The Concept of Semantic Space

D. Gueorguiev, 10/5/21

The Concept of Semantic Space

Every V-particle is adorned with a semantic signature which can be represented by a DAG where each node contents can be represented as an integer number and each arc is assigned a weight which is a positive real number. Equivalently, the semantic signature can be represented as a matrix (refer to the document The Signature of Semantic Structures for details). Every semantic structure is represented by its semantic signature. The weights of the arcs in its semantic signature represent confidence values which are not normalized. The Semantic space is a metric space where the metric (norm) is the semantic distance denoted by $sdist$. The semantic distance is defined recursively as:

Let us denote the semantic distance between two particles V_{a_1} and V_{b_1} with $sdist(V_{a_1}, V_{b_1})$.
Let us evaluate the semantic distance between V_{a_1} and V_{b_1}.

$$\begin{array}{c} V_{a_1} \\ / \, w \\ V_{a_2} \end{array}$$

Here V_{a_1} is connected to V_{a_2} with an arc having a weight w. Let us assume that the $sdist\left(V_{a_1}, V_{b_1}\right) < \varepsilon$ where ε is a small positive number. Let us denote the new compound V particle with $V_{new} = [V_{a_1} \overset{w}{\rightarrow} V_{a_2}]$. We want the following asymptotic behavior to hold true when we make the weight arbitrary small:

$$sdist(V_{new}, V_{a_1}) < \varepsilon \text{ when } w \rightarrow 0$$

We want also $sdist(V_{new}, V_{a_1}) < \varepsilon$ when $sdist(V_{a_2}, V_{\emptyset})$ is small enough. Here $V_{\emptyset}$ represents the null semantic particle which has no meaning i.e. it is arbitrarily close in terms of semantic distance to any other semantic structure or particle. The last asymptotic relation is equivalent to disregarding V-particles which do not enrich the semantic structure of the resulting compound particle.
Example:

$$text(V_{new}) = "Yes, he\ is\ Dimitar, yup"$$
$$text\left(V_{a_1}\right) = "Yes, he\ is\ Dimitar", \ text\left(V_{a_2}\right) = "yup"$$

Possible ways to define semantic distance and equivalence between the semantic DAG G and signature matrix S

We want to create the signature matrix S from the semantic DAG G in such way that we preserve the asymptotic closeness properties of the semantic space defined earlier. Let us define the signature matrix of V particle in such a way that each row corespond to a property from the property graph $G(V)$ of V particle traversed *in order*. For details on the Property Graph of semantic particle consult *Properties and Dependent Properties* paragraph in the document Inference and Execution.
Using the definition of signature of semantic structure we will define ε-closeness of two semantic structures. The semantic structures S1 and S2 are ε-close if there exist a difference matrix $D = ssig(S_1) -$

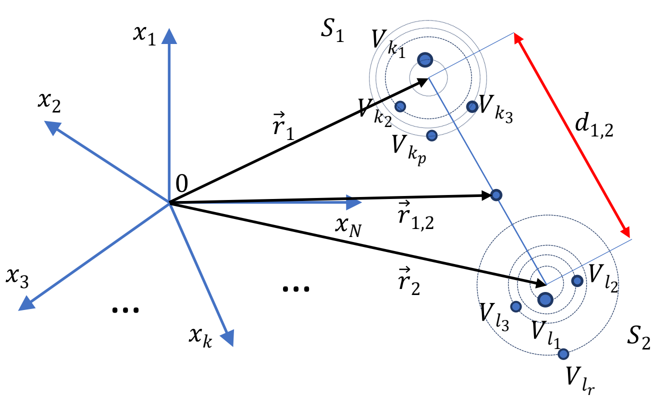
Notes on Semantic Distance

Semantic distance between two structures is a measure of how close those two structures are semantically. ~~One criterion measuring the structural alignment is the connectedness of each structure to the enclosing context.~~ If the two structures are to be close semantically with respect to the enclosing context the connections between each structure and the enclosing context should be similar. The other obvious criterion is the similarity of the semantic signatures. If the two structures are to have short semantic distance the semantic signatures should be similar. Here we should elaborate on an important detail of the last criterion - the semantic signatures of which structures should be compared? The signatures of the ***augmented*** structures are those which matter and those which should be compared. The augmented structure of a given semantic structure S relative to some context C is obtained by replacing all inbound similarity links, denoted by $SA^{-}(S)$ with the structures they refer to within the context C. For more details on the similarity links see the paragraph *The notion of effective mass of a semantic structure* in the document Connecting Semantically Related Structures.

Let us study the following problem-
We have an enclosing semantic structure (context) C_1 and we have associated with C_1 two separate semantic structures S_1 and S_2. Let the semantic structure S_1 is represented by the DAG G_1 composed of p V-particles and q A-particles. The set of the V-particles in G_1 will be denoted by $\mathfrak{S}_V(G_1) = \left\{V_{k_1}, V_{k_2}, \dots, V_{k_p}\right\}$. The set of the A-particles in in G_1 will be denoted by $\mathfrak{S}_A(G_1) = \left\{A_{k_1}, A_{k_2}, \dots, A_{k_q}\right\}$.
Let the semantic structure S_2 is represented by the DAG G_2 composed of r V-particles and s A-particles. The set of the V-particles in G_2 will be denoted by $\mathfrak{S}_V(G_2) = \{V_{l_1}, V_{l_2}, \dots, V_{l_r}\}$. The set of the A-particles in in G_2 will be denoted by $\mathfrak{S}_A(G_2) = \{A_{l_1}, A_{l_2}, \dots, A_{l_s}\}$.
Let us for a moment assume that neither S_1 nor S_2 have outbound or inbound similarity links i.e.

$$SA^{+}(S_1) = SA^{-}(S_1) = SA^{+}(S_2) = SA^{-}(S_2) = \emptyset$$

Then we define the semantic distance between S_1 and S_2 to be the distance between *the semantic centers of masses* of S_1 and S_2. For discussion on a semantic mass center of particles and structures refer to On The Semantic Position of Properties, Primitive Particles, and Semantic Structures.



Let us denote with $\vec{r}_1$ the semantic mass center of the structure S_1 and with $\vec{r}_2$ the semantic mass center of S_2. For the vector $\vec{r}_1$ we have the following expression:

$\vec{r}_1 = \sum_{k=k_1}^{k_p} \frac{M_{V_k}}{M_{S_1}} \vec{r}_k$. Here M_{V_k} denotes the semantic mass of particle V_k from S_1. M_{S_1} denotes the total semantic mass of the structure S_1 and it is given with $M_{S_1} = \sum_{k=k_1}^{k_p} M_{V_k}$. Obviously, the vector $\vec{r}_k$ represents the semantic position of the particle V_k after it was added to the ensemble and has moved toward the center of mass over a distance d_k. Recall, $|\vec{r}_1 - \vec{r}_k| = \varepsilon_k$ where $\varepsilon_k \sim C{M_{V_k}}^{-\alpha}, \alpha > 0$. The original position of the particle V_k before it was added to the ensemble is $\vec{r}_k^0 = \vec{r}_1 + \frac{\vec{r}_k - \vec{r}_1}{|\vec{r}_k - \vec{r}_1|}(d_k + \varepsilon_k)$.

//TODO: finish this