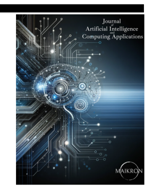
MAIKRON

Journal
of
Artificial Intelligence
and
Computing Applications



Math for AI Capsules

An element-wise contribution-based vector similarity measure for artificial intelligence applications: a brief exploration

Mauricio G. Orozco-del-Castillo[1]

[1]Tecnológico Nacional de México / IT de Mérida, Mérida, Yucatán, México

ABSTRACT

In Artificial Intelligence (AI), the ability to accurately assess the similarity between data points is fundamental to a myriad of applications, from recommendation systems and semantic analysis to case-based reasoning. Traditional similarity measures, however, often fall short in capturing the relationships inherent in complex data, particularly when the relevance of individual features varies or opposes. This article briefly introduces a novel element-wise contribution-based vector similarity measure that dynamically weighs the importance and directional relevance of features, offering a more refined approach to similarity assessment in AI. By normalizing vectors within specific ranges and employing a modulating vector to adjust feature contributions, our measure facilitates a more contextually aware and adaptable comparison process. The proposed measure is poised to have wide-ranging implications across diverse AI domains, suggesting its potential to enrich personalized, intelligent systems. This work contributes in similarity measurement, proposing a pathway for future research into AI methodologies that necessitate a personalised interpretation of data.

Keywords: similarity measure, element-wise analysis, feature importance modulation

1. Introduction

In the rapidly evolving field of Artificial Intelligence (AI), accurately quantifying the similarity between data points is foundational to the success of numerous applications, ranging from machine learning models, case-based reasoning and data clustering to information retrieval and pattern recognition [1]. These measures serve as critical components in algorithms that categorize, group, or differentiate data based on underlying patterns and relationships [2]. As AI systems handle increasingly complex datasets across diverse domains, the demand for more contextually aware similarity measures becomes fundamental [3]. This demand highlights the importance of developing similarity metrics that not only capture the essence of data relationships but also adapt to the unique requirements and dimensions of varied AI tasks.

Despite the widespread use of traditional similarity measures in AI, such as the Euclidean distance and cosine similarity, these methods often fall short when confronted with the intricate structure of high-dimensional data or when specific dimensions carry disparate significance [4, 5]. This limitation is particularly pronounced in domains where the relevance of features varies widely or explanation of such features is paramount, necessitating a more flexible approach to similarity assessment [6]. The inherent rigidity of conventional measures, which treat all dimensions uniformly, overlooks the underlying

E-mail address: mauricio.orozco@itmerida.edu.mx
https://doi.org/10.5281/zenodo.10846197

relationships and importance differentials among features, potentially leading to suboptimal performance in tasks such as clustering, classification, and recommendation systems [7, 8].

In response to these challenges, we propose a novel similarity measure that integrates element-wise contribution and directional relevance, leveraging a modulating vector to dynamically weigh the importance and oppositional alignment of features of two given vectors. This approach not only addresses the dimensional disparity but also introduces an understanding of 'opposite' directionality in vector similarity, thereby enriching the measure's applicability across a spectrum of AI tasks [4, 6].

2. Concept overview

The cornerstone of our proposed similarity measure is the interpretation of vector components within the context of their application domain. Specifically, the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$, representing entities to be compared, are normalized to lie within the range of $[-0.5, 0.5]$. This normalization reflects the premise that elements within these vectors span opposite ends of a spectrum, allowing for a more balanced and interpretable comparison. Central to our formulation is the vector $\mathbf{v}$, with elements constrained within $[-1, 1]$, which serves as a modulator for the contribution of each corresponding feature in $\mathbf{v}_1$ and $\mathbf{v}_2$ towards the overall similarity score. This arrangement ensures that each feature's influence is not only accounted for in terms of magnitude but also in terms of its directional relevance, with negative values indicating an opposing contribution to the similarity measure.

The similarity measure is articulated through a formula that combines the normalized vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ with the modulating influence of $\mathbf{v}$. The formula employs an element-wise operation modulated by $\mathbf{v}$, where each element's contribution is scaled according to its value in $\mathbf{v}$, signifying the feature's relative importance or oppositional alignment. This is achieved by adjusting the elements of $\mathbf{v}_2$ in accordance with the sign and magnitude of the corresponding elements in $\mathbf{v}$, thereby reflecting their directional contribution to similarity. The measure then calculates the absolute difference between the adjusted $\mathbf{v}_2$ and $\mathbf{v}_1$, which is subsequently weighted by the normalized version of $\mathbf{v}$, ensuring that each feature's significance is proportionally represented. The culmination of this process is a norm calculation, aggregating the weighted differences into a single scalar value that encapsulates the multidimensional similarity between the vectors, accounting for both magnitude and directional alignment of the constituent features. The formal expression of this measure is given by:

$$S = \left\| (1 - |\mathbf{v}_1 - (\mathbf{v}_2 \odot \operatorname{sign}^*(\mathbf{v}))|) \odot \frac{\mathbf{v}}{\|\mathbf{v}\|} \right\|, \qquad (1)$$

where $\odot$ denotes element-wise multiplication, and $\operatorname{sign}^*(\mathbf{v})$ is an element-wise operation that maps positive elements and zero in $\mathbf{v}$ to 1 and negative elements to -1. This formula encapsulates the core principles of our similarity measure, taking into account the normalized differences between $\mathbf{v}_1$ and $\mathbf{v}_2$, modulated by the importance and directional relevance assigned by $\mathbf{v}$. The measure yields a value between 0 and 1, corresponding to low and high similarity, respectively.

3. Application in AI

The versatility of our similarity measure could be evidenced by its broad applicability across various AI domains, notably in areas where in-depth analysis and interpretation of data are paramount. In personalized recommendation systems, this measure could discern subtle user preferences from activity logs, significantly enhancing user experience by weighting features based on their relevance and directionality [9]. Our approach enables the identification of items that more closely align with a user's implicit preferences, even when not overtly expressed.

In semantic text analysis, our measure could facilitate the evaluation of thematic closeness among documents, where the normalization of $\mathbf{v}_1$ and $\mathbf{v}_2$ accommodates a comparison sensitive to the semantic scale of features. The modulation by $\mathbf{v}$ ensures that each term's specific importance is accurately reflected, aiding in the clustering or classification of texts based on underlying topics or sentiments [10].

Moreover, case-based reasoning (CBR) systems stand to benefit from our similarity measure. CBR hinges on the retrieval of relevant cases for problem-solving based on past experiences. By finely tuning the contribution of each feature in the comparison process, our measure offers a more adaptable and context-sensitive means to match new cases with the most pertinent historical instances, thereby optimizing the problem-solving process in CBR applications [11, 12].

4. Conclusions

In this work, we have introduced a novel similarity measure that leverages the nuanced contributions of individual vector elements $\mathbf{v}_1$ and $\mathbf{v}_2$, modulated by a third vector $\mathbf{v}$, to assess the similarity between two entities. This measure stands out by its ability to account for both the magnitude and directional relevance of features within vectors $\mathbf{v}_1$ and $\mathbf{v}_2$, thus offering a more sophisticated and contextually aware approach to similarity assessment. The normalization of vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ within a defined range, alongside the modulation provided by $\mathbf{v}$, ensures that the similarity measure is adaptable to various domains and applications within AI, including but not limited to recommendation systems, semantic text analysis, and case-based reasoning.

The versatility and intuitiveness of our similarity measure could benefit AI systems. By providing a method to account for feature importance and oppositional alignment, our measure introduces a level of adaptability that is crucial for personalized, context-aware AI solutions. Future work may explore the integration of this similarity measure into different AI methodologies or its adaptation to specific challenges within emerging domains of AI. Furthermore, empirical validation across diverse datasets and comparative studies with existing measures would provide deeper insights into its efficacy and applicability.

References

[1] A. S. Shirkhorshidi, S. Aghabozorgi, T. Y. Wah, and A. Dalby, "A comparison study on similarity and dissimilarity measures in clustering continuous data," *PLoS ONE*, vol. 10, 2015.

[2] J. Liu, L. Hou, X. Pan, Q. Guo, and T. Zhou, "Stability of similarity measurements for bipartite networks," *Scientific Reports*, vol. 6, 2015.

[3] S. Santini and R. Jain, "Similarity measures," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, pp. 871–883, 1999.

[4] M. Kryszkiewicz, "The cosine similarity in terms of the euclidean distance," 2014.

[5] D. Liu, X. hong Chen, and D. Peng, "Some cosine similarity measures and distance measures between q-rung orthopair fuzzy sets," *International Journal of Intelligent Systems*, vol. 34, pp. 1572 – 1587, 2019.

[6] X. Mian, J. Ling, and W. Guanqin, "Chaotic local weighted linear prediction algorithms based on the angle cosine," *Systems Engineering Procedia*, vol. 5, pp. 334–339, 2012.

[7] M. Kryszkiewicz, "Determining cosine similarity neighborhoods by means of the euclidean distance," 2013.

[8] B. Li and L. Han, "Distance weighted cosine similarity measure for text classification," 2013.

[9] K. B. Fard, M. Nilashi, M. Rahmani, and O. Ibrahim, "Recommender system based on semantic similarity," *International Journal of Electrical and Computer Engineering*, vol. 3, no. 6, pp. 751–761, 2013.

[10] P. Resnik, "Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language," *J. Artif. Intell. Res.*, vol. 11, pp. 95–130, 1999.

[11] T. Liao, Z. Zhang, and C. Mount, "Similarity measures for retrieval in case-based reasoning systems," *Appl. Artif. Intell.*, vol. 12, pp. 267–288, 1998.

[12] G. Coletti and B. Bouchon-Meunier, "A study of similarity measures through the paradigm of measurement theory: the classic case," *Soft Computing*, 2019.