Deductive parsing and semirings: exercises

7 Oct 2004

Background Pharaoh (Koehn et al., NAACL 2003; Koehn, AMTA 2004) is a decoder for statistical phrase-based translation models. Here is a simplified explanation: the decoder takes as input a French sentence $f = f_1 \cdots f_n$ and translates it into an English sentence via the following steps:

- Segment f into a sequence of *phrases*
$$f = \bar{f}_1 \cdots \bar{f}_m$$
where $\bar{f}_k = f_{i_k+1} \cdots f_{j_k}$ and $0 = i_1 < j_1 = i_2 < \ldots < j_{m-1} = i_m < j_m = n$.
- Rearrange the French phrases into a new pseudo-French sentence
$$f' = \bar{f}_{\pi(1)} \cdots \bar{f}_{\pi(m)}$$
where π is a permutation of $\{1, \ldots, m\}$. This step has score
$$\prod_{k=1}^{m} d(i_{\pi(k)} - j_{\pi(k-1)})$$
where d (for *distortion*) is a function given to the decoder, and $j_{\pi(0)}$ is defined to be 0. The argument to d is just the distance in f from the end of the $(k-1)$st reordered phrase to the beginning of the kth reordered phrase.
- Translate each phrase $\bar{f}_{\pi(k)}$ into an English phrase $\bar{e}_k$ drawn from a set $\mathcal{E}$. This step has score
$$\prod_{k=1}^{m} \phi(\bar{f}_{\pi(k)}, \bar{e}_k)$$
where ϕ (for *phrase*) is again a function given to the decoder.

The decoder considers all possible segmentations, rearrangements, and translations, and outputs the best result.

For example, it might take the French sentence

Marie mange souvent les pommes,

segment it into

(Marie) (mange) (souvent) (des pommes),

reorder it as

(Marie)	(souvent)	(mange)	(des pommes)
$d(0)$	$d(1)$	$d(-2)$	$d(1)$

(with the distortion scores shown underneath), and finally translate it as

(Mary)	(often)	(eats)	(apples)
$\phi(\text{Marie}, \text{Mary})$	$\phi(\text{souvent}, \text{often})$	$\phi(\text{mange}, \text{eats})$	$\phi(\text{des pommes}, \text{apples})$

(with the translation scores shown underneath).

Exercise Now we present a deductive system for this simplified Pharaoh. The general form of an item is $[F, j]$, where $F \subseteq \{1, \ldots, n\}$ indicates which French words have been translated, and j means that f_j was the last French word to have been translated. Then the system is as follows:

$$
\begin{array}{lcl}
\text{Axiom} & [\emptyset, 0] & \\[1em]
\text{TranslatePhrase} & \dfrac{[F, j]}{[F \cup \{i', \ldots, j'\}, j']} & \begin{array}{l} j' \geq i',\ F \cap \{i', \ldots, j'\} = \emptyset,\ \bar{e} \in \mathcal{E} \\ \text{value: } \phi(f_{i'+1} \cdots f_{j'}, \bar{e}) d(i' - j - 1) \end{array} \\[1em]
\text{Goal} & [\{1, \ldots, n\}, j] &
\end{array}
$$

The value given for TranslatePhrase is for the Viterbi semiring.

1. What is the time/space complexity of this system without pruning?
2. How should we bucket the items (Goodman, pp. 597–8)? Are there any looping buckets?
3. All the system does so far is output "yes" or "no" with a score, which isn't very useful. Write down the semiring that reconstructs the English sentences.
4. The real Pharaoh multiplies in a trigram language model for English,

$$\prod_{i=1}^{n'+1} p_{\text{LM}}(e_i \mid e_{i-2} e_{i-1})$$

   where $e_0 = \textbf{START}$, $e_{-1} = \epsilon$, and $e_{n'+1} = \textbf{STOP}$. Make our system do the same. Hint: you have to add some information to the items.
5. The real Pharaoh can also output a tuple of separate component scores (ϕ, d, p_{LM}). Write down the semiring to do this.
6. Something the real Pharaoh doesn't do but should is compute a list of the k best translations with their component scores. (This is useful for finding optimal weights for the components.) Write the semiring to do this.