Semantic tree operations

D. Gueorguiev 1/6/2021

Semantic subtree expansion and node comparison

We are considering m-ary tree T which is a tree in which each node has at most m children. Let us introduce the *tree tuple* (k, v) where v represents a *tree node* which is a semantic particle or a *subtree* of nodes (i.e. semantic particles).
Here k is a tree tuple factor which encodes uniquely the position of the node or the root of the subtree v in the parent tree T. More precisely, the tuple factor k encodes uniquely the path from the root of the tree to the node v or the root of subtree associated with k. Each tuple factor k representing the root can be decomposed into an ordered set of primitive tuple factors $\{k_1, k_2, \ldots, k_m\}$

Definition: *Arc* of semantic tree

The following operations are defined for tree tuple factors:

Multiplication operation for semantic tuple factors

One possible implementation for the primitive tuple factors $k_1, k_2, \ldots\ k_m$ is to define them as the base of m digits greater than 0 of $(m + 1)$-nary number system such that $0 = k_0 < k_1 < k_2 < \ldots < k_m$.
We define an operation `$*$` denoting digit concatenation $k_i * k_j = (m + 1)k_i + k_j$. Obviously,
$k_i * k_j > k_i$ for any pair $i, j = 1..m$
Note that the latter implies that
$k_{i_1} * k_{i_2} * \ldots * k_{i_n} > k_{j_1} * k_{j_2} * \ldots * k_{j_{n-1}}$ for any tuple where $i_p, j_q = 1..m, p = 1..n, q = 1..n - 1$

$$\left(k_i, \left(k_j, v_j\right)\right) = (k_i * k_j, v_j)$$

Encoding a complete m-ary tree T of height h with the algebraic notation above:
$T = (k_0, v_0^0)$. Further we will assume that $k_0 = 0$.
$v_0^0 = (k_0, v_0^1) + (k_1, v_1^1) + (k_2, v_2^1) + \ldots + (k_m, v_m^1)$

In general, we have:
$v_q^p = \left(k_0, v_{(q-1)m}^{p+1}\right) + \left(k_1, v_{(q-1)m+1}^{p+1}\right) + \left(k_2, v_{(q-1)m+2}^{p+1}\right) + \ldots + \left(k_m, v_{qm}^{p+1}\right)$
where $q = 1..m^h, p = 1..h$
Obviously, we have at most $\frac{(m^{h+1}-1)}{m-1}$ distinct terms v_q^p which represent nodes i.e. semantic values.

Tuple factor comparison operator

The expression for the tree also can be written as:
$T = \sum_{i=0}^{N}(k_i^*, v_i)$ where $N \le \frac{(m^{h+1}-1)}{m-1}$ and k_i^* are the *node factors* given with $k_i^* = k_{i_1} * k_{i_2} * \ldots * k_{i_n}; n \le h$. The node values v_i are the values v_q^p ordered in increasing order of k_i^*. This order

corresponds to *level order traversal* of the m-ary tree. Note that with appropriately defined comparison operation `<` we can model different ways of traversing the m-ary tree. For instance, if we define `<` as the comparison for the values of $(m+1-k_{i_1})m^{n-1} + (m+1-k_{i_2})m^{n-2} + \cdots + (m+1-k_{i_n})$ we will have ordering which corresponds to the *preorder traversal* of the tree.

Weighted semantic trees

<u>*Example*</u>
Peter is Dimitar's son.
Dimitar's son has a friend in the neighborhood and his friend's name is James.

- ⇨ *James is Peter's friend*

Peter is the son of Dimitar.
V_1 V_2 V_3 V_4 V_5 V_6
The son of Dimitar has a friend in the neighborhood and the name of his friend is James.
V_7 V_8 V_9 V_{10} V_{11} V_{12} V_{13} V_{14} V_{15} V_{16} V_{17} V_{18} V_{19} V_{20} V_{21} V_{22} V_{23} V_{24}

T_1

V_2
/ \
V_1 V_3
/ | \
V_4 V_5 V_6

T_2

________V_{17}________
/ \
V_{11}________ V_{23}
/ \ \ / \
V_8 V_{12} V_{14} V_{19} V_{24}
/ | \ | / \ / \
V_7 V_9 V_{10} V_{13} V_{15} V_{16} V_{18} V_{20}
/ \
V_{21} V_{22}

T

V
/ \
V_{24} V_{13}
/ | \
V V V_1

Expressing T_1 with the algebraic notation discussed earlier:
$T_1 = (k_0, v_2) + (k_1, v_1) + \big(k_2, \big((k_0, v_3) + (k_1, v_4) + (k_2, v_5) + (k_3, v_6)\big)\big)$
which is expanded to:
$T_1 = (k_0, v_2) + (k_1, v_1) + (k_2 k_0, v_3) + (k_2 k_1, v_4) + (k_2 k_2, v_5) + (k_2 k_3, v_6)$

Expressing T_2 with the algebraic notation yields:

$$T_2 = (k_0, v_{17}) + \left(k_1, \left((k_0, v_{11}) + \left(k_1, \left((k_0, v_8) + (k_1, v_7) + (k_2, v_9) + (k_3, v_{10})\right)\right) + \left(k_2, \left((k_0, v_{12}) + (k_1, v_{13})\right)\right) + \left(k_3, \left((k_0, v_{14}) + (k_1, v_{15}) + (k_2, v_{16})\right)\right)\right)\right) + \left(k_2, \left((k_0, v_{23}) + \left(k_1, \left((k_0, v_{19}) + (k_1, v_{18}) + \left(k_2, \left((k_0, v_{20}) + (k_1, v_{21}) + (k_2, v_{22})\right)\right)\right)\right) + (k_2, v_{24})\right)\right)$$

which is expanded to:

$$\begin{aligned} T_2 = {} & (k_0, v_{17}) + (k_1 k_0, v_{11}) + (k_1 k_1 k_0, v_8) + (k_1 k_1 k_1, v_7) + (k_1 k_1 k_2, v_9) + (k_1 k_1 k_3, v_{10}) \\ & + (k_1 k_2 k_0, v_{12}) + (k_1 k_2 k_1, v_{13}) + (k_1 k_3 k_0, v_{14}) + (k_1 k_3 k_1, v_{15}) + (k_1 k_3 k_2, v_{16}) \\ & + (k_2 k_0, v_{23}) + (k_2 k_1 k_0, v_{19}) + (k_2 k_1 k_1, v_{18}) + (k_2 k_1 k_2 k_0, v_{20}) + (k_2 k_1 k_2 k_1, v_{21}) \\ & + (k_2 k_1 k_2 k_2, v_{22}) + (k_2 k_2, v_{24}) \end{aligned}$$

Semantic Tree Difference

Let us have two trees represented as:

$T = \sum_{i=0}^{N} (k_i, v_i)$ and $T^* = \sum_{i=0}^{N} (k_i^*, v_i)$
Here $v_i, i = 1..N$ denote the nodes of the two trees (not subtrees) which are semantic particles with signatures $ssig(v_i)$. The sequences k_i and k_i^* denote the sequences of tuple factors which encode the position of each node v_i in each of the two trees.

Definition: Matching arcs of trees

Let us assume that the two trees are weighted so that for each tree there is weight function which maps each tuple factor to a real number which is the weight corresponding the arc associated with the specified tuple factor. Let us denote by f and f^* the two weight functions corresponding to T and T^*. Generally, $f(k_i) \neq f^*(k_i)$. If $f \equiv f^*$ then T and T^* have the same weights on their matching arcs.

We have the same set of semantic particles but they are arranged differently in two trees. We would like to define metric how different are the two trees.