Article Title: Conceptual Semantic Evaluation Metric Using Taxonomy

Authors: İlknur Dönmez & Nur Bengisu Çam

Presenter: Nur Bengisu Çam, Scientific and Technological Research Council of Türkiye (TUBITAK BILGEM)

Presenter Email: nur.cam@tubitak.gov.tr

Nur Bengisu Çam

- Research & Development Engineer at TUBITAK (Scientific and Technological Research Council of Türkiye
- MSc at Bogazici University, Computer Engineering Department under the supervision of Arzucan Özgür.
- MSc Thesis Topic: Relation Extraction From Biomedical Text Using Transformer-based Models
- More than 3 years of software development experiment
- Currently working as a Deep Learning / Natural Language Processing research engineer.
  - LLM Finetuning & training from scratch
  - Machine Translation
  - Fraud Detection
  - Multi-modal system for emotion recognition from audio and text

WHAT IS THE PROBLEM?

- Conceptual method is one of the important technique to calculate semantic similarity.
- In earlier times, methods evolved from looking at sentences word by word as a distinguishing feature to using grammatical rules to represent sentences [21].
- WordNet has changed the whole literature by providing the ontology to compute semantic similarity between words [22].
- Semantic similarity can be used for many different purposes:
  - estimating similarity between documents [3],
  - ontology-based text clustering [4][5],
  - text summarization[6],
  - entity disambiguation [7],
  - developing recommender systems [8],
  - semantic annotation [9],
  - ontology merging [10],
  - ontology segment matching [11],
  - information retrieval [12],
  - personalized support [13]- [15],
  - and the graph editör similarity search problem [16], etc.

OUR PROPOSED METRIC – TAXONOMY TREE

- We propose a taxonomy based formula for calculating the conceptual similarity of sentences.
- We used WordNet taxonomy.
- By considering the distance of these two words from their common ancestor and the position of the common ancestor in the ontology tree.
- If the nodes are not equal, the distance is correlated with the height difference of the nodes to the common parent.
- If both children are closer to the common parent, it means that the concepts of the children's nodes are also closer and similar.
- When the distance to the common parent is small, the similarity is high.

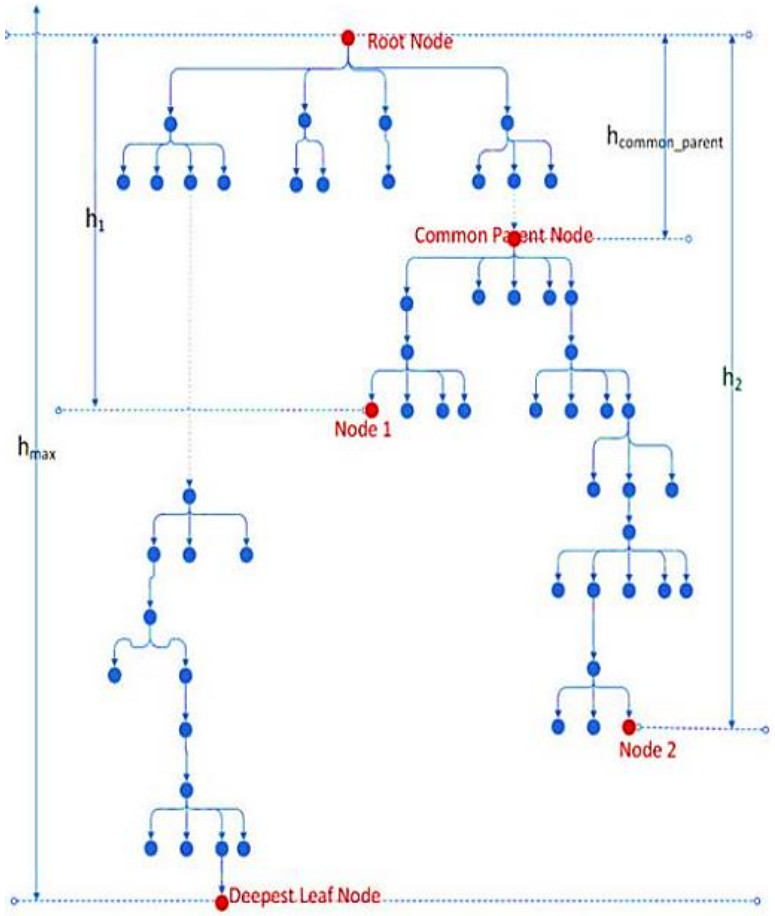

Figure 1. Nodes and their heights.

OUR PROPOSED METRIC - FORMULA

- Find the similarity value for each pair of words, we can average them to calculate the similarity of the sentences.
- We compare each word in one sentence with the words in the other sentence, and the most similar pairs of words are included in the calculation of the average.

$$sim_i = max_j \begin{cases} 1, & n_{1,i} = n_{2,j} \\ \dfrac{h_{cp,i,j}}{(h_{cp,i,j} - h_{1,i} + 1)(h_{cp,i,j} - h_{2,j} + 1)}, & n_{1,i} \neq n_{2,j} \end{cases} \quad (1)$$

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (2)$$

$$sim_{i\_s} = max_j \begin{cases} 1, & n_{1,i} = n_{2,j} \\ \dfrac{h_{cp,i,j}}{(h_{cp,i,j} - h_{1,i} + 1)(h_{cp,i,j} - h_{2,j} + 1)(h_{max})}, & n_{1,i} \neq n_{2,j} \end{cases} \quad (3)$$

$$sim = \frac{1}{n}\sum_{i=1}^{n}(sim_i) \quad (4)$$

OUR PROPOSED METRIC - ALGORITHM

- Step 1:

  For each pair of sentences in the dataset, remove the stop-words and the punctuations.

- Step 2:

  For each sentence in a pair, extract the part-of-speech (POS) tags of each word.

- Step 3:

  Create a combination of the words in the pair according to their POS tags, then calculate the similarity score of the word pairs, using (3).

- Step 4:

  From the previous step, we have many similarity score for a word. Accept the maximum similarity score.

DATASET

- Train split of the English STS benchmark dataset [37].
- It is a collection of data given in SemEval tasks between 2012 and 2017.
- 5749 sentence
- The given similarity scores range from 0 to 5
- These scores are annotated by human judges
- We normalized the similarity scores using the min-max normalization function of scikit-learn [38].

EVALUATION

TABLE II. NORMALIZED PAIR SIMILLARITY SCORES EXAMPLES

Pairs	Sentences	STS Similarity Score	Noun Only Similarity Score	Noun + Verb Similarity Score	Noun + Verb + Adjective Similarity Score
1	A woman is dancing and singing with other women. A girl is dancing and singing in the rain.	0.60	0.73	0.86	0.56
2	Two men are packing suitcases into the trunk of a car. The men are putting suitcases into the car's trunk.	0.88	1.00	0.75	0.50
3	The woman picked up the kangaroo. A woman picks up a baby kangaroo.	0.75	1.00	0.50	0.33
4	Two foxes are eating from a plate on a brick patio. Foxes are eating from a plate.	0.56	0.51	0.75	0.50
5	Two zebras are playing. Zebras are socializing.	0.85	1.00	0.50	0.34
6	A group of people dance on a hill. A group of people are dancing.	0.64	0.67	0.33	0.22
7	A car is moving through a road. A car is driving down the road.	0.80	1.00	0.50	0.33
8	The man is shooting an automatic rifle. A man is shooting a gun.	0.76	0.58	0.79	0.52
9	A woman is cutting up a chicken. A woman is slicing meat.	0.55	0.54	0.27	0.18
10	Butter is being put into a bowl. A man cutting butter into a mixing bowl.	0.85	1.00	0.50	0.33

- If any of the sentences of the pairs do not contain adjectives, then the similarity between the adjectives is zero. Therefore, the similarity score in such a case is drastically lower.

EVALUATION - CONT'D

TABLE IV. EVALUATION

	Results of the Proposed Method	
Noun Only	Percentage of pairs in the range of 1.00– 0.50	77.09%
	Percentage of pairs in the range of 1.00 – 0.75	72.42%
	Percentage of pairs in the range of 0.75 – 0.50	82.05%
	Percentage of pairs in the range of 0.50 – 0.00	56.85%
	Percentage of pairs in the range of 0.50 – 0.25	42.38%
	Percentage of pairs in the range of 0.25 – 0.00	69.45%
Noun + Verb	Percentage of pairs in the range of 1.00 – 0.50	56.58%
	Percentage of pairs in the range of 1.00 – 0.75	50.74%
	Percentage of Pairs in the range of 0.75 – 0.50	62.77%
	Percentage of pairs in the range of 0.50 – 0.00	71.94%
	Percentage of pairs in the range of 0.50 – 0.25	59.37%
	Percentage of pairs in the range of 0.25 – 0.00	82.88%
Noun + Verb + Adjective	Percentage of pairs in the range of 1.00 – 0.50	46.35%
	Percentage of pairs in the range of 1.00 – 0.75	19.47%
	Percentage of pairs in the range of 0.75 – 0.50	74.88%
	Percentage of pairs in the range of 0.50 – 0.00	81.99%
	Percentage of pairs in the range of 0.50 – 0.25	54.76%
	Percentage of pairs in the range of 0.25 – 0.00	95.71%

- For different *thresholds*, to see if our proposed method also labels these pairs similarly we checked *what percentage* of pairs the proposed method finds in this range.

TABLE III. CORRELATION RESULTS

Task Name	Pearson Correlation Score
Noun Only	0.51
Noun + Verb	0.47
Noun + Verb + Adjective	0.48

CONCLUSION

- We have proposed a formula for calculating the conceptual similarity of sentences.
- By looking at the distance of these two words to their common ancestor and the location of the common ancestor in the ontology tree, we calculated the sentence similarity.
- If the compared words are close to their common ancestor, they are more likely to be similar.
- If the common ancestor is far from the root, the similarity of the compared words increases according to its position closer to the root node.
- Even if WordNet also has a taxonomical structure of adjectives, they name an attribute of a noun (a concept). Thus, they are not the actual concepts [39]. The inclusion of the similarity contribution between verbs or adjectives in our study negatively affected the results.
- In the future, we would like to use a 5-level human-annotated similarity dataset, where the judges will be asked to classify the pairs into high similarity, low similarity, different, completely different, and no idea classes.
- We expect our method to give better results on the human tagged datasets, since our proposed method simulates the human mind to find the conceptual relationship.

References

1. L. W. Barsalou, “The human conceptual system,” The Cambridge handbook of psycholinguistics, pp. 239-258, 2012.
2. G. Pirró, and D. Talia, “UFOme: An ontology mapping system with strategy prediction capabilities,” Data & Knowledge Engineering, 69(5), pp. 444-471, 2010.
3. F. Benedetti, D. Beneventano, S. Bergamaschi, and G. Simonini, “Computing inter-document similarity with context semantic analysis,” Information Systems, vol 80, pp. 136-147, 2019.
4. J. Nasir, I. Varlamis, A. Karim, and G. Tsatsaronis, “Semantic smoothing for text clustering,” Knowledge-Based Systems, vol 54, pp. 216-229, 2013.
5. W. Song, J. Liang, and S. Park, “Fuzzy control GA with a novel hybrid semantic similarity strategy for text clustering,” Information Sciences, vol. 273, pp. 156-170, 2014.
6. S. Kumar, and K. Bhatia, “Semantic similarity and text summarization based novelty detection,” SN Applied Sciences, vol 2, pp. 1-15, 2020.
7. A. Vretinaris, C. Lei, V. Efthymiou, X. Qin, and F. Özcan, “Medical entity disambiguation using graph neural networks,” In Proceedings of the 2021 International Conference on Management of Data, pp. 2310-2318, 2021.
8. V. Demertzi, and K. Demertzis, “A hybrid adaptive educational eLearning project based on ontologies matching and recommendation system,” arXiv preprint, arXiv:2007.14771, 2020.
9. A. Chikkamannur, “Semantic Annotation of IoT Resource with ontology orchestration,” In 2020 Third International Conference on Advances in Electronics, Computers and Communications (ICAECC- IEEE) pp. 1-7, 2020.
10. S. Mhammedi, H. El Massari, and N. Gherabi, “Composition of large modular ontologies based on structure,” In Advances in Information, Communication and Cybersecurity: Proceedings of ICI2C’21, pp. 144-154, 2022.
11. A. Belhadi, Y. Djenouri, G. Srivastava, and J. Lin, “Fast and Accurate Framework for Ontology Matching in Web of Things,” ACM Transactions on Asian and Low-Resource Language Information Processing, 2023.
12. G. Yang, et al., “CCGIR: Information retrieval-based code comment generation method for smart contracts,” Knowledge-Based Systems, 237, 107858, 2022.
13. M. Sreenivasan, S. Dhar, and A. Chacko, “PCPS: Personalized Care through Patient Similarity,” In 2022 IEEE Region 10 Symposium (TENSYMP-IEEE), pp. 1-6, 2022.
14. Y. Liu, and M. Ijaz, “Personalized auxiliary information presentation system for mobile network based on multimodal information,” Mobile Networks and Applications, pp. 1-11 2022.
15. **F. Liu and S. Li, “Research on personalized user-centered product improvement based on sentiment mining of online reviews and competitor analysis,” Home, http://www.researchsquare.com/article/rs-1829215/v1 (Accessed Aug. 19, 2023).**
16. S. Babalou, A. Algergawy, and B. KönigRies, “SimBio: Adopting Particle Swarm Optimization for ontology-based biomedical term similarity assessment,” Data & Knowledge Engineering, 102137, 2023.
17. M. Landolsi, L. Hlaoua, and L. Ben Romdhane, “Information extraction from electronic medical documents: state of the art and future research directions,” Knowledge and Information Systems, vol. 65(2), pp 463-516, 2023.
18. B. Yang, et al., “Classification of Medical Image Notes for Image Labeling by Using MinBERT,” Tsinghua Science and Technology, vol. 28(4), pp. 613-627, 2023.
19. D. Tian, M. Li, Y. Shen, and S. Han, “Intelligent mining of safety hazard information from construction documents using semantic similarity and information entropy,” Engineering Applications of Artificial Intelligence, 119, 105742, 2023.

References

20. N. B. Cam, sementicSimilarity. Github. https://github.com/bengisucam/semanticSimilarity . (accessed Aug. 19, 2023).
21. R. P. Honeck, “Semantic similarity between sentences,” *Journal of Psycholinguistic Research*, vol. *2*, pp. 137-151, 1973.
22. **G. A. Miller, “WordNet: A Lexical Database for English.,” In Human Language Technology: Proceedings of a Workshop held at Plainsboro, New Jersey, March 8-11, 1994.**
23. M. Sussna, “Word sense disambiguation for free-text indexing using a massive semantic network,” In Proceedings of the second international conference on Information and knowledge management, pp. 67-74, 1993.
24. J. J. Jiang and D. W. Conrath “Semantic similarity based on corpus statistics and lexical taxonomy,” arXiv preprint cmp-lg/9709008, 1997.
25. N. Seco, T. Veale, and J. Hayes, “An intrinsic information content metric for semantic similarity in WordNet,” In *Ecai* vol. 16, p. 1089, 2004.
26. D. Yang and D. M. Powers, “Measuring semantic similarity in the taxonomy of WordNet,” Australian Computer Society, 2005.
27. X. Y. Liu, Y. M. Zhou and R. S. Zheng, “Measuring semantic similarity in WordNet,” In *2007 international conference on machine learning and cybernetics*, vol. 6, pp. 3431-3435, 2007.
28. Z. Zhou, Y. Wnag, and J. Gu, “New model of semantic similarity measuring in wordnet,” In *2008 3rd International Conference on Intelligent System and Knowledge Engineering*, vol. 1, pp. 256-261, 2008.
29. P. Sravanthi and B. Srinivasu, “Semantic similarity between sentences,” International Research Journal of Engineering and Technology (IRJET), vol. 4, no. 1, pp. 156-161, 2017.
30. A. Selvarasa, N. Thirunavukkarasu, N. Rajendran, C. Yogalingam, S. Ranathunga, and G. Dias, “Short Tamil sentence similarity calculation using knowledge-based and corpus-based similarity measures,” In 2017 Moratuwa Engineering Research Conference, pp. 443-448, 2017.
31. SemEval 2016 Dataset: https://altqcri/semeval2016/task2. (Accessed Aug. 19, 2023)
32. M. N. Jeyaraj and D. Kasthurirathna, “Mnet-SIM: A multi-layered semantic similarity network to evaluate sentence similarity,” International Journal of Engineering Trends and Technology, vol. 69, no. 7, pp. 181–189, 2021. doi:10.14445/22315381/ijett-v69i7p225
33. M. C. Lee. “A novel sentence similarity measure for semantic-based expert systems,” Expert Systems with Applications, vol. 38, no. 5, pp. 6392-6399, 2011.
34. N. Love, “Translational semantics: A discussion of the second edition of Geoffrey Leech's Semantics: the Study of Meaning,” Stellenbosch Papers in Linguistics, vol. 11, pp. 115-136, 1983.
35. D. Geeraerts, “Theories of lexical semantics,” OUP Oxford, 2009.
36. T. A. Van Dijk, “Society and discourse: How social contexts influence text and talk,” Cambridge University Press, 2009.
37. Cer, Daniel, et al., "Semeval-2017 task 1: Semantic textual similarity-multilingual and cross-lingual focused evaluation," arXiv preprint arXiv:1708.00055, 2017.
38. Pedregosa, Fabian, et al., "Scikit-learn: Machine learning in Python," The Journal of machine Learning research 12, pp. 2825-2830, 2011.
39. D. Gross and K. J. Miller, “Adjectives in wordnet,” International Journal of lexicography, vol. 3, no. 4, pp. 265-277, 1990.