

STOCHASTIC REPRESENTATION OF CONCEPTUAL STRUCTURE IN THE ATIS TASK

Roberto Pieraccini, Esther Levin, Chin-Hui Lee

AT&T Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ, 07974

ABSTRACT

We propose a model for a statistical representation of the conceptual structure in a restricted subset of spoken natural language. The model is used for segmenting a sentence into phrases and labeling them with concept relations (or cases). The model is trained using a corpus of annotated transcribed sentences. The performance of the model was assessed on two tasks, including DARPA ATIS class A sentences.

$$\begin{aligned} c_6 = c_7 &= C_1 \\ c_8 = c_9 &= C_2 \\ c_{10} = \dots = c_{12} &= C_3 \end{aligned}$$

In this paper we tackle the problem of decoding the words constituting the spoken sentence and the corresponding sequence of case labels, from the speech signal.

INTRODUCTION

The goal of a speech understanding system is generally that of translating a sequence of acoustic measurements of the speech signal into some form that represents the meaning conveyed by the sentence. One of the knowledge representation paradigms, known as *semantic networks* [2] establishes relations between conceptual entities using a graph structure. These *concept relations*, or *linguistic cases*, can be used to label different parts of a sentence in order to obtain its interpretation. The task itself defines the set of relevant cases. For instance, for the task of assigning the origin, the destination and the departure time of a flight, a convenient representation is in terms of the following set of cases: $C = \{C_1, C_2, C_3, C_4\}$, where $C_1 = \text{ORIGIN}$, $C_2 = \text{DESTINATION}$, $C_3 = \text{DEPARTURE_TIME}$, and $C_4 = \text{DUMMY}$. The introduction of a DUMMY case is useful for covering all the parts of the sentence that are not relevant to the task. A sentence like *I would like to fly from Boston to Chicago next Saturday night* can be analyzed as:

- DUMMY: *I would like to fly*
- ORIGIN: *from Boston*
- DESTINATION: *to Chicago*
- DEPARTURE.TIME: *next Saturday night*.

Note that although the first phrase (*I would like to fly*) conveys important information, it is considered irrelevant to this particular task, and therefore assigned to the DUMMY case. The segmentation of a sentence into cases (*conceptual segmentation*) can be described by labeling each word in the sentence with the index of the case it expresses. In the example above, the conceptual segmentation is represented by the following sequence of labels:

$$\mathbf{C} = (c_1, c_2, c_3 \dots c_{12}) \quad (1)$$

where:

$$c_1 = c_2 = \dots = c_5 = C_4 \quad (2)$$

MAP DECODING OF CASES

Let us denote by

$$\mathbf{A} = a_1, a_2 \dots a_N, \quad (3)$$

the sequence of acoustic observations extracted from a spoken sentence, by

$$\mathbf{W} = w_1, w_2 \dots w_M, \quad (4)$$

the sequence of words constituting the sentence, and by

$$\mathbf{C} = c_1, c_2 \dots c_M, \quad (5)$$

the sequence of case labels, where c_i takes its values from a pre-defined set of conceptual relations $C = \{C_1, C_2, \dots, C_K\}$. The problem of finding \mathbf{W} and \mathbf{C} given \mathbf{A} can be approached using the maximum a posteriori decoding (MAP). Following this criterion we want to find the sequence of words $\tilde{\mathbf{W}}$ and the sequence of cases $\tilde{\mathbf{C}}$ that maximizes the conditional probability

$$P(\tilde{\mathbf{W}}, \tilde{\mathbf{C}} | \mathbf{A}) = \max_{\mathbf{W} \times \mathbf{C}} P(\mathbf{W}, \mathbf{C} | \mathbf{A}). \quad (6)$$

This conditional probability can be written using the Bayes inversion formula as:

$$P(\mathbf{W}, \mathbf{C} | \mathbf{A}) = \frac{P(\mathbf{A} | \mathbf{W}, \mathbf{C})P(\mathbf{W} | \mathbf{C})P(\mathbf{C})}{P(\mathbf{A})}. \quad (7)$$

In this formula $P(\mathbf{C})$ represents the a-priori probability of the sequence of cases, $P(\mathbf{W} | \mathbf{C})$ is the probability of a sentence expressing a given sequence of cases, and $P(\mathbf{A} | \mathbf{W}, \mathbf{C})$ is the acoustic model. We can reasonably assume that the acoustic representation of a word is independent of the conceptual relation it belongs to, hence:

$$P(\mathbf{A} | \mathbf{W}, \mathbf{C}) = P(\mathbf{A} | \mathbf{W}), \quad (8)$$

and this is the criterion that is usually maximized in stochastic based speech recognizers, for instance those using hidden Markov

modeling [1] for the acoustic/phonetic decoding. In this paper we deal with the remaining terms

$$P(W | C)P(C) = \prod_{i=2}^M P(w_i | w_{i-1} \dots w_1, C)P(w_1 | C) \prod_{i=2}^M P(c_i | c_{i-1} \dots c_1)P(c_1). \quad (9)$$

We proceed by assuming that:

$$P(w_i | w_{i-1} \dots w_1, C) = P(w_i | w_{i-1} \dots w_{i-n}, c_i), \quad (10)$$

and

$$P(c_i | c_{i-1} \dots c_1) = P(c_i | c_{i-1} \dots c_{i-m}). \quad (11)$$

These are Markov processes of order n and m respectively, and if n and m are large we don't lose any generality by making this assumption. For practical purposes n and m should be small enough to allow a reliable estimation of the probabilities from a finite set of data. An additional assumption in equation (10) is that a given word in the sentence, used for expressing a certain case, is independent of the case of the preceding words. Assuming that the sequence of words could be directly observed (for instance providing a transcription of the uttered sentence), and the sequence of cases is unknown, equations (10) and (11) describe a hidden Markov process, where the states of the underlying model correspond to the cases, the observation probabilities of each state are represented by equation (10) in the form of state local ($n + 1$)-gram language models, and the transitions between the states are described by equation (11).

THE FROM-TO TASK

A first evaluation of the model was performed based on a set of 825 sentences artificially generated by a finite state grammar [3] using a vocabulary of 41 different words. The sentences express different ways of making requests to travel between two cities. A typical example is:

I want to travel into Boston and I am interested in flights between Boston and Washington

The task consisted of identifying the origin and destination cities of the flight. The relevant cases of this task are then *flight origin* and *flight destination*. However the model has three states, ORIGIN, DESTINATION and DUMMY. 50 sentences, randomly selected out of the 825, were used to estimate the parameters of the model, i.e. the transition probabilities (equation 11) and the state local language models (equation 10), with $n = 1$ and $m = 1$ (i.e. the underlying Markov process was a 1st order process and the state local language models were bi-grams). The training sentences were hand-labeled with the appropriate cases. The remaining 775 sentences were decoded using Viterbi decoding algorithm. The performance was assessed by counting the number of sentences that were segmented assigning the correct words (i.e. the correct city names) to the DESTINATION and ORIGIN states. We observed that 7% of the sentences (55 out of

775) had a wrong origin/destination assignment. In some of the wrong segmentations one of the relevant states was missing, the other state containing both the real destination and origin cities. In other examples, similar to the sentences shown above, both the destination and the origin states were assigned to the same city name, that appeared twice in the sentence. To improve the performance we imposed some additional constraints in the decoding procedure. For a given sentence the decoded state sequence was searched among those sequences of states where both the origin and destination states were visited only once (i.e. when one of those states was left, the current partial path was not allowed to enter that state again). In addition, the phrases assigned to the origin and destination states had to include different city names. These constraints, representing a higher level a priori knowledge of the task, were imposed in the Viterbi decoding by keeping track of the past sequence of states for each partial candidate solution, and duplicating the partial solutions when two (or more) candidates merged at the same state and showed conflicting constraints. This approach resulted in a substantial improvement of the performance. Only one error was observed out of the 775 test sentences (0.13% error rate). The same level of performance was obtained in experiments using a 1-gram language model inside each state, but increasing the number of states to five:

ORIGIN, DESTINATION, DUMMY, FROM, TO.

The last two states accounted for the expressions that usually precede the origin and destination city names respectively. For example the FROM state was associated to expressions of the kind: *from, depart out of, leaving, etc.*, and the TO state was associated to expressions like: *to, going to, arriving into, etc.* This experiment indicates that there is a tradeoff between the number of states and the complexity (order) of the state language models. Expanding the set of states to reflect the linguistic structure of the sentences may result in a reduction of the number of parameters to be estimated during training, giving a more robust model.

THE ATIS TASK

The technique of case decoding is being applied to the class A sentences of the DARPA ATIS task. A sentence of this task can be analyzed in terms of 7 general cases, that are QUERY, generally associated to the phrases expressing the kind of request, OBJECT expressing the object of the query, ATTRIBUTE that describes some attributes of the object, RESTRICTION describing the restrictions on the values of the answer, Q_ATTR describing possible attributes of the query, AND including connectives like *and, or, also*, indicating that the sentence may have more than one query. Of course we include a DUMMY state like in the above mentioned examples. For example, a sentence like:

What type of economy fare could I get from San Francisco to Dallas on the 25th of April

is segmented as:

- QUERY: *What type of*
- ATTRIBUTE: *economy*
- OBJECT: *fare*

- Q_ATTR: *could I get*
- RESTRICTION: *from San Francisco to Dallas on the 25th of April*

We can further analyze some of the cases into more detailed conceptual relations, giving the following representation:

- ATTRIBUTE
 - ◊ a_fare: economy
- RESTRICTION
 - ◊ origin: *from San Francisco*
 - ◊ destination: *to Dallas*
 - ◊ date : *on the 25th of April*

We defined 44 different cases for describing the whole set of 547 class A training sentences. The complete list of cases is shown in Table 1. The training sentences (covered by a vocabulary of 501 words) were hand-labeled according to this set of states and the transition probabilities and the state local bigram models were estimated using the maximum likelihood criterion. Table 2 shows examples of the phrases used for estimating the bigram language models for some of the defined states. Considering the large number of parameters to be estimated (i.e. the transition probabilities between the 44 states of the model and the 44 bigram models extended to the entire vocabulary of 501 words), and considering the small number of training sentences, this estimation poses robustness problems. One way to alleviate these problems consists of grouping the words in the vocabulary into equivalence classes. For example all the city names can be grouped in the same class, as well as the airport names, the numbers, the airline names, etc.

The testing of the system was performed on the transcribed Jun-90 and Feb-91 class A test sentences. New words were allocated to a *new-word* category that was assigned a small probability within each state. Table 3 reports the number of sentences, along with the statistics on the correctly assigned cases. Table 4 shows examples of correct segmentations from the FEB-91 test set. It is interesting to notice the allocation of the connective *and* to different cases in sentences 1),3), and 4). Although sentences 1) and 3) contain similar expressions (*between ... and ...*), the system recognizes that in the first case the phrase refers to a period of time, while in the second case it refers to origin and destination cases. Moreover, sentence 3) shows that the concept relations origin and destination are not necessarily referred to the origin and destination of a flight, but can be referred to other events, like ground transportation in this case. This sensitivity to the context (to the value of the OBJECT in the example above) shown by certain cases must be taken into account by the module that will interpret the conceptual segmentation and generate the SQL query. In sentence 4) the word *and* is clearly interpreted as connecting two distinct restrictions on the query. The same phenomenon is shown in sentence 5) where the word *or* connects two alternative possible origins of the flight. Table 5 shows examples of incorrect segmentations from the FEB-91 test set. In sentence 1) the phrase *used for Eastern* should be assigned to the airline case. The error is due to the fact that the word *Eastern* was not observed in the training set. In sentence 2) the phrase *through Dallas Fort Worth* should have been labeled with the connect case, but this case has very few examples in the training set

QUERY	RESTRICTION	date
OBJECT		origin
ATTRIBUTE	attribute	destin
	a_date	time
	a_origin	airline
	a_destin	flcode
	a_time	meal
	a_airline	ground
	a_flcode	aircraft
	a_aircraft	class
	a_class	fare
	a_fare	stop
	a_stop	atplace
	a_atplace	dept_time
	a_way	arvl_time
	a_restrict	way
	a_table	restrict
	a_body	table
Q_ATTR		range
AND		speed
DUMMY		body
		day
		connect

Table 1: The set of cases in the ATIS task

QUERY	<i>I would like can I have a list of it give me a description of</i>
OBJECT	<i>the flights the fare on a price on a ticket</i>
origin	<i>arriving from Dallas from Atlanta airport between airport BWI departing Atlanta</i>
destin	<i>and Boston arriving in San Francisco going to San Francisco returning to Atlanta</i>
dept_time	<i>leaving after 1:00 pm that depart in the afternoon</i>
way	<i>round-trip return that are round-trip</i>
class	<i>a class QW ticket a 1st class ticket which have 1st class service available</i>

Table 2: Examples of phrases assigned to cases in the training sentences

TEST	Number of sentences	Sentences correct	Number of cases	Cases correct
JUN-90	98	87 (88.7%)	419	398 (95.0%)
FEB-91	148	119 (80.4%)	713	671 (94.1%)

Table 3: Results with two different test sets

1) Please list all flights between Baltimore and Atlanta on Tuesdays between 4 in the afternoon and 9 in the evening
DUMMY: Please QUERY: list all OBJECT: the flights origin: between Baltimore destin: and Atlanta day: on Tuesdays time: between 4 in the afternoon and 9 in the evening
2) What's the cheapest round-trip airfare on American flight 1074 from Dallas to Philadelphia
QUERY: What's a.fare: the cheapest a.way: round-trip OBJECT: airfare airline: on American flcode: flight 1074 origin: from Dallas destin: to Philadelphia
3) What kind of ground transportation is there between the airport and downtown Atlanta
QUERY: What kind of OBJECT: ground transportation Q_ATTR: is there origin: between the airport destin: and downtown Atlanta
4) What are the restrictions on the cheapest fare from Pittsburgh to Denver and from Denver to San Francisco
QUERY: What are OBJECT: the restrictions fare: on the cheapest fare origin: from Pittsburgh destin: to Denver AND: and origin: from Denver destin: to San Francisco
5) Display flights from Oakland or San Francisco to Denver
QUERY: Display OBJECT: flights origin: from Oakland AND: or origin: San Francisco destin: to Denver

Table 4: Examples of correctly decoded test sentences from FEB-91 test set

1) What kind of aircraft is used for Eastern flight 205
QUERY: What kind of OBJECT: aircraft Q_ATTR: is flcode: used for Eastern flight 205
2) Is there a flight from Denver through Dallas Fort Worth to Philadelphia
QUERY: Is there OBJECT: a flight origin: from Denver through Dallas Fort Worth destin: to Philadelphia
3) Can you please tell me the type of plane that my client would be flying on from Baltimore to Pittsburgh
DUMMY: Can you please QUERY: tell me the type of plane that my client would be OBJECT: flying on origin: from Baltimore destin: to Pittsburgh

Table 5: Examples of incorrectly decoded test sentences from FEB-91 test set

with a consequent poor estimation of the parameters related to it. The same problem, i.e. inadequate training, is also the cause of the wrong segmentation of sentence 3.

Future Work

The goal of the understanding system is to retrieve the information in the ATIS database. In order to do this we are developing a module that translates the conceptual representation of the sentence obtained with the described method into an SQL query. Since the ambiguity of the sentence is resolved by the conceptual segmentation, this module implements a deterministic mapping.

CONCLUSIONS

We proposed a very simple semantic grammar for the ATIS task. The grammar was designed to be rich enough to handle most queries, but limited in certain ways so as to facilitate parsing by very simple and well-understood HMM methods. The advantages of this approach are its straightforward integration with an HMM based speech recognizer, and its capability of learning from examples. Even with an extremely small training set, the system was able to assign the correct analysis to more than 80% of the class A sentences in both the JUN-90 and FEB-91 test sets.

The authors gratefully acknowledge the helpful advice and consultation of Ken Church, Alexandra Gertner, Al Gorin, Fernando Pereira, and Evelyne Tzoukerman.

REFERENCES

- [1] Jelinek, F., "Continuous Speech Recognition by Statistical Methods" *Proceedings of IEEE*, vol. 64, no. 4, pp. 532-556, 1976
- [2] Sowa, J., F. *Conceptual Structures: Information Processing in Mind and Machine*, Addison-Wesley, Reading, MA, 1984.
- [3] Gertner, A. N., Gorin, A. L., Roe, D. B., "Adaptive Language Acquisition from a Subset of the Airline Reservation Task," *paper in preparation*.