SCORING TECHNIQUES FOR PHOENIX PARSES

Paul C. Constantinides

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA. 15213
pcc@cs.cmu.edu

ABSTRACT

This document presents a metric for characterizing semantic frame parses generated from the Phoenix parser. This scoring technique produces measurements indicating coverage, fragmentation and relevance of the parse. These measurements describe the relative structure of the parse, particularly with respect to expectation, allowing comparable measurements between parses. The scoring process is intended as an extension to the implicit scoring method used for parse disambiguation within Phoenix.

1. INTRODUCTION

The Phoenix parser segments and labels strings using a context-free semantic grammar. The tokenized parse consists of a sequence of semantic parse trees, where each tree is rooted by a slot and populated by nets, and each slot belongs to a particular frame. The semantic grammar contains slot and net rewrite rule information. The parser searches the space of parses by maximizing the coverage while minimizing the number of slots, and frames.

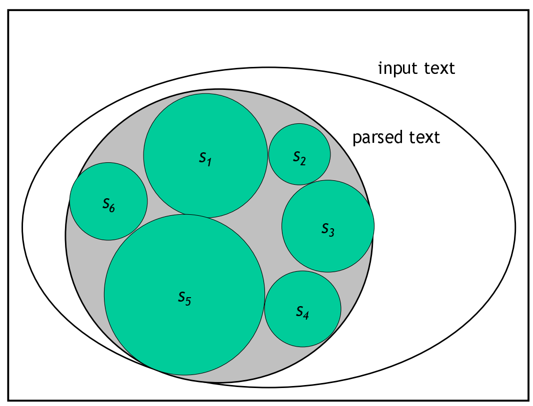
With complicated grammars, it is often possible that a single input string can have several valid, equally scoring top parses. In the travel-planning domain, we have found that spoken phrases often have several parse interpretations. Thus, we recognize the need for descriptive metrics for evaluating parses, to facilitate dialog control. These descriptions should reflect the overall parse quality, and the relevance of the parse and its slots with respect to expectation.

We present a scoring technique as an add-on to the Phoenix parse disambiguation process. This process does not modify the internal Phoenix process for eliminating parses, but rather annotates the top parses with quality and relevance information, allowing the parse consumer to define their own criteria for selecting which is the most germane parse. As a post process, no valid top scoring parses are eliminated from the top list, thus leaving all options open different (often, transient) factors to be weighed against each other.

2. PARSE SCORING

We present for discussion a set of measurements that describe parse structure, including coverage, fragmentation, and germaneness of the phrase and its components. We use features of the parse such as the number of words in the input text, the number of parsed words, the number of slots in the parse, number of nets, etc. to motivate these measurements for scoring. Additionally, relevance scores are motivated by the counts of occurrences of focus tokens. The consumer is given a handle for conveying the current context, allowing for sets of frames, slots, and nets to be brought in to focus. Table 1 enumerates the symbols we will use when describing the scoring functions:

Description	***Symbol***
Total number of words in the input string	t
Number of words accounted for in the parse	p
Number of frames	f
Number of slots	s
Total number of nets in the parse	n_t
Number of nets for slot s	n_s
Number of slots matching focus	m_s
Number of (slot, net) pairs matching focus	m_{sn}
Number of slots belonging to a focus frame	${}_f m_s$
Number of focus slots under particular slot s	m_n^s



2.1 Quality Metrics

Two measurements characterize the overall structure of the parse, indicating the coverage of the parse over the input string, and the fragmentation of the parse. The coverage measurement expresses the proportion of the input string that is accounted for in the parse. This is simply calculated as the ratio of the number words in the parse to the number of words from the input string. A parse with low coverage is characterized as only accounting for a small number of words in the input string. This can be seen graphically in Figure 1 as the size of the set of the parsed text, over the size of the input text, or calculated as:

$$Cov(P) = \frac{p}{t}$$

Another measurement quantifies the fragmentation of the parse, specifically with respect to the slots of the parse. A highly fragmented parse will have many slots where each covers small portions of the parse, whereas a parse with low fragmentation consolidates the coverage into a few number of slots. This is calculated as the ratio of slots per word, or:

$$Frag(P) = \frac{s}{p}$$

These measurements are easily computed from the parse and can be used for rejecting parses. Generally, because of the way Phoenix eliminates parse paths, all remaining parses will have the same *Cov* and *Frag* scores; these scores reflect the parseability of the utterance.

2.2 Relevance Metrics

We've found that different top scoring parses often contain contradictory information with respect to the segmented phrase annotations. Knowledge of the topical context, expectation, narrows the selection of top parses, reducing or eliminating ambiguity or inconsistency. The metrics below account for semantic expectation, to score the parses accordingly.

The parse score consumer can bring into focus sets of frames, slots, and nets, to indicate that these tokens match the consumer's expectation for the semantic interpretation of the parse. Generally, these scores present the proportion of the tokens that match the expectation.

On the entire parse, we check the proportion of the slots that belong to a focus frame, and check the proportion of slots that match the slot expectation. These are calculated as:

$$r_f = \frac{{}_f m_s}{s}, \quad r_s = \frac{m_s}{s}$$

Also, we calculate the proportion of focus nets under focus slots, compared with all of the nets. We limit our net count to those occurring under focus slots, because a net's meaning is modified by its slot parent value. Alternately stated, the same nets may have different meaning when they are children of different slots. This is ratio is calculated as:

$$r_{sn} = \frac{m_{sn}}{n}$$

These ratios characterize the entire parse based on context. It is also possible to evaluate each of the slots in the parse for its contribution to the context. For each particular slots within a parse we can check the proportion of the context matching nets for which this slot is parent:

$$given\ s,\ r_n^s = \frac{m_{sn}}{m_n}$$

This metric can be useful for locating information rich portions of the parse.

3. REFERENCES

1. Ward, W. "Understanding Spontaneous Speech: the Phoenix System", Proceedings of International Conference on Acoustics, Speech, and Signal Processing, May 1991, pp. 365-367.