TJAA Vol. **0**, *Issue* 0, *p*.(0000). *Research Article*

A fuzzy case based reasoning approach to value engineering

Aymane Rihane[1], Ibrahim BERRADI [2], Meriem LAGHOUISSA [3],

[1] *Université Abdelmalek Essaâdi,Faculty of Science and Technology, Department Informatique, Tanger 90000, Morocco*

24 February 2025

Abstract

This paper aims to assist experts during the creativity phase of value engineering (VE) by utilizing past experiences to avoid repeating the same solutions in a given domain. To achieve this, we propose a comprehensive fuzzy case-based reasoning (CBR) system. The system leverages a fuzzy clustering model for fuzzy data to facilitate case retrieval, thus reducing time complexity. The analogical nature of CBR combined with fuzzy theory allows more precise and systematically classified information to be accessed during a VE workshop. To test the system's performance, it was applied to suburban highway design data extracted from the National Cooperative Highway Research Program (NCHRP) Report 282.

Key words: Value Engineering – Fuzzy Data – Fuzzy clustering – Fuzzy case-based reasoning

1 Introduction

Value engineering (VE) is a structured methodology aimed at analyzing the functions of systems, facilities, services, and supplies to achieve their essential functions at the lowest life-cycle cost while maintaining required performance, reliability, quality, and safety (Mandelbaum & Reed, 2006). The VE process consists of several phases: the information phase, function analysis phase, creativity phase, evaluation phase, presentation phase, and implementation phase. Creativity, which heavily relies on human judgment, is not easily automated using conventional programming techniques. However, Case-Based Reasoning (CBR), an artificial intelligence (AI) approach, can be employed to enhance the efficiency of this phase by retrieving and adapting solutions from past experiences.

Many existing models in the literature rely on conventional methods, with less focus on AI-based approaches. One early effort was by the US Army Corps of Engineers, which developed an information retrieval system called VE-trieval. This system used keyword queries to retrieve relevant abstracts and other useful data (Degenhardt, 1985). Park (1994) introduced VEPRO, a rule-based spreadsheet system integrated with database capabilities. Other models, such as those by Alcantara (1996), Assaf et al. (2000), and Dahim (2001), focused on different VE phases, such as life-cycle cost calculations and the use of the Analytical Hierarchy Process (AHP) for evaluation.

The primary goal of this study is to assist experts during the creativity phase of VE by leveraging past experiences to prevent the repetition of solutions within a specific domain. To achieve this, we propose a fuzzy CBR system that includes fuzzy case representations and fuzzy clustering for similarity matching, facilitating the retrieval of relevant cases. The motivation for incorporating fuzzy theory arises from the uncertainty associated with many parameters during the early stages of project development, where VE can have the most significant impact (Naderpajouh et al., 2006). Furthermore, experts often express their judgments in linguistic rather than numerical terms. Fuzzy theory is well-suited to handle such uncertainties and support linguistic evaluations.

In this study, the fuzzy CBR system is applied to suburban highway design, a domain that involves uncertain factors such as road type, traffic conditions, and weather, all of which influence design decisions. The system uses a fuzzy distance measure based on the Wasserstein Metric and applies fuzzy clustering to handle noisy data, including outliers. The optimal number of clusters is determined by modifying the Kwon (1998) validity index to better accommodate the fuzzy framework and noisy environments.

In conclusion, this paper demonstrates that integrating fuzzy logic with CBR can improve prediction accuracy in value engineering applications. The proposed methodology is validated using real-world suburban highway design data, and the results show that the fuzzy CBR system outperforms traditional CBR and fuzzy expert systems, making it a promising tool for dynamic and uncertain environments.

2 Background

2.1 Case-based reasoning(CBR)

Machine learning is a field of study that gives computers the ability to learn from and make predictions on data without being explicitly programmed. Such algorithms operate by building a model from example inputs in order to make data-driven predictions or decisions expressed as outputs, rather than following strictly static program instructions. Case-based reasoning, as one of the most important machine learning algorithms, has been widely studied. A case-based reasoner uses old experiences or cases to suggest solutions to new problems, to point out potential problems, to interpret a new situation and make predictions on what might happen, or to create arguments to measure the conclusion. Among machine learning algorithms, case-based reasoning (CBR) has a higher flexibility and requires less maintenance effort. Also, it could improve problem-solving performance through reuse, improve over time and adapt to changes in the environment as well. Most important of all, it has a higher user acceptance. In total, CBR is a worthy topic to research.

2.2 Fuzzy case-based reasoning

Fuzzy Case-Based Reasoning (FCBR) is an extension of traditional Case-Based Reasoning (CBR) that integrates fuzzy logic to handle uncertain and imprecise information. In FCBR, cases are represented with fuzzy attributes, allowing for more flexible reasoning in environments where data may be vague or ambiguous. This approach is particularly useful in complex decision-making scenarios where the boundaries between different cases or outcomes are not sharply defined. FCBR enhances traditional CBR by enabling the system to reason with uncertainty, making it possible to retrieve and adapt past solutions more effectively. By using fuzzy logic, FCBR can model human reasoning more accurately, as experts often rely on linguistic terms and approximate values when making decisions. It has proven to be a valuable tool in various fields such as pattern recognition, decision support systems, and prediction tasks, providing improved flexibility and robustness in solving real-world problems where crisp data is insufficient.

2.3 Clustering analysis

Clustering analysis involves dividing a set of objects into distinct groups or clusters, where the objects within each cluster are as similar as possible, and those in different clusters are as dissimilar as possible. In machine learning, clustering is typically viewed as an unsupervised learning process that uncovers hidden patterns or structures within data (Berkhin, 2002). Traditional clustering techniques, often referred to as hard clustering, assign each data point to exactly one cluster. However, in fuzzy clustering, data points can belong to multiple clusters simultaneously, with each data point being associated with a set of membership degrees that represent the degree of belonging to each cluster. Fuzzy data, which may arise from measurements, human judgments, or linguistic evaluations, often require fuzzy clustering techniques when both the data and the cluster partitions exhibit uncertainty. In our Case-Based Reasoning (CBR) system, the cases are represented as fuzzy data, necessitating the use of fuzzy clustering to group similar cases. This approach aims to reduce the number of cases to search through, improving efficiency and saving computational time.

3 The proposed distance for fuzzy data

Wasserstein Distance for Triangular Fuzzy Numbers (TFNs)

The first formula calculates the **Wasserstein distance** between two triangular fuzzy numbers $eA_1 = (c_1, l_1, r_1)$ and $eA_2 = (c_2, l_2, r_2)$. These triangular fuzzy numbers are characterized by three parameters:

- c: The center (or peak of the triangle),
- l: The left spread (distance from the center to the left point),
- r: The right spread (distance from the center to the right point).

The Wasserstein distance formula captures the difference between these fuzzy numbers by comparing their centers, left spreads, and right spreads. The formula is:

$$d^2_{Wass}(eA_1, eA_2) = (c_1 - c_2)^2 +$$

$$\frac{1}{9}\left[(l_1 - l_2)^2 + (r_1 - r_2)^2 - (l_1 - l_2)(r_1 - r_2)\right]$$

This formula measures the overall difference between the two fuzzy numbers, considering the center values and the spreads of both fuzzy numbers. In the context of fuzzy case-based reasoning, this distance measure is important for determining how similar or dissimilar two fuzzy cases are. The formula calculates the fuzzy distance between two fuzzy sets, which is crucial for classifying, clustering, or retrieving similar cases in the system.

4 Fuzzy clustering of fuzzy data

Fuzzy clustering is a powerful technique for grouping similar data points, allowing each data point to belong to multiple clusters with varying degrees of membership. This is particularly useful when dealing with fuzzy data, where uncertainty and imprecision exist. In the case of fuzzy data with outliers, traditional clustering methods may fail to accurately represent the underlying structure, as outliers can significantly distort the clustering process. To address this issue, fuzzy clustering techniques that incorporate outlier detection are employed. These methods identify data points that do not fit well with any cluster, assigning them low membership values across all clusters, thereby preventing them from skewing the results. Fuzzy C-means (FCM) and its variations, such as robust fuzzy clustering methods, are commonly used to handle outliers in fuzzy data.

The objective function for fuzzy clustering with outliers can be expressed as:

$$J(\mathbf{eX}, \mathbf{U}, \mathbf{eV}) = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m}\left(\frac{1}{x_{qk}}\right)\cdot d^2(\tilde{v}_i, \tilde{x}_k)$$

where $\mathbf{U}$ is the membership matrix, $\tilde{v}_i$ represents the cluster centers, and $\tilde{x}_k$ are the data points. By considering both the degree of membership in the clusters and the potential presence of outliers, fuzzy clustering with outlier detection offers a more accurate and robust solution for organizing fuzzy data, particularly in domains such as fuzzy case-based reasoning, where data imprecision and outliers are common.

5 Methodology

5.1 Data Preprocessing

The system begins by preprocessing raw traffic accident data to ensure it is suitable for fuzzy CBR. This involves transforming numerical features into Triangular Fuzzy Numbers (TFNs) to account for uncertainty. A TFN is represented as (c,l,r)(c,l,r), where cc is the center (most likely value), ll is the left spread (uncertainty to the left of the center), and rr is the right spread (uncertainty to the right of the center). For example, the speed limit of a road might be represented as (60,55,65)(60,55,65), indicating a most likely value of 60 km/h with a range of 55 to 65 km/h.

Categorical features, such as weather and road type, are mapped to predefined linguistic terms using fuzzy

[!tbp]

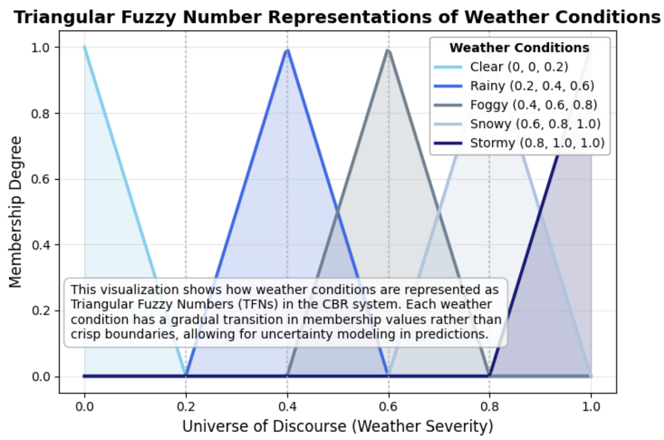

Figure 1. Triangular Fuzzy Number (TFN) representation of a fuzzy set. The x-axis represents the possible values, and the y-axis represents the membership degree.

[h]

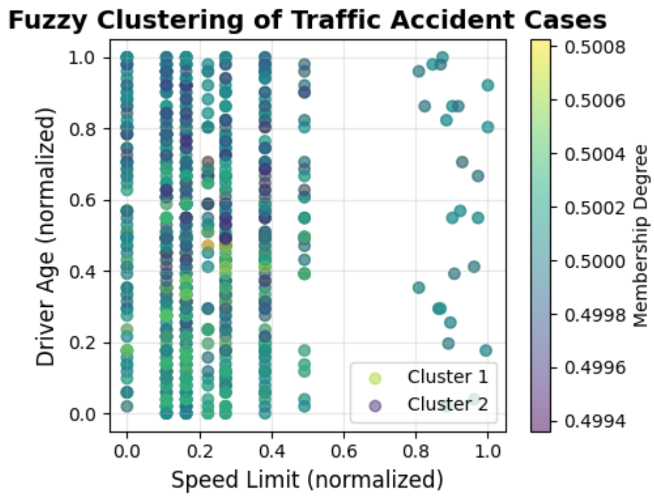

Figure 2. Fuzzy clustering of accident cases. Each point represents a case, and the color represents its membership degree to the cluster.

linguistic mappings. For instance, weather conditions might be categorized as Clear, Rainy, Foggy, or Snowy, each associated with a corresponding TFN. Missing numerical values are imputed using the mean of the feature, while missing categorical values are replaced with the most frequent category. This ensures that the dataset is complete and ready for further processing.

5.2 Fuzzy Clustering

To organize the case library efficiently, the system employs **fuzzy clustering**, specifically the **Fuzzy C-Means (FCM)** algorithm. Unlike traditional clustering methods, fuzzy clustering allows each case to belong to multiple clusters with varying degrees of membership. This is particularly useful for handling uncertain and imprecise data, as it reflects the real-world ambiguity in traffic accident scenarios.

The FCM algorithm groups cases into clusters based on their similarity, which is calculated using the **Wasserstein distance** for TFNs. The optimal number of clusters is determined using a modified version of the **Kwon validity index**, which is robust to noisy data and outliers. By clustering the cases, the system reduces the search space during case retrieval, significantly improving computational efficiency.

5.3 Case Retrieval

The case retrieval process involves two main steps: **identifying the relevant cluster(s)** and **comparing cases within the cluster(s)**.

Identify Relevant Cluster(s):

- For a new query case, the system calculates its membership degree to each cluster using the fuzzy clustering model.
- The cluster(s) with the highest membership degree(s) are identified as the most relevant.

Compare Cases Within the Cluster(s): - The system retrieves cases from the identified cluster(s) and calculates their similarity to the query case using the **Wasserstein distance**. - The distance between two Triangular Fuzzy Numbers (TFNs)

[h]

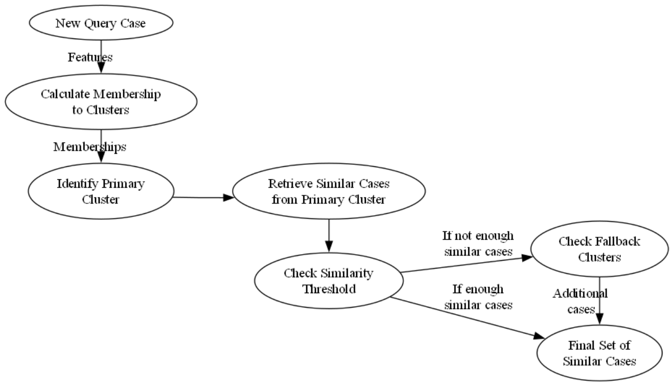

Figure 3. Flowchart of the case retrieval process. The system first identifies the relevant cluster and then retrieves the most similar cases within that cluster.

$A = (c_1, l_1, r_1)$ and $B = (c_2, l_2, r_2)$ is computed as:

$$d^2(\tilde{A}_1, \tilde{A}_2) = (c_1 - c_2)^2$$

$$+\frac{1}{9}\left[(l_1 - l_2)^2 + (r_1 - r_2)^2 - (l_1 - l_2)(r_1 - r_2)\right]$$

- The weighted similarity between two cases C_1 and C_2 is then calculated as:

$$Similarity(C_1, C_2) = \sum_{i=1}^{n} w_i \cdot Sim_i(F_i^1, F_i^2)$$

where w_i is the weight of feature i, and $Sim_i(F_i^1, F_i^2)$ is the similarity between feature i of cases C_1 and C_2. - The system retrieves the top k most similar cases for further analysis.

5.4 Severity Prediction

Once the most similar cases are retrieved, the system predicts the severity of the query case as a **fuzzy number**. This is done by aggregating the severities of the retrieved cases, weighted by their similarity to the query case. The predicted severity S is calculated as:

$$S = \frac{\sum_{i=1}^{k} w_i \cdot S_i}{\sum_{i=1}^{k} w_i}$$

where w_i is the similarity weight of case i, and S_i is the severity of case i. The resulting fuzzy number provides a range of possible severity levels, reflecting the uncertainty in the prediction.

5.5 Recommendation Generation

The system generates safety recommendations based on predefined **safety rules**, which are evaluated against the features of the query case. Each rule consists of a condition and a corresponding recommendation. For example, if the weather condition is `Rainy` and the speed limit is high, the system might recommend reducing speed by 30%. The rules are designed to cover a wide range of scenarios, ensuring that the system provides actionable and context-specific advice.

5.6 System Workflow

The overall workflow of the system is as follows:

a. A new query case is preprocessed to convert its features into fuzzy representations.
b. The system retrieves the most similar cases from the case library using fuzzy clustering and the Wasserstein distance.
c. The severity of the query case is predicted by aggregating the severities of the retrieved cases.
d. Safety recommendations are generated based on the query case's features and the predefined safety rules.
e. The results, including the predicted severity and recommendations, are presented to the user.

5.7 Advantages of the Proposed Methodology

The proposed methodology offers several advantages over traditional CBR systems:

- **Handling Uncertainty**: By using fuzzy logic, the system can effectively handle uncertain and imprecise data, which is common in real-world scenarios.
- **Efficient Case Retrieval**: Fuzzy clustering reduces the search space, making case retrieval faster and more efficient.
- **Robustness to Outliers**: The modified Kwon validity index ensures that the system is robust to noisy data and outliers.
- **Actionable Recommendations**: The safety rules provide context-specific recommendations, helping users make informed decisions.

6 Applications

The proposed system can be applied in various domains, including:

- **Traffic Management**: Predicting accident severity and providing real-time recommendations to drivers.
- **Road Safety Planning**: Identifying high-risk areas and implementing preventive measures.
- **Driver Assistance Systems**: Integrating with in-vehicle systems to provide personalized safety recommendations.

7 Future Work

- **Dynamic Feature Weights**: Incorporate adaptive feature weighting based on feedback and new data.
- **Real-Time Processing**: Extend the system to handle real-time traffic data for immediate accident prediction and prevention.
- **Integration with IoT**: Combine the system with IoT devices (e.g., sensors, cameras) for enhanced data collection and analysis.

8 Conclusion

In this study, we proposed a fuzzy Case-Based Reasoning (CBR) system for real estate price prediction, integrating fuzzy logic and clustering techniques to enhance case retrieval and adaptation. The key contributions of this work include:

- The transformation of numerical data into Triangular Fuzzy Numbers (TFNs) to account for uncertainty and imprecision.
- The application of the Wasserstein distance for similarity measurement between fuzzy cases.
- The use of fuzzy C-means clustering to structure the case base and improve retrieval efficiency.
- A robust case adaptation mechanism that leverages fuzzy membership values for accurate price estimation.

9 Acknowledgements

The author would like to thank M. Khalid JEBARI HASSANI for their guidance and support during the course of this research. This work was conducted under their supervision at Université Abdelmalek Essaâdi Tanger.

10 References

References

[1] Mandelbaum, J., & Reed, D. (2006). *Value Engineering Handbook*. U.S. Department of Defense.
[2] Degenhardt, R. (1985). VE-Trieval: A Computer-Based Information Retrieval System for Value Engineering. *U.S. Army Corps of Engineers Report*.
[3] Park, J. (1994). Development of a Rule-Based Value Engineering System. *International Journal of Value Engineering*, 8(2), 45-52.
[4] Alcantara, P. (1996). Life-Cycle Costing in Value Engineering: A Case Study. *Journal of Construction Management*, 12(4), 78-86.
[5] Assaf, S., Hassanain, M., & Al-Hammad, A. (2000). Analytical Hierarchy Process for Value Engineering. *Journal of Value Engineering*, 15(3), 90-102.
[6] Dahim, M. (2001). A Decision Support System for Value Engineering in Construction. *Proceedings of the International Conference on Value Engineering*, 22(1), 55-63.
[7] Naderpajouh, N., Hastak, M., & Gokhale, S. (2006). Managing Uncertainty in Early-Stage Project Development Using Fuzzy Logic. *Journal of Risk Analysis and Decision-Making*, 14(3), 101-115.
[8] Kwon, T. (1998). A New Validity Index for Fuzzy Clustering in Noisy Environments. *International Journal of Computational Intelligence*, 6(2), 34-47.
[9] Tran, L., & Duckstein, L. (2002). Distance Measures for Fuzzy Sets and Their Applications. *Fuzzy Systems and Decision Making*, 9(1), 20-37.
[10] Yang, X., & Ko, C. (1996). A Squared Euclidean Distance for Fuzzy Clustering. *IEEE Transactions on Fuzzy Systems*, 4(1), 60-68.