Edge dependent pathway scoring for calculating semantic similarity in ConceptNet

Steve Spagnola
Cornell University
sps34@cs.cornell.edu

Carl Lagoze
Cornell University
lagoze@cs.cornell.edu

Abstract

Most techniques that calculate the relatedness between two concepts use a semantic network, such as Wikipedia, WordNet, or ConceptNet, to find the shortest intermediate pathway between two nodes. These techniques assume that a low number of edges on the shortest pathway indicates conceptual similarity. Although this technique has proven valid in conforming to psychological data, we test the usefulness of additional pathway variables in ConceptNet, such as edge type and user-rated score. Our results show strong evidence for the application of additional pathway variables in calculating semantic similarity.

1 ConceptNet Pathways

ConceptNet 3 is one of the largest commonsense semantic networks in existence, relying on its users to make conceptual assertions and collectively vote on the legitimacy of other users' assertions. ConceptNet is valuable as a semantic resource because it suggests transitive inference between ideas, enabling dissimilar concepts to share a semantic, indirect relationship. Unlike Wikipedia and WordNet, the edges in ConceptNet contain additional semantic information between two concepts. Each edge is assigned a relation type (such as "Is A" or "Located At") and a score that correlates to how well ConceptNet users believe in the validity of the relation [Havasi and Alonso (2007)]. Previous work on calculating semantic relatedness between two concepts ignores these extra edge features, using only the inverse of number of edges on a short path [Rada and Blettner (1989); Wubben and A. (2009)]. Instead of only looking for the shortest pathway from one concept to another we assess all pathways in measuring the semantic similarity of an association. A simple example is shown in Figure 1: two nodes (cat and dog) with two pathways containing varying intermediary nodes and edge types.

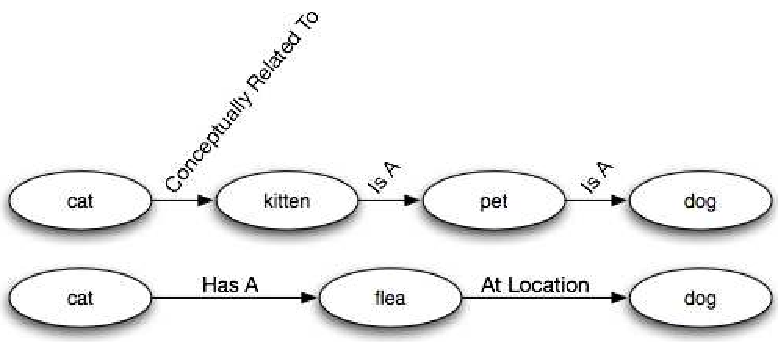

Figure 1: Transitive links between nodes in ConceptNet may occur through a variety of pathways, some being more appropriate than others.

Each possible pathway from one concept to another contains a set of edges, where each edge has a predefined edge type and validity score. Hence, in addition to edge count, we can use the sum of edge scores within a pathway as an additional feature. It may be the case that some pathways have a low number of edges (implying a high relatedness metric by inverse edge count), yet only contain poorly scoring edges that are incorrect or noisy inferences, yielding low aggregate score. In such cases the simple use of edge count may lead to false ”low confidence” inferences, which our technique avoids.

In addition to edge score, we also use the 27 predefined edge types in the network to calculate conditional edge type transition probability. We use these edge types to calculate the overall distribution of the 27 edge types as the independent probability of encountering a given edge type on the initial edge traversal within a pathway. Furthermore, we also calculate the conditional probabilities of changing edge types along any given pathway. For example, given that we traverse an ”Is A” edge to a concept X, the probability of traversing another ”Is A” edge from concept X is .13, whereas the probability of changing to a ”Has A” relation is only .05, as shown in Figure 2. These transition probabilities are calculated for all sequential edge pairs within the entire graph, constructing the following Markov Model.

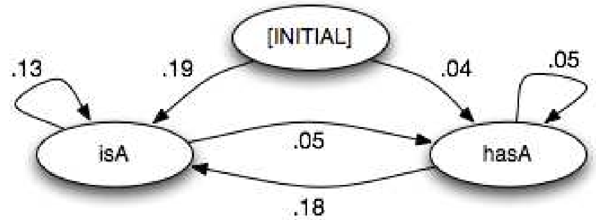

Figure 2: Learned edge transition probabilities (only 2 of 27 edge types shown).

We believe that the inclusion of these two additional variables helps identify edges that not only have a high consensus with respect to the individual triple, but also are contextually appropriate relative to the prior edge traversed. Hence, we are combining the specific attributes of edges with the characteristics of the larger, overall network, edge transition probabilities to assess pathways.

2 Pathway Scoring

We now have two properties for each edge on a given pathway of N nodes: score and transition probability. Using these, we can compose three vectors that can describe any given pathway: SCORES, TRANSITIONS, and EDGES (see equations 1, 2, and 3 respectively).

$$SCORES = [\sqrt[3]{EdgeScore_{1,2}}, \sqrt[3]{EdgeScore_{2,3}}, ..., \sqrt[3]{EdgeScore_{N-1,N}}] \quad (1)$$

SCORES is a vector of size N-1 with the cube root of each edge score in each cell. In initial testing, we found that using the cube root of user ranked scores was necessary to prevent high scoring edges from overshadowing lower scoring edges. We believe that this particular normalization technique is unique to the score distribution of ConceptNet 3, but that the general technique is expandable to other networks based on score distributions.

$$TRANSITIONS = [P(Edge_{1,2}), P(Edge_{2,3}|Edge_{1,2}), ..., P(Edge_{N,N-1}|Edge_{N-1,N-2})] \quad (2)$$

TRANSITIONS is another vector of size N-1 containing the conditional probabilities of observing each edge type within each cell. The first cell in the vector is just the independent probability of encountering the first edge, whereas every cell beyond that calculates the conditional probability of encountering the edge type based on the prior edge type.

$$EDGES = [N - 1] \tag{3}$$

EDGES is a vector of size 1 that simply contains N-1 as its value, or the number of edges in the pathway.

After associating these three vectors with all possible pathways from one concept to another, we now introduce a scoring function for each pathway. The function assigns each possible pathway a score, and the pathway with the highest score is used as the ”shortest” semantic pathway from one concept to another. Hence, if a given concept has several pathways to another given concept, we select the pathway with the maximum score and use this score as the similarity metric. We use the maximum score as a metric to follow similar work in the field rather than experimenting with average or minimum pathway strength. Note, that because ConceptNet is directed, the similarity metric between concepts corresponds to a specific direction.

$$PathScore = \frac{< SCORES, TRANSITIONS >}{||EDGES||^2} \tag{4}$$

The scoring function shown in equation 4 is the inner product of SCORES and TRANSITIONS divided by the EDGES, or number of edges, squared. This approach introduces the SCORES and TRANSITIONS values into the traditional inverse edge-count function. We use the inner product to obtain an interaction effect between scores and transitions, where each edge has its score multiplied by its transition probability before being summed. Hence, the inner product will reward edges with both high scores and high transition probabilities as the product grows higher. Our intuition is that this interaction effect will be useful for finding relevant pathways because it ensures that the edges traversed have a high social confidence score, and are ordered in an edge-type sequence that agrees with the overall structure of the network.

We define our function in terms of vectors so we can easily test the effect of ignoring certain features by substituting vectors with all 1s instead of the actual values. For example, to disable the use of the transition probabilities in scoring, we simply set all values in the TRANSITIONS vector to 1. This approach allows us to test all combinations of pathway features to determine which ones are useful, and whether or not an interaction effect exists between scores and transitions.

3 Experiments

We tested all possible combinations of enabling or disabling our three pathway feature vectors in computing pathway scores. To measure how well a scoring function performs, we followed Wubben's approach of comparing ConceptNet shortest path conceptual relatedness to the Finkelstein-353 psychological dataset [Finkelstein and Ruppin (2002)]. This dataset contains 353 word pairs, each with a similarity score from 0 to 10 based on psychological free word association between two words. Because ConceptNet does not contain nodes for all words used in the Finkelstein-353, we ignored word pairs that could not be found in ConceptNet. Wubben calculated the correlation between the word pair rankings in the Finkelstein-353 and ConceptNet similarity (using simple edge count as pathway cost) to obtain a Pearson's correlation of .47 [Wubben and A. (2009)]. Studies using Wikipedia are able to obtain even higher correlations than .47, however we find this work incomparable to our study due to the wider coverage and different structure of Wikipedia [Strube and Ponzetto (2006)].

Our experimental design differs from that of Wubben's on two fronts. Wubben used bidirectional pathways, whereas our work only focuses on directed pathways for each pair in the dataset, ranking the directed pathways from word A to word B in the Finkelstein-353. We chose this approach because psychological subjects were presented with word A first, followed by word B, in which we assume that subjects were more prone to make the directed connections. Furthermore, because we are not using a simple edge-counting algorithm, we were not able to efficiently examine all of ConceptNet due to computational limitations. Instead, we ran several cases in which we examine all pathways between concepts in the Finkelstein set within ConceptNet, subject to a maximum number of pathway nodes. For example, if we only allow a maximum of three pathway nodes, then we only analyze the set of found pathways containing up to three nodes. If a pathway is not found between two concepts within a three node pathway, but we know that the concept exists in ConceptNet, then we set the score equal to 0. We were able to reach a maximum of six pathway nodes before calculations became infeasible.

4 Results

Our experiments were designed to test the usefulness for pathway score and transition probabilities, as well as an interaction effect between the two features, in addition to inverse edge-count. Hence, we aim to prove that using all three vectors in Equation 4 will outperform the traditional approach of only using EDGES. We tested these cases, in addition to all other feature usage combinations, on allowed maximum pathway lengths up to six, hoping to see how allowed pathway length additionally affects our scoring function's usefulness.

Our results from Table 1 show a significant performance gain when using all three vectors in tandem, as opposed to just the number of edges. Furthermore, our results were able to outperform Wubben's results of .47. We also see that the performance gains over the naive edge-only approach increase with the maximum allowed number of nodes in the pathway.

Although we are able to show improvements over edge-only scoring, we replicate previous results that demonstrate that using the inverse of edge length is both necessary and sufficient to obtain acceptable results [Wubben and A. (2009)]. Results become unusable for rows (S), (T), and (ST) beyond a pathway length of 3 when pathway length (E) is not considered. Furthermore, edge only scoring (E) is sufficient by itself to obtain acceptable results, outperforming all other cases aside from (EST).

The most important finding is that if we only add scores or transition probabilities to edge count scoring in cases (ES) and (ET), then performance declines from using only edge counts (E). This suggests that edge scores and transition probabilities are noisy features when considered in isolation. It is only when they are combined, in conjunction with edge count (EST) that we see performance gains, confirming our proposed interaction effect.

5 Discussion

Our results suggest that the interaction effect between edge score and edge type transitioning is useful for improving conceptual similarity calculations in ConceptNet. Our scoring formula provides a strong reward for edges that contain a high user rated score and edge transition probability. We believe that these properties produce stronger results because isolated edge scores are contextualized based on their relative sequence of edge-types within a pathway. For example, a path traversal across an "is A" edge and then to a "has A" edge may not be meaningful if the network as a whole does not contextually support this edge type transition pattern.

For future studies, we would like to extend our experiments beyond pathways of length 6 and see if

Maximum Pathway Length	2	3	4	5	6
Edge (E)	.285	.391	.414	.416	416
Score (S)	.280	.347	.079	-.166	-.120
Transition (T)	.281	.357	.041	-.268	-.097
Edge & Score (ES)	.280	.383	.400	.400	.401
Edge & Transition (ET)	.281	.399	.387	.407	.414
Score & Transition (ST)	.275	.363	.070	-.223	-.093
Edge & Score & Transition (EST)	.275	.416	.473	.496	.507
(EST) Absolute Improvement over (E)	-.010	.025	.059	.080	.091
(EST) Relative Improvement over (E)	-3.51%	6.39%	14.25%	19.23%	21.88%

Table 1: Results: Pearson's correlation for varying formulas and improvement over the traditional edge-only approach. Row labels indicate which feature vectors were activated for the experiment.

performance will begin sinking or will flatten after a given number of allowed nodes. Furthermore, we believe that deeper analysis of ConceptNet is warranted. We noticed that there are many noisy relationships that are illogical at first glance. However, deeper analysis shows that these noisy assertions are often the result of word stemming (such as shortening ”building” to ”build”), and word sense disambiguation issues. In future work, we would like to revert to the pure lexical assertions within ConceptNet to remove the stemming from the words, preserving the full meaning of the intended relations. We believe that a lexical parser would be able to detect plurals and address stemming issues better than the approach used in ConceptNet. Furthermore, we believe that word sense ambiguity adds noise to the network, and that our results could improve if ambiguous nodes were split into their corresponding senses.

Despite the noise present in ConceptNet, we believe that our approach demonstrates a strong improvement over traditional semantic similarity computations. We hope that future work may resolve the noise and ambiguity issues present in ConceptNet, in which our methodology may provide even stronger results in accurately calculating semantic similarity.

References

Finkelstein, L., G. E. M. Y. R. E. S. Z. W. G. and E. Ruppin (2002). Placing search in context: The concept revisited. In *ACM Transactions on Information Sys- tems*.

Havasi, C. Speer, R. and J. Alonso (2007). Conceptnet 3: a flexi- ble, multilingual semantic network for common sense knowledge. In *Recent Advances in Natural Language Processing*.

Rada, R.; Mili, H. B. E. and M. Blettner (1989). Development and application of a metric to semantic nets. In *IEEE Transactions on Systems, Man and Cybernetics*, Volume 19 of *1*, pp. 17–30.

Strube, M. and S. Ponzetto (2006). Wikirelate! computing semantic relatedness using wikipedia. In *AAAI'06: proceedings of the 21st national conference on Artificial intelligence*, Boston, Massachusetts, pp. 1419–1424. AAAI Press.

Wubben, S. and B. A. (2009). A semantic relatedness metric based on free link structure. In *Proceedings of the 8th International Conference on Computational Semantics*, pp. 355–358. International Conference on Computational Semantics.