourteen forty three* (atis0 - b600c1sx).

combine to more complex *semantic expressions.* In the example *AA* is both the value of the case airline (v:airline) and a marker for the flight-number (m:flight_num) *1443*. In both the case grammar and the rule-based method, the semantic annotation is not exhaustive. It considers only those words of the input query that are related to the concept and its constraints. However, in order to correctly estimate the model parameters, the stochastic approach requires a complete annotation of the input query. Each contextual unit of the input query must have a corresponding a semantic label. To assure this the label (dummy) is introduced for those contextual word units that are judged to be not needed for the task. In the example query, *show*, which was transformed to the class *(fill)* corresponds to the semantic label (dummy).

3.3. Stochastic Model

The segmented corpus contains a total of 330 different semantic expressions, defined to be the states of a first order HMM. The state transitions probabilities are bigrams which can model only the adjacent marker-value relations, but not longer distance relations. We use a simple ergodic topology, allowing all semantic expressions to follow each other. The observations correspond to the 737 conceptually preprocessed lexical entries.

Semantic expressions (states)	*Conceptually preprocessed words (observations)*
(<flight>)	*(flight),(leave),(arrive), time,flight-number*
(<airfare>)	*(fare),ticket*
(m:order_arriv)(<flight>)	*(arrive)*
(v:order_arriv)	*earliest,early,first,same*
(v:stop-nonstop)	*nonstop,stop,direct,connect*
(v:stop-city)	*ddfw,dden,matl,ppit,pphl*
(v:to-city)	*ssfo,dden,matl,bbos,pphl*
(m:stop-city)	*stop*
(m:to-city)	*to,and,in,for,(arrive)*

Table 1: Examples of semantic expressions (considered as the states in the stochastic model) along with the corresponding conceptually preprocessed words (the observations).

Table 1 shows examples of state-observation correspondencies. Various observations are attributed to different semantic expressions, e.g. *(stop)* is associated with both (v:stop-nonstop) and (m:stop-city). City codes (*ddfw, dden, matl, ...*) are attributed to the semantic expressions (v:stop-city), (v:to-city) depending on the adjoining marker (m:stop-city), (m:to-city). The (dummy) - *(fill)* and (dummy) - *(ood)* correspondencies are removed from the training data since they do not provide any meaningful information.

4. CORPUS ESTABLISHMENT

The stochastic model has been trained using the 6,439 answerable type A+D[1] queries of the ARPA ATIS0 and ATIS2 corpora. Prior to training and testing the corpora were semantically annotated. The test data consist of the transcriptions of the 402 type A queries in February 1992 ATIS ARPA Benchmark test. The English rule-based understanding component of L'ATIS [9] was used to produce a semantic frame for each query and along with a preliminary sequential representation (Figure 2). Given that the rule-based understanding component is not error-free, the preliminary labels must be verified. In order to simplify this task, all semantic representations that have judged incorrect according to the database response evaluation [1] are flagged for manual correction.

5. MULTI-LEVEL EVALUATION

A multi-level performance evaluation method is used to measure the performance of the understanding component at different stages. The ARPA ATIS paradigm [1] for the natural language systems evaluation was carried out on the SQL database response. Even though the this paradigm allows comparison of results in the natural language processing community, it does not directly reflect the performance of the understanding component itself. Evaluating the semantic representation at various levels as shown in Figure 3 enables a more refined error analysis.

The most severe evaluation is applied to the *semantic sequence*, the output of the semantic analyzer. A scoring program compares the accuracy of the hypothesized sequence to that of the reference sequence. All labels - concepts, markers and constraining values - are compared. Semantic sequence evaluation is the equivalent of the commonly used word accuracy measure for speech recognition. This measure may in fact be stricter than is necessary and a more appropriate evaluation may be to consider only errors on *concepts* and *values*, since these are relevant for database access. *Database response* is evaluated using the ARPA ATIS evaluation paradigm [1].

Approach	*Evaluation level*				
	sequence		*concept/value*		*response*
RULE-BASED	85.6	(96.4)	85.6	(94.8)	83.8
STOCHASTIC	58.2	(91.4)	65.2	(88.7)	67.9

Table 2: Multi-level evaluation of the rule-based and the stochastic NL understanding components using the type A queries in the ATIS February 1992 Benchmark test data. Sentence-level semantic accuray and response accuracy (%); in parenthesis the accuracy is given for the individual semantic expressions.

Table 2 shows the accuracies on the complete semantic sequences, as well as the sequences of concepts and values output by rule-based and the statistical understanding components. The accuracy of the individual semantic expressions (given in parentheses) of the rule-based model is 96% and the concept/value accuracy is 95%. For the stochastic approach the accuracies are lower (91% and 89%) which is to be expected given the rather simple model topology.

The query *please list the prices for the flights from Dallas to Baltimore on June twentieth* (feb92-e80042sx), is preprocessed to *the (fare) for the (flight) from ddfw to bbwi on june 20*. It

[1] Following the ARPA classification, type A signifies context-independent queries and type D signifies context-dependent queries.

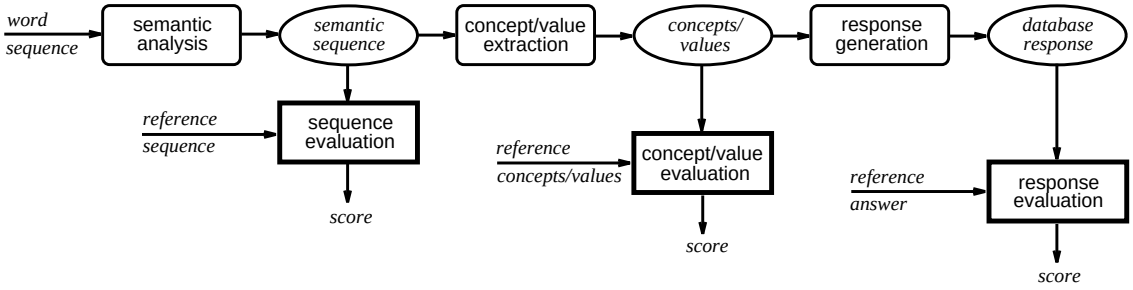


Figure 3: Multi-level evaluation of the natural language understanding component.

contains two keywords corresponding to the different concepts (<airfare>) and (<flight>). In the rule-based approach, the identification of the appropriate concept is guided by the order in which the keywords appear in the query and by the rule application order of the case grammar. Once a keyword is chosen, other keywords within the query are ignored. In the current implementation of the stochastic system the word units output from the conceptual preprocessor are considered as the observations, and are modeled independently of their context. The system therefore fails on the example query because it identifies the two concepts. 25.8% of the errors on the individual semantic expressions were related to this type of problem.

The difference in the accuracy on the semantic sequences and the concept/value sequences for the stochastic system indicates that the markers and values are less tightly coupled than in the rule-based system. This means that an incorrect case marker may still be followed by a correct value. In the rule-based system where an incorrect case marker leads to an incorrect case value, the performance result not change.

A priori we may expect that the database response evaluation should yield the highest performance, as even an incorrect semantic representation can potentially yield a correct database response. However, there is not a large difference, and for the case-frame analysis the results are worse. We attribute this difference to the difficulty of matching the response generator to the "rules of interpretation" adopted in the ARPA community.

6. CONCLUSION AND FUTURE DIRECTIONS

In this paper, we have presented a stochastic case frame approach for natural spoken language understanding as an alternative to the system of rules for semantic analysis previously described. The strength of the stochastic method is that it limits the human effort in system development to the tasks of data labeling and maintenance of the conceptual preprocessing component. The labeling task is much simpler than maintainance (and extension) of the case-frame grammar rules. A multi-level evaluation method has been used, that involves performance tests on different semantic levels, including the database response level adopted in the ARPA community.

Error analysis of this simple stochastic system revealed an essential problem related to the lack of contextual information, as well as the difficulty to update the response generation part in global systems performance. We are now planning to introduce broad contextual information into the stochastic model to improve performance. We also are investigating the use of a domain-independent morphologic analysis to replace the conceptual preprocessing in order to further increase the flexibility and portability of the system towards new domains and languages.

7. REFERENCES

1. M. Bates, S. Boisen, and J. Makhoul. Developing an Evaluation Methodology for Spoken Language Systems. In *Proceedings of DARPA Speech and Natural Language Workshop*, February 1992.
2. S. K. Bennacef, H. Bonneau-Maynard, J. L. Gauvain, L. F. Lamel, and W. Minker. A Spoken Language System For Information Retrieval. In *Proceedings of ICSLP*, September 1994.
3. B. Bruce. Case Systems for Natural Language. *Artificial Intelligence*, 6:327–360, 1975.
4. S. Issar and W. Ward. CMU's Robust Spoken Language Understanding System. In *Proceedings of the European Conference on Speech Technology, EUROSPEECH*, September 1993.
5. S. M. Katz. Estimation of Probabilities from Sparse Data for the Language Model Component of a Speech Recognizer. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 35(3):400–401, 1987.
6. E. Levin and R. Pieraccini. CHRONUS - The Next Generation. In *Proceedings of ARPA Workshop on Human Language Technology*, January 1995.
7. MADCOW. Multi-Site Data Collection for a Spoken Language Corpus. In *Proceedings of DARPA Speech and Natural Language Workshop*, February 1992.
8. S. Miller, M. Bates, R. Bobrow, R. Ingria, J. Makhoul, and R. Schwartz. Recent Progress in Hidden Understanding Models. In *Proceedings of ARPA Workshop on Human Language Technology*, January 1994.
9. W. Minker and S.K. Bennacef. Compréhension et Évaluation dans le Domaine ATIS. In *Proceedings of the Journées d'Études en Parole, JEP*, June 1996. English version: W. Minker. An English Version of the LIMSI L'ATIS System Technical Report 9512, LIMSI-CNRS, April 1995. Notes et Documents LIMSI.