Cutting-Edge Neurofuzzy Approaches for Semantic Textual Similarity

Jorge Martinez-Gil

Abstract

This study presents our research focused on developing innovative models designed to effectively and efficiently process textual information. We demonstrate that neural networks and fuzzy logic possess distinct characteristics that render them suitable for specific problem domains while presenting limitations in others. Neural networks excel in pattern recognition but lack inherent interpretability for decision-making. Conversely, fuzzy logic systems offer interpretability but cannot derive decision-making rules automatically. These limitations have motivated the creation of an intelligent hybrid system, merging both techniques to address the drawbacks above at the individual level.

Keywords: Text and Unstructured Data, Neurofuzzy systems, Deep Learning Applications

1. Introduction

Assessing the level of semantic similarity that exists between different pieces of text is a difficult task. We are using here Deep Learning models because, in most areas where artificial intelligence may assist in the automation of processes, these models have shown to generate the most significant outcomes. However, these models also have some unique flaws, such as an inability to make it easier for human operators to comprehend the insights they provide and an excessive requirement for a considerable quantity of data in order for them to be trained. Because of these two factors, their application is not nearly as ubiquitous as its potential would appear to promise it would be. On the other hand, our working hypothesis is that if these methods are appropriately combined with other methods that are better equipped to facilitate explainability, such as fuzzy logic, it is possible to arrive at an acceptable conclusion.

The requirement for data information frequently needs the interpretability of models to comprehend the derived knowledge. Most of the solutions that implement some eXplainable Artificial Intelligence (XAI) are focused on technical solutions-oriented for operators with profound mathematical knowledge [19, 21]. However, another family of alternative methods is based on symbolic AI. Therefore, XAI is the research field that tries to make systems more transparent and understandable. Following this approach, we look at neurofuzzy models [27]. Therefore, the contribution of this work is that we propose a neurofuzzy model result of a fuzzy system that operates concurrently with a neural network intending to soften the deficiency of each of these systems. In this way, we aim to achieve a model being more efficient, robust, and easy to understand how pieces of text are being compared.

The rest of this work is structured in the following way: Section 2 presents the related works on using neurofuzzy systems for textual information processing. Section 3 presents our proposal's technical details for using neurofuzzy systems. Finally, we draw the essential conclusions of this work and outline the potential lines for future work.

2. State-of-the-art

Fuzzy systems are usually characterized by the structure and the number of fuzzy rules they implement [20]. This starkly contrasts with artificial neural networks, often known as ANNs, designed to address pattern recognition, regression, or computer vision problems. ANNs cannot learn anything outside the training data input into the model. Incorporating fuzzy logic into neural networks is one approach that might be taken to meet the challenge of training a machine to make decisions in a way that is analogous to that of a person [16].

Nowadays, fuzzy systems and neural networks are widely used for universal approximation [24]. Fuzzy systems are interpretable since their behavior can be explained using rules close to natural language [7]. Thus their performance can be modified by changing those rules. However, knowledge acquisition is complicated, and each input variable's universe of discourse must be divided into several intervals. Furthermore, the number of input variables is also limited. Therefore, fuzzy system applications are often limited to areas where expert or domain knowledge is available. In order to overcome this problem, several methods for extracting fuzzy rules from numerical data have been developed, among which the neurofuzzy model stands out [29].

Lastly, concerning semantic similarity, there is a large body of literature on methods and tools to address the problem [3, 25]. From the classical Latent Semantic Analysis techniques or equivalent distributional approaches, which prevailed for decades [9, 23], to the most innovative techniques based on word embeddings [12], through the dictionary-based techniques from the mid-2000s [5, 11, 13, 15] or semantic similarity controllers [22] and other kinds of aggregators [17, 18].

Today, Deep Learning is predominant as it can obtain the best results in many use cases [4]. However, its effectiveness has not yet been demonstrated in more specific domains with unique characteristics. It is widely assumed that models based on whole-sentence understanding, rather than traditional word-by-word approaches, are expected to be better. In order to fill the mentioned gap and test our hypothesis, we present a novel neurofuzzy computational model to address the problem of semantic similarity in texts with the double goal of being accurate and interpretable simultaneously.

3. How does a neurofuzzy system work?

In this project, we design a concurrent neurofuzzy system that automatically considers some peculiarities to compare text. Our system is composed of a neural and fuzzy part designed independently but must be coupled to work together. In the following, we describe each of these components.

3.1. Neural component

The neural component leverages transformers that are appropriate models for transformations between abstract representations [28]. The foundation of the transformer models is an encoder-decoder architecture, which allows the encoder to learn how to represent the input data and then transfer this representation onto the decoder. The decoder's responsibility is to obtain the representation and then provide the output data for the user.

3.2. Fuzzy component

The development of inference algorithms can benefit from the computational tools that fuzzy logic can give. We are looking at Mamdani fuzzy inference because it makes it easier to build systems controlled by a set of rules similar to natural language. Any rule implemented in Mamdani fuzzy inference systems will always provide a fuzzy set as its output. Because of its rule base, which is relatively easy to understand and is very similar to natural language [6], these systems are widely implemented in expert systems to mimic the knowledge of human experts [1].

We will work with the IEC 61131-7 standard since it inspires a specific-purpose programming language and does not have primitives unrelated to fuzzy logic.

3.3. Coupling and mode of operation

We work here with a concurrent neurofuzzy model, where the input data is fed to a neural network. This way of working helps determine the system's best membership functions to be processed later. Combining neural networks with fuzzy logic does not optimize the fuzzy system but improves the system's overall performance.

The advantage of analytically derived fuzzy systems over other kinds of solutions is clear: the operator no longer needs to specify the structure of the model before operating. However, this advantage comes with its drawbacks, particularly computation. The search space of all possible configurations cannot feasibly be searched in its entirety. Thus, clever algorithms must be employed to search this space intelligently to discover the correct model to fit the data.

A neurofuzzy model is created by combining the two methods correctly. We want to get satisfactory results when calculating the semantic similarity between two terms. Furthermore, these systems benefit from being able to be trained alone or collectively. If it is necessary to model a trade-off between model accuracy and interpretability, it is possible to resort to multi-objective evolutionary techniques [8]. As a result, they have a lot of flexibility and variety.

4. Discussion

When considered independently, both neural networks and fuzzy systems offer considerable benefits and drawbacks. Knowledge may be typically acquired through the use of neural networks. The learning process, however, is quite long since it needs to collect vast amounts of data; it is also typically tough to comprehend the completed model. In addition to this, extracting structured knowledge from the model that is produced is challenging, and it is also hard to insert domain knowledge to ease the learning process.

The presence of better features facilitates the interpretability of fuzzy systems. In addition, these systems' performance may be modified using IF-THEN rules. However, learning this information may be highly challenging; numerous factors must be adjusted, so it is frequently necessary to seek the assistance of a domain expert. In order to solve all of these issues, neurofuzzy systems, which integrate the advantages of both classical and modern approaches, have been offered as a solution.

Our approach is pioneering in bringing this computational model here because neurofuzzy systems have been traditionally used in control applications or industrial environments. We think this approach opens a field of possibilities to investigate the application of these hybrid systems in a wide range of language-related applications.

5. Conclusions and Future Work

Neurofuzzy systems have been relatively well studied in engineering and industrial applications. However, their application in text processing has been very little explored. New techniques based on neural-based solutions allow text to be transformed into numerical vectors in a sophisticated way that can even preserve positional information about words. This makes them a clear candidate for processing complicated sentences and paragraphs. Thus, combining neural-based and fuzzy systems with raising neurofuzzy solutions is now possible. In addition, its performance makes it a good choice for developing even more sophisticated text models. So, in principle, it seems clear that neurofuzzy models combine the best features of neural networks and fuzzy logic to produce systems that overcome the shortcomings of both methodologies.

In future work, we intend to explore further the possibilities of a neurofuzzy model for working with domain-specific texts. In this project, we focus exclusively on combining popular models such as BERT[10], ELMo [26] or USE [2] using Mamdani inference [14]. However, for example, one could also study the degree of success that could be achieved with Takagi Sugeno-type models [30]. All this is to open up a new field of research and forget the old idea that neurofuzzy systems belong exclusively to industrial application fields.

6. Acknowledgments

The NEFUSI project is funded in the project NGI Zero Discovery by the NLnet Foundation and the European Commission. Project number: 2021-04-069.

References

[1] Angelov, P. P., & Buswell, R. A. (2003). Automatic generation of fuzzy rule-based models from data by genetic algorithms. *Inf. Sci.*, *150*, 17–31. doi:10.1016/S0020-0255(02)00367-5.

[2] Cer, D., Yang, Y., Kong, S., Hua, N., Limtiaco, N., John, R. S., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., Strope, B., & Kurzweil, R. (2018). Universal sentence encoder for english. In E. Blanco, & W. Lu (Eds.), *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018: System Demonstrations, Brussels, Belgium, October 31 - November 4, 2018* (pp. 169–174). Association for Computational Linguistics. URL: https://doi.org/10.18653/v1/d18-2029. doi:10.18653/v1/d18-2029.

[3] Chaves-Gonzalez, J. M., & Martinez-Gil, J. (2013). Evolutionary algorithm based on different semantic similarity functions for synonym recognition in the biomedical domain. *Knowl. Based Syst.*, *37*, 62–69. doi:10.1016/j.knosys.2012.07.005.

[4] Chen, C., Huang, X., Ji, F., Wang, C., Qiu, M., Huang, J., & Zhang, Y. (2021). CAT-BERT: A context-aware transferable BERT model for multi-turn machine reading comprehension. In C. S. Jensen, E. Lim, D. Yang, W. Lee, V. S. Tseng, V. Kalogeraki, J. Huang, & C. Shen (Eds.), *Database Systems for Advanced Applications - 26th International Conference, DASFAA 2021, Taipei, Taiwan, April 11-14, 2021, Proceedings, Part II* (pp. 152–167). Springer volume 12682 of *Lecture Notes in Computer Science*. URL: https://doi.org/10.1007/978-3-030-73197-7_10. doi:10.1007/978-3-030-73197-7_10.

[5] Cilibrasi, R., & Vitányi, P. M. B. (2007). The google similarity distance. *IEEE Trans. Knowl. Data Eng.*, *19*, 370–383. URL: https://doi.org/10.1109/TKDE.2007.48. doi:10.1109/TKDE.2007.48.

[6] Cingolani, P., & Alcalá-Fdez, J. (2013). jfuzzylogic: a java library to design fuzzy logic controllers according to the standard for fuzzy control programming. *Int. J. Comput. Intell. Syst.*, *6*, 61–75. doi:10.1080/18756891.2013.818190.

[7] Cordón, O. (2011). A historical review of evolutionary learning methods for mamdani-type fuzzy rule-based systems: Designing interpretable genetic fuzzy systems. *Int. J. Approx. Reason.*, *52*, 894–913. URL: https://doi.org/10.1016/j.ijar.2011.03.004. doi:10.1016/j.ijar.2011.03.004.

[8] Deb, K., Agrawal, S., Pratap, A., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.*, *6*, 182–197. URL: https://doi.org/10.1109/4235.996017. doi:10.1109/4235.996017.

[9] Deerwester, S. C., Dumais, S. T., Landauer, T. K., Furnas, G. W., & Harshman, R. A. (1990). Indexing by latent semantic analysis. *J. Am. Soc. Inf. Sci.*, *41*, 391–407.

[10] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, .

[11] Han, L., Kashyap, A. L., Finin, T., Mayfield, J., & Weese, J. (2013). Umbc_ebiquity-core: Semantic textual similarity systems. In M. T. Diab, T. Baldwin, & M. Baroni (Eds.), *Proceedings of the Second Joint Conference on Lexical and Computational Semantics, *SEM 2013, June 13-14, 2013, Atlanta, Georgia, USA* (pp. 44–52). Association for Computational Linguistics. URL: `https://www.aclweb.org/anthology/S13-1005/`.

[12] Lastra-Díaz, J. J., Goikoetxea, J., Taieb, M. A. H., García-Serrano, A., Aouicha, M. B., & Agirre, E. (2019). A reproducible survey on word embeddings and ontology-based methods for word similarity: Linear combinations outperform the state of the art. *Eng. Appl. Artif. Intell.*, *85*, 645–665. URL: `https://doi.org/10.1016/j.engappai.2019.07.010`. doi:`10.1016/j.engappai.2019.07.010`.

[13] Li, Y., Bandar, Z., & McLean, D. (2003). An approach for measuring semantic similarity between words using multiple information sources. *IEEE Trans. Knowl. Data Eng.*, *15*, 871–882. URL: `https://doi.org/10.1109/TKDE.2003.1209005`. doi:`10.1109/TKDE.2003.1209005`.

[14] Mamdani, E. H., & Assilian, S. (1999). An experiment in linguistic synthesis with a fuzzy logic controller. *Int. J. Hum.-Comput. Stud.*, *51*, 135–147. doi:`10.1006/ijhc.1973.0303`.

[15] Martinez-Gil, J. (2014). An overview of textual semantic similarity measures based on web intelligence. *Artif. Intell. Rev.*, *42*, 935–943. doi:`10.1007/s10462-012-9349-8`.

[16] Martinez-Gil, J. (2015). Automated knowledge base management: A survey. *Comput. Sci. Rev.*, *18*, 1–9. URL: `https://doi.org/10.1016/j.cosrev.2015.09.001`. doi:`10.1016/j.cosrev.2015.09.001`.

[17] Martinez-Gil, J. (2016). CoTO: A novel approach for fuzzy aggregation of semantic similarity measures. *Cogn. Syst. Res.*, *40*, 8–17. URL: `https://doi.org/10.1016/j.cogsys.2016.01.001`. doi:`10.1016/j.cogsys.2016.01.001`.

[18] Martinez-Gil, J. (2019). Semantic similarity aggregators for very short textual expressions: a case study on landmarks and points of interest. *J. Intell. Inf. Syst.*, *53*, 361–380. URL: `https://doi.org/10.1007/s10844-019-00561-0`. doi:`10.1007/s10844-019-00561-0`.

[19] Martinez-Gil, J., & Aldana-Montes, J. F. (2013). Semantic similarity measurement using historical google search patterns. *Inf. Syst. Frontiers*, *15*, 399–410. URL: `https://doi.org/10.1007/s10796-012-9404-7`. doi:`10.1007/s10796-012-9404-7`.

[20] Martinez-Gil, J., & Chaves-Gonzalez, J. M. (2019). Automatic design of semantic similarity controllers based on fuzzy logics. *Expert Syst. Appl.*, *131*, 45–59. URL: `https://doi.org/10.1016/j.eswa.2019.04.046`. doi:`10.1016/j.eswa.2019.04.046`.

[21] Martinez-Gil, J., & Chaves-Gonzalez, J. M. (2020). A novel method based on symbolic regression for interpretable semantic similarity measurement. *Expert Syst. Appl.*, *160*, 113663. doi:10.1016/j.eswa.2020.113663.

[22] Martinez-Gil, J., & Chaves-Gonzalez, J. M. (2021). Semantic similarity controllers: On the trade-off between accuracy and interpretability. *Knowl. Based Syst.*, *234*, 107609. URL: https://doi.org/10.1016/j.knosys.2021.107609. doi:10.1016/j.knosys.2021.107609.

[23] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States.* (pp. 3111–3119).

[24] Nauck, D. D., & Kruse, R. (1999). Neuro-fuzzy systems for function approximation. *Fuzzy Sets Syst.*, *101*, 261–271. URL: https://doi.org/10.1016/S0165-0114(98)00169-9. doi:10.1016/S0165-0114(98)00169-9.

[25] Navigli, R., & Martelli, F. (2019). An overview of word and sense similarity. *Nat. Lang. Eng.*, *25*, 693–714. URL: https://doi.org/10.1017/S1351324919000305. doi:10.1017/S1351324919000305.

[26] Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. In *Proceedings of NAACL-HLT* (pp. 2227–2237).

[27] Rutkowski, L., & Cpalka, K. (2003). Flexible neuro-fuzzy systems. *IEEE Trans. Neural Networks*, *14*, 554–574. URL: https://doi.org/10.1109/TNN.2003.811698. doi:10.1109/TNN.2003.811698.

[28] Rychalska, B., Pakulska, K., Chodorowska, K., Walczak, W., & Andruszkiewicz, P. (2016). Samsung poland NLP team at semeval-2016 task 1: Necessity for diversity; combining recursive autoencoders, wordnet and ensemble methods to measure semantic similarity. In *Proceedings of the 10th International Workshop on Semantic Evaluation, SemEval-NAACL-HLT 2016, San Diego, CA, USA, June 16-17, 2016* (pp. 602–608). URL: https://www.aclweb.org/anthology/S16-1091/.

[29] Singh, H., & Lone, Y. A. (2020). *Deep Neuro-Fuzzy Systems with Python*. Springer.

[30] Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans. Syst. Man Cybern.*, *15*, 116–132. URL: https://doi.org/10.1109/TSMC.1985.6313399. doi:10.1109/TSMC.1985.6313399.