1 Semiring formulation

So yeah, the highlights are: an edge connects three nodes, a parent and two CHILDREN, each node is a labelled SPAN; you need to identify the scoring function for an edge, let's call it $w(e)$, in this case we have

$$w(e) = f(\textsc{head}(e)) \bigotimes_{c \in \textsc{children}(e)} g(\textsc{span}(c)) \tag{1}$$

where f and g are parametric functions; then you can compute the Inside recursion for a node v

$$I(v) = \bigoplus_{e \in BS(v)} w(e) \otimes \bigotimes_{c \in \textsc{children}(e)} I(c) \tag{2}$$

where I'm using $BS(v)$ to denote the set of edges incoming to v; note that BS here basically enumerates the different ways to segment the string under (i, j) into two adjacent parts and the different labels of each child SPAN (let's call these a and b, each an element in the labelset L), thus we can write

$$I(v = [i, j, l]) = \bigoplus_{\substack{e=[i,j,l,k,a,b]: \\ a \in L, \\ b \in L, \\ k \in \{i+1, \ldots, j-1\}}} w(e) \otimes I([i, k, a]) \otimes I([k+1, j, b]) \tag{3}$$

Now the key is to realise that $w(e)$ factorises and therefore we can rewrite this as

$$I(v = [i, j, l]) = f(i, j, l) \otimes \bigoplus_{k=i+1}^{j-1} g(i, k) \otimes g(k+1, j) \tag{4}$$

$$\otimes \bigoplus_{a \in L} I([i, k, a]) \tag{5}$$

$$\otimes \bigoplus_{b \in L} I([k+1, j, b]) \tag{6}$$

and this finally motivates having an inside table for the SPANs (with labels summed out), let's call that

$$S(i, j) = \bigoplus_{l \in L} I(i, j, l) \tag{7}$$

and then we have the result

$$I(i, j, l) = f(i, j, l) \otimes \bigoplus_{k=i+1}^{j-1} g(i, k) \otimes g(k+1, j) \otimes S(i, k) \otimes S(k+1, j). \tag{8}$$

2 Alternative formulation

In this derivation we follow Michael Collins notes on the Inside-Outside Algorithm.darkblue[1]

Let a sentence be $x_1, \ldots, x_n$, where each x_i is a word. We are given a CFG (N, Σ, R, S) in Chomsky normal form. Let ψ be a function that maps any rule which is how I derived that **??** is the correct formula for the rule score.

We obtain our CRF objective when we normalize this score globally

$$P(T) = \frac{\prod_{r \in T} \psi(r)}{\sum_{T' \in \mathcal{T}} \prod_{r' \in T'} \psi(r')} \tag{16}$$

$$\tag{17}$$

or equivalently

$$\log P(T) = \sum_{r \in T} \log \psi(r) - \log \sum_{T \in \mathcal{T}} \prod_{r \in T} \psi(r) \tag{18}$$

$$\tag{19}$$

From the aforementioned notes we get the following general result for the inside value α. For all $A \in N$, for all $0 \leq i < n$

$$\alpha(A, i, i+1) = \psi(A, i, i+1) \tag{20}$$

and for all (i, j) such that $1 \leq i < j \leq n$:

$$\alpha(A, i, j) = \sum_{A \to BC} \sum_{k=i+1}^{j-1} \psi(A \to B\ C, i, k, j) \cdot \alpha(B, i, k) \cdot \alpha(C, k, j) \tag{21}$$

Note that we are considering a CFG in which the rule set is complete, i.e.

$$\langle A \to B\ C \rangle \in R \text{ for each } (A, B, C) \in N^3, \tag{22}$$

and recall that the labels B and C do not appear in the scoring functions in **??**. These facts will allow us to simplify the expression in formula **??** as

$$\alpha(A, i, j) = \sum_{B \in N} \sum_{C \in N} \sum_{k=i+1}^{j-1} \tilde{s}_{label}(i, j, A) \cdot \tilde{s}_{span}(i, j) \alpha(B, i, k) \cdot \alpha(C, k, j) \tag{23a}$$

$$= \tilde{s}_{label}(i, j, A) \cdot \tilde{s}_{span}(i, j) \sum_{k=i+1}^{j-1} \sum_{B \in N} \alpha(B, i, k) \cdot \sum_{C \in N} \alpha(C, k, j) \tag{23b}$$

$$= \tilde{s}_{label}(i, j, A) \cdot \tilde{s}_{span}(i, j) \sum_{k=i+1}^{j-1} S(i, k) \cdot S(k, j) \tag{23c}$$

where we've introduced a number of notational abbreviations

$$\tilde{s}_{label}(i, j, A) = \exp(s_{label}(i, j, A)) \tag{24}$$

$$\tilde{s}_{span}(i, j) = \exp(s_{span}(i, j)) \tag{25}$$

$$S(i, j) = \sum_{A \in N} \alpha(A, i, j) \tag{26}$$

Note that this is the exact same formula as **??**.

From equation **??** we can deduce that we in fact do even need to store the values $\alpha(i, j, A)$ but that it suffices to only store the marginalized values $S(i, j)$. In this case, the recursion simplifies even further:

$$\begin{aligned} S(i,j) &= \sum_{A \in N} \alpha(A, i, j) && \text{(27a)} \\ &= \sum_{A \in N} \tilde{s}_{label}(i, j, A) \cdot \tilde{s}_{span}(i, j) \sum_{k=i+1}^{j-1} S(i,k) \cdot S(k,j) && \text{(27b)} \\ &= \left[\sum_{A \in N} \tilde{s}_{label}(i, j, A) \cdot \tilde{s}_{span}(i, j) \right] \left[\sum_{k=i+1}^{j-1} S(i,k) \cdot S(k,j) \right] && \text{(27c)} \end{aligned}$$

where we put explicit brackets to emphasize that independence of the sub-problems of labeling and splitting. We can now recognize this as the 'inside' equivalent of the expression from the paperdarkblue[2]

$$s_{best}(i, j) = \max_{\ell}[s_{label}(i, j, \ell)] + \max_{k}[s_{split}(i, k, j)]. \tag{28}$$

The recursions are the same; the semirings are different. The viterbi recursion given above is in the ViterbiSemiring, which uses the max operator as $\oplus$; the inside recursion given in **??** has standard addition (+) instead.

[2]I believe there is actually an error in this equation: it should read $s_{label}(i, j, \ell) + s_{span}(i, j)$ instead of just $s_{label}(i, j, \ell)$. This is implied by the score for a single node, which is given by equation **??**, taken directly from the paper.

3 Outside

$$
\begin{aligned}
\beta(A,i,j) &= \sum_{B \to CA \in R} \sum_{k=1}^{i-1} \psi(B \to CA, k, i-1, j) \cdot \alpha(C,k,i-1) \cdot \beta(B,k,j) \\
&\quad + \sum_{B \to AC \in R} \sum_{k=j+1}^{n} \psi(B \to A, C, i, j, k) \cdot \alpha(C, j+1, k) \cdot \beta(B,i,k) \\
&= \sum_{B \in N} \sum_{C \in N} \sum_{k=1}^{i-1} \psi(B,k,j) \cdot \alpha(C,k,i-1) \cdot \beta(B,k,j) \\
&\quad + \sum_{B \in N} \sum_{C \in N} \sum_{k=j+1}^{n} \psi(B,i,k) \cdot \alpha(C,j+1,k) \cdot \beta(B,i,k) \\
&= \sum_{k=1}^{i-1} \left[\sum_{B \in N} \psi(B,k,j) \cdot \beta(B,k,j) \right] \cdot \left[\sum_{C \in N} \alpha(C,k,i-1) \right] \\
&\quad + \sum_{k=j+1}^{n} \left[\sum_{B \in N} \psi(B,i,k) \cdot \beta(B,i,k) \right] \cdot \left[\sum_{C \in N} \alpha(C,j+1,k) \right] \\
&= \sum_{k=1}^{i-1} S'(k,j) \cdot S(k,i-1) + \sum_{k=j+1}^{n} S'(i,k) \cdot S(j+1,k)
\end{aligned}
$$

where

$$
\begin{aligned}
S(i,j) &= \sum_{A \in N} \alpha(A,i,j) \\
S'(i,j) &= \sum_{A \in N} \psi(A,i,j) \beta(A,i,j)
\end{aligned}
$$