Chapter 1

The Credibility Value of the Similarity Relations

In chapter **??** the *SBM* approach[**?**] itself and its shortcomings were discussed in detail. Our new methodology was introduced in chapter **??** which overcomes the flows of the preceding approach via the path it adopts for representing the knowledge base, the algorithm it embraces in order to search the data, and the structured function it contains for evaluating the result.

With all its promising traits, the new methodology is not without flaws. Mind that when calculating the final degree of similarity between two predicates, solely the predefined similarity relations on subconcepts are taken into account in the evaluation algorithm. However in real world applications, knowledge bases do not always come with complete information. Thus a notion that emphasizes the approximate error in the accuracy of the result of the algorithm should be introduced, for the sake of the highly probable possibility of encountering missing information.

Example 1.0.1. Consider the following example:

$$
\begin{aligned}
&decent\_child: &&(Person) \\
&eats\_well: &&(Feature) \\
&does\_sports: &&(Feature) \\
&tidies\_room: &&(Feature) \\
&studies\_hard: &&(Feature) \\
&respects\_elders: &&(Feature) \\
&role\_model\_youngster: &&(Person) \\
&good\_dietary: &&(Feature) \\
&fit: &&(Feature) \\
&neat: &&(Feature) \\
&good\_grades: &&(Feature) \\
&obedient\_toParents: &&(Feature)
\end{aligned}
$$

$$decent\_child(X) \quad \overset{1.0,..}{\longleftarrow} prod \quad eats\_well(X), does\_sports(X), tidies\_room(X),$$
$$studies\_hard(X), respects\_elders(X).$$
$$role\_model\_youngster(X) \quad \overset{1.0,..}{\longleftarrow} prod \quad good\_dietary(X), fit(X), neat(X),$$
$$good\_grades(X), obedient\_toParents(X).$$

$$Sim(tidies\_room, neat) = 0.8$$

$$Sim(respects\_elders, obedient\_toParents) = 0.9$$

When the proximity of similarity between the predicates *decent_child* and *role_model_youngster* is queried, our algorithm take into consideration only the two predefined similarity relations between the subconcepts. Namely, the ones between *tidies_room* and *neat*, and *respects_elders* and *obedient_toParents*.

However if one reasons on the real-world semantic meanings of the predicates, s/he will realize that the subconcepts pairs

- *eats_well, good_dietary*

- *studies_hard, good_grades*

- *does_sports, fit*

seem to be immensely related, or in other terms *similar* to each other. So indeed one in this case suspect the knowledge base to be missing information. With this motivation, we propose three possible models for evaluating a credibility approximation for our algorithm in the following sections. Namely:

1. Vertex Approach

2. Edge Approach

3. Hybrid Approach

In section **??** we have seen that our method has the means of representing similarity relations with fuzzy credibility values. With that idea in mind, here we will propose the following:

Assume that for two predicates p_0 and p_1, our algorithm concludes the similarity value S with the corresponding credibility value C. Then the following fuzzy rule is added into the knowledge base when our method terminates.

$$Sim(p_0, p_1) \quad \overset{C}{\longleftarrow} S.$$

1.0.1 Vertex Approach

The most naive method for computing an estimate of credibility for our problem seems to be finding the percentage of the subconcepts which appear in a similarity relation. The equation is as follows:

$$Credibility = \frac{\#\_of\_subconcepts\_participating\_in\_a\_similarity\_relation}{\#\_of\_total\_subconcepts} \tag{1.1}$$

In the previous example there were five subconcepts for each of the predicates, and two from each side were included in a similarity relation. Hence the degree of credibility of our algorithm for the case of example 1.0.1 is going to be:

$$\begin{aligned} Credibility &= \frac{\#\_of\_subconcepts\_participating\_in\_a\_similarity\_relation}{\#\_of\_total\_subconcepts} \\ &= \frac{4}{10} \\ &= \mathbf{0.4} \end{aligned} \tag{1.2}$$

Then this means the credibility value that we attain to our result is *0.4* for this example. The evaluation function would conclude the similarity degree between the predicates *decent_child* and *role_model_youngster* as *0.85* hence in the end the following fuzzy rule would be added into the knowledge base:

$$Sim(decent\_child, role\_model\_youngster) \xleftarrow{0.4} 0.85.$$

1.0.2 Edge Approach

There seems to be one point of concern with the approach in section 1.0.1. We want to kindly remind the reader that in section **??** we had shown that our algorithm was fully able to handle the cases where a specific combination of concepts are included in more than one similarity relations.

Example 1.0.2. Suppose that we extend the example 1.0.1 with the following two similarity relations:

$$Sim(tidies\_room, obedient\_toParents) = 0.6$$

$$Sim(respects\_elders, neat) = 0.55$$

Considering the existing similarity relations, now we have four in total. Another important point that demands attention is that four subconcepts have been included in more

than one fuzzy similarity relations. All in all at this points we have more information with respect to the first version of the problem. Thus we would expect the credibility that we attain to our estimation to be bigger. Let us do the calculation with section 1.0.1 approach once again:

$$
\begin{aligned}
Credibility &= \frac{\#\_of\_subconcepts\_participating\_in\_a\_similarity\_relation}{\#\_of\_total\_subconcepts} \\
&= \frac{4}{10} \\
&= \mathbf{0.4}
\end{aligned}
\tag{1.3}
$$

As can be seen from the results, nothing has changed. That is because the prior approach only considers the number of predicates that are contained in similarity rules, or vertices in graph terms. It does not able to realize the information of the additional info introduced by the proximity of similarity between different pairings of these nodes, or speaking in graph terms, number of added edges between the two trees. Thus here we propose a second equation which would handle this shortcoming.

$$
Credibility = \frac{\#\_of\_similarity\_pairings\_between\_subconcepts}{\#\_of\_possible\_similarity\_pairings\_between\_subconcepts}
\tag{1.4}
$$

Let us compute the results of both versions of the problem via utilizing this approach respectively.

$$
\begin{aligned}
Credibility &= \frac{\#\_of\_similarity\_pairings\_between\_subconcepts}{\#\_of\_possible\_similarity\_pairings\_between\_subconcepts} \\
&= \frac{2}{5 \times 5} \\
&= \mathbf{0.08}
\end{aligned}
\tag{1.5}
$$

$$
\begin{aligned}
Credibility &= \frac{\#\_of\_similarity\_pairings\_between\_subconcepts}{\#\_of\_possible\_similarity\_pairings\_between\_subconcepts} \\
&= \frac{4}{5 \times 5} \\
&= \mathbf{0.16}
\end{aligned}
\tag{1.6}
$$

As expected the second result proved to be bigger, which means the approach was able to attain a higher credibility value for the case where the knowledge base consisted more information.

The only problem with this approach appears to be the big decrease in all of the credibility values compared to the first approach. Here we are dividing the number of existing

similarity pairings, to the number of all possible pairings which actually proves to be a relatively huge number. And indeed in reality we wouldn't normally expect for every subconcept to be related, or similar to every other subconcept from the second predicate tree. Thus it seems that this equation seems to go under some sort of normalization process in order to depict more accurate results. This is the motivation for the last approach which is described in section 1.0.3.

1.0.3   Hybrid Approach

As a quick recap, let us see what are the advantages and disadvantages of the approaches presented in sections 1.0.1 and 1.0.2.

	PROS	CONS
Vertex Approach	Realistic results	No means to differentiate shared concepts
Edge Approach	Utilizes complete info	Unrealistic results

Vertex Approach evaluates results that are in harmony with the common sense. However it can not realize the extra knowledge gained by the shared concepts. On the other hand, *Edge Approach* is able to tackle this problem, whereas the credibility values it returns are dramatically lower than the expected ones.

Somehow we would like to come up with an algorithm which would have both approaches' strong points, meanwhile it should avoid the shortcomings of the two. In the light of this motivation, we propose a hybrid approach which merges two approaches together by attaining them some weights:

$$Credibility = (w \times [\textbf{Vertex\_Approach}]) + ((1 - w) \times [\textbf{Edge\_Approach}]) \qquad (1.7)$$

where $w \in [0..1]$.

The intuition tells to select w relatively high (*i.e. bigger then 0.5*) as the first approach was the one with more realistic results, but we wanted to integrate the second approach as we wanted to make use of all the info that the knowledge base contain.

Hence with the assumption of $w = 0.8$, let us calculate both versions of the problem with the latest approach:

$$Credibility = (w \times [\textbf{Vertex\_Approach}]) + ((1 - w) \times [\textbf{Edge\_Approach}])$$
$$= (0.8 \times \frac{4}{10}) + (0.2 \times \frac{2}{5 \times 5})$$
$$= \textbf{0.336}$$

(1.8)

$$Credibility = (w \times [\textbf{Vertex\_Approach}]) + ((1 - w) \times [\textbf{Edge\_Approach}])$$
$$= (0.8 \times \frac{4}{10}) + (0.2 \times \frac{4}{5 \times 5})$$
$$= \textbf{0.352}$$

(1.9)

Even though it's small, we are able to differentiate the effect of the added knowledge from the distinct results. Moreover final values seem to be meaningful, on contrary of the case in section 1.0.2.

All in all, with the help of our advanced similarity and credibility approximation algorithms, confidently we would be able to add the following fuzzy rules into the corresponding knowledge bases.

$$Sim(decent\_child, role\_model\_youngster) \overset{0.336}{\longleftarrow} 0.85.$$

$$Sim(decent\_child, role\_model\_youngster) \overset{0.352}{\longleftarrow} 0.713.$$

1.0.4   Current Directions

As seen in the preceding section, we adopt an intuitionistic way for defining the weights in the credibility evaluation equation of the *Hybrid Approach*. This is a naive first look to an intriguing potential area of research.

In *RFuzzy* framework [**?**], they utilize Multi Adjoint-Logic in order to compute the credibility values from real-world information in an automized fashion. The focus of our ongoing work is centered on this promising idea of developing an atomization method for attaining weights to the evaluation function, thus computing credibility values directly from the examples of real-world data. With this inspiration, in the following chapter we introduce our automized methodology for the concerning problem, illustrating it on a real-world case. There exist many more approaches in other scientific fields that we could observe as a source of inspiration like the methods utilized in artificial neural networks.