

# FRF-FHG: Constructing Fuzzy Hypergraphs Based on FRF

Ziqian Zhao\*, Xiaotong Zhang

*College of Science, North China University of Science and Technology, Tangshan, China*

*\*Corresponding Author*

**Abstract:** The partitioning mechanisms of fuzzy hypergraphs predominantly rely on empirical rules, lacking explicit traceable paths that constrain model credibility and generalizability. To address this, we propose FRF-FHG, a fuzzy hypergraph construction method based on Fuzzy Random Forest (FRF). By building FRF from data and extracting complete branch paths from root to leaf nodes for each tree, we define fuzzy hyperedges as sample sets satisfying all path feature constraints, achieving effective mapping from rule space to hypergraph structure. To mitigate path redundancy, exponential growth of fuzzy hyperedges, and "densification" issues caused by FRF's inherent branching mechanism, we design a fuzzy hyperedge reduction strategy to precisely retain core association information between samples. Comparative and ablation experiments conducted on four cross-domain datasets demonstrate that FRF-FHG achieves robust and superior classification performance.

**Keywords:** FRF; Fuzzy Hypergraph; Fuzzy Hyperedge Reduction; Classification Model

## 1. Introduction

In the era of big data, the complexity and multi-level nature of data have grown exponentially. Traditional graphs can only depict pairwise interactions between nodes, failing to capture all relational information [1]. As an extension of traditional graphs, hypergraphs [2] connect multiple nodes through hyperedges, providing a visual representation of multi-dimensional relationships and offering a novel perspective for multi-level, multi-dimensional data analysis. However, traditional hypergraphs often rely on rigid decision-making rules to model node associations, making them ill-suited for data with fuzzy characteristics. To address this, researchers have integrated fuzzy set theory with hypergraph theory, proposing fuzzy hypergraphs

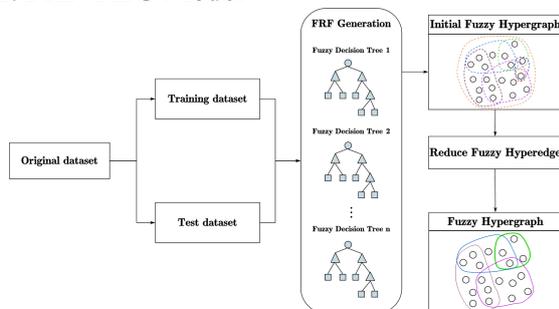
[3].

Although there are various approaches to the construction of current hypergraphs and fuzzy hypergraphs, each has its own limitations. Firstly, a domain name hypergraph is constructed based on a decision tree, and the output result of the second-to-last node of the decision tree is used as a prior condition to form a hyperedge [4], however, single-tree learning is vulnerable to noise and there is a risk of overfitting, which limits the generalization performance of the hypergraph. Secondly, a hypergraph penetration model is constructed based on the factor graph message passing method [5], this method focuses more on the characterization of the hypergraph structure and has limited processing capabilities for data uncertainty and fuzziness. Thirdly, the neighborhood hypergraph based on incomplete information systems [6] has solved the problem of processing discrete attributes in traditional hypergraphs, however, it is limited by the hard partition logic of rough sets, has a weak ability to adapt to the fuzziness of data, and the decision-making basis for the evolution of hypergraphs is not transparent. Fourth, the N-element generalized fuzzy hypergraph construction method [7] expands the representation dimension of fuzzy relations, but the modeling complexity of multivariate relations is relatively high, and the traceability of hypergraph partitioning rules is weak.

The partitioning mechanism of fuzzy hypergraphs mostly relies on empirical rules and lacks traceable paths, which limits the credibility and generalization of the model. Building upon the random forest framework [8], FRF incorporates a fuzzy membership mechanism to model data's ambiguity and uncertainty with precision. This approach not only enhances the model's adaptability to complex scenarios but also improves classification accuracy and robustness. Furthermore, by leveraging fuzzy rules and feature importance analysis, it enhances the interpretability of decision paths, effectively addressing the core limitations of the

fuzzy hypergraph partitioning mechanism. Based on this, this paper proposes an FRF-FHG model for constructing fuzzy hypergraphs based on FRF and verifies it in classification applications. This model takes the branch path feature association rules of FRF as the core basis for fuzzy hyperedge partitioning, extracts the complete decision paths of each tree to generate fuzzy hyperedges. It not only enhances the generalization performance of the fuzzy hypergraph by leveraging the integration advantages of FRF and strengthens the traceability of the fuzzy hyperedge partitioning mechanism through its branch path rules, but also relies on the representation ability of the fuzzy hypergraph for multiple associations. Enhance the interpretability and transparency of the decision-making logic of the FRF model. Due to the inherent characteristics of the FRF model such as "multi-tree parallelism" and "path diversity", the number of generated branch paths is relatively large, resulting in a corresponding increase in the number of fuzzy hyperedges. To avoid the problem of "densest" in hypergraphs, the fuzzy hyperedges are further reduced to achieve efficient modeling of uncertainties and multiple associations in complex data.

**2. FRF-FHG Model**



**Figure 1. Illustrates the Workflow for Constructing the FRF-FHG Model.**

In order to provide clear theoretical support and traceable logic for the division of fuzzy hyperedges, this paper proposes the FRF-FHG fuzzy hypergraph construction method. By constructing the FRF model, the branch paths from the root node to the leaf node of each tree are extracted, and the sample set that satisfies all the features on this path is defined as a fuzzy hyperedge. Aiming at the problems of increased number of fuzzy hyperedges and "denser" of hypergraphs caused by the branch mechanism of the FRF model, the fuzzy hyperedges are further reduced to achieve structural optimization and core association retention of the fuzzy

hypergraph. The specific process is shown in **Figure 1**.

**2.1. FRF Construction**

When constructing an FRF model, the original dataset should be preprocessed first. Then, the parameters of the FRF model are determined using common empirical values in the field of machine learning. Multiple sub-datasets  $D_1, D_2, \dots, D_T$  are randomly drawn from the original dataset D. For each sub-dataset  $D_T$ , the Gini coefficient is used to determine the optimal node splitting point and complete the node splitting. Until the stop condition is met, output the fuzzy decision tree  $T_t$ . Repeat this process to generate multiple fuzzy decision trees until the parameter requirements of the FRF model are met. Output the FRF model  $FRF = \{T_1, T_2, \dots, T_t\}$ .

**2.2 Construction of Initial Fuzzy Hypergraph**

The FRF model outputs prediction results through voting from multiple trees, but its internal decision logic is implicit and decentralized, making it difficult to analyze intuitively. In order to make the decision logic of the FRF model explicit and associate it with the structure of the fuzzy hypergraph, and provide a computable foundation for subsequent redundancy reduction and weighted prediction, a method based on the branching of the FRF model is adopted to construct the fuzzy hypergraph, as follows.

Assuming that the internal nodes of the fuzzy decision tree T have a fuzzy feature splitting condition, and the leaf nodes correspond to the fuzzy class membership distribution of the samples, extracting the branch path P from the root node to the leaf node can represent a finite sequence of fuzzy feature splitting conditions:

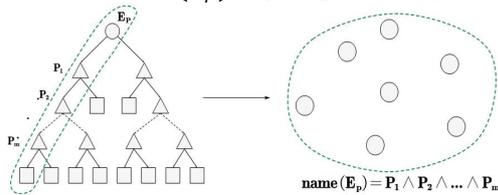
$P = \{p_1, p_2, \dots, p_m\}$ , Where m is the number of split nodes contained in path P,  $p_i$  representing the feature corresponding to the i-th split node on path P and its corresponding fuzzy membership function is  $\mu_{p_i} \in [0, 1]$ .

As illustrated in **Figure 2**, for each branch path P, the set of samples that satisfy all the feature conditions on that path is defined as a fuzzy hyperedge  $E_p$ ,

$$E_p = \{x \in D | x \text{ satisfies all conditions on the } P\}$$

where D is the dataset. In order to intuitively reflect the path logic corresponding to fuzzy hyperedges, the naming rule of fuzzy hyperedge

$E_p$  is the conjunctive form of "characteristic-fuzzy membership degree", which is composed of a combination sequence of all the features and their corresponding fuzzy partitions that the path passes through. The fuzzy membership degree of hyperedges is expressed in the form of a set, specifically represented as an ordered set of fuzzy membership degrees of each feature on the path, that is  $name(E_p)=p_1 \wedge p_2 \wedge \dots \wedge p_m$ .



**Figure 2. Naming Method of Fuzzy Hyperedge and Sample Nodes Contained in Fuzzy Hyperedge.**

This construction method is based on the idea of complementary advantages, taking the internal decision-making logic of the FRF model as the core mechanism of fuzzy hypergraph partitioning, and proposes the FRF-FHG model. It not only takes advantage of the integration of FRF to enhance the generalization performance of fuzzy hypergraph construction, but also strengthens the traceability of the fuzzy hyperedge partitioning mechanism through its inherent branch path rules. Moreover, by utilizing the representation ability of fuzzy hypergraphs for multivariate correlations, the interpretability and transparency of the decision-making logic of the FRF model are enhanced, effectively alleviating the "black box" problem. Due to the inherent characteristics of the FRF model's branch mechanism, the number of branch paths it generates is relatively large, which leads to a corresponding increase in the number of fuzzy hyperedges. To avoid the "denser" problem caused by a large number of fuzzy hyperedges, it is necessary to further reduce the fuzzy hyperedges.

**2.3. Fuzzy Hyperdimensional Reduction**

In fuzzy hypergraph modeling, the "densest" feature is prone to occur. This not only increases the computational complexity but also may introduce noise due to redundant fuzzy hyperedges, masking the core correlations between samples. Therefore, when the number of fuzzy hyperedges far exceeds the number of sample nodes, hyperedge reduction becomes a necessary step in model optimization. The core logic is to eliminate redundant and low-

contribution fuzzy hyperedges while retaining key information. This study follows the following optimization logic for the fuzzy hyperedge reduction process, aiming to enhance the effectiveness and predictive ability of the fuzzy hyperedge set:

- 1) Eliminate extreme granularity fuzzy overedges: Eliminate fuzzy overedges that only contain a single sample or cover all samples. Such fuzzy overedges exhibit extreme characteristics in classification granularity, lack generalization ability and statistical significance, and their classification results are easily disturbed by random factors, making it difficult to reflect the regular association of the sample population. Therefore, they should be eliminated.
- 2) Filtering low confidence fuzzy hyperedges: Calculate the prediction accuracy of fuzzy hyperedges as shown in Equation (1):

$$Acc(\tilde{E}) = \frac{1}{|\tilde{E}|} \sum_{x \in \tilde{E}} I(\tilde{y}(x)=y(x)) \tag{1}$$

Among them,  $|\tilde{E}|$  represents the total number of samples covered by the fuzzy hyperedge  $\tilde{E}$ ,  $\tilde{y}(x)$  is the predicted label of the fuzzy hyperedge for sample  $x$ ,  $y(x)$  is the true label, and  $I(\cdot)$  is the indicator function. At that time,  $I(\cdot)=1$  indicates that the fuzzy hyperedge's prediction of this sample is correct. Otherwise, it indicates that the prediction of this sample is incorrect. If the prediction accuracy of a certain fuzzy hyperedge  $Acc(\tilde{E})=0$ , the prediction confidence of the fuzzy hyperedge is determined to be close to zero, and has no effective contribution to the overall classification. In order to screen out the highly consistent fuzzy hyperedge and strengthen the prediction reliability of the set, it is eliminated.

- 3) Merge redundant fuzzy hyperedges and weight them: Merge fuzzy hyperedges containing completely consistent sample sets into a single fuzzy hyperedge, and accumulate their occurrence frequencies to eliminate information redundancy. For each merged fuzzy hyperedge, use the logarithm of its frequency as the weight to avoid excessive dominance of high-frequency fuzzy hyperedges.

Set the merged fuzzy hyperedge  $\tilde{E}=\{\tilde{E}_1, \tilde{E}_2, \dots, \tilde{E}_L\}$ ,  $\forall \tilde{E}_j \in \tilde{E}$ , the weight formula is shown in Equation (2):

$$w(\tilde{E}_j) = \log(freq(\tilde{E}_j) + \alpha) \tag{2}$$

Among them,  $freq(\tilde{E}_j)$  represents the number of redundant fuzzy hyperedges that have been merged into the fuzzy hyperedge, where  $\alpha$  is the

smoothing coefficient and takes the value of 1.

For each sample, calculate the sum of its weights in all fuzzy hyperedges predicted as "yes" and "no", forming score vectors for two types of predictions. The formula for calculating sample weights is shown in Equation (3):

$$W(x_i) = \sum_{j=1}^L w(\widetilde{E}_j^*) \cdot I(x_i \in \widetilde{E}_j^*) \quad (3)$$

Among them,  $I(\cdot)$  is the indicator function, when  $I(\cdot)=1$  indicates the sample  $x_i \in \widetilde{E}_j^*$ , otherwise it indicates the sample  $x_i \notin \widetilde{E}_j^*$ . The final attribute prediction result of the sample is determined by the category with higher scores, that is, the aggregation judgment of sample attributes is achieved through a weighted voting mechanism.

### 3. Experiments and Result Analysis

#### 3.1 Experimental Environment and

**Table 1. Parameter Comparison.**

types of models	Precision	Recall	F1	AUC	Accuracy
FRF-FHG model	0.8889	1.000	0.9412	0.9167	0.9412
No-eliminating	0.7273	1.000	0.8421	0.9028	0.8235
No-filtering	0.8889	1.000	0.9412	0.9804	0.8824
No-merging	0.7273	1.000	0.8421	0.9028	0.8235

**Table 1** shows the influence of different reduction fuzzy hyperedge modules on the method proposed in this paper. Eliminating extreme granularity fuzzy hyperedges and merging redundant fuzzy hyperedges have a significant overall impact on the model, with the reduction of three core indicators exceeding 10%. The AUC value of the model without filtering low-confidence fuzzy overedges rose from 91.67% to 98.04%, but the Accuracy dropped from 94.12% to 88.24%, indicating that low-confidence overedges are prone to misjudgment of some samples. However, filtering low-confidence overedges achieves a more stable actual prediction effect by sacrificing a small amount of potential discrimination ability. From the overall effect, the process of reducing the fuzzy hyperedge is necessary.

#### 3.3 Comparative Experiment

To evaluate the adaptability and stability of

**Table 2. Comparison of Parameters for Each Classification Model.**

data set	types of models	Precision	Recall	F1	AUC	Accuracy
Breast Cancer Diagnosis	RF	0.9444	0.9533	0.9488	0.9898	0.9357
	FRF	0.9450	0.9626	0.9537	0.9895	0.9415
	<b>FRF-FHG model</b>	<b>0.9450</b>	<b>0.9626</b>	<b>0.9537</b>	<b>0.9889</b>	<b>0.9415</b>
	Statistical model	0.7000	1.0000	0.8235	0.6379	0.7000

#### Evaluation Indicators

The research process was conducted on devices with Intel (R) Core (TM) i7-8750H CPU @ 2.20GHz 2.21 GHz and Windows 10 64 bit operating system, running in Anaconda 4.13.0. FRF models were generated and fuzzy hypergraphs were reduced in this environment. Five core evaluation metrics were introduced to analyze model performance, including Precision, Recall, F1 score, AUC value (stability metric), and Accuracy.

#### 3.2. Ablation Experiment

To evaluate the impact of simplified fuzzy hyperedges on the FRF-FHG model's performance, this section performed an ablation experiment on the Watermelon dataset by removing one simplification step during the hyperedge reduction process. The resulting performance changes are presented in **Table 1**.

different classification models in complex scenarios, this study selected four typical cross-domain tasks: breast cancer diagnosis (healthcare), shopping cart purchase behavior recognition (consumption recommendation), credit evaluation (risk control), and Titanic survival prediction (emergency decision-making). A comparative framework was established, incorporating five models: traditional Random Forest (RF), Fuzzy Random Forest (FRF), FRF-FHG model, statistical model [9], and hypergraph neural network model [10]. Performance was analyzed using five core evaluation metrics. The baseline experimental parameters remained consistent, with fuzzy membership values categorized into "low/middle/high" fuzzy sets and normalized. While fuzzy interval divisions differed slightly, core parameters remained identical. Model parameters are detailed in **Table 2**.

	Hypergraph neural network model	0.8367	0.9200	0.8764	0.8652	0.8133
Market Basket Analysis	RF	0.8000	0.1905	0.3077	0.9886	0.9400
	FRF	0.7407	0.9524	0.8333	0.9944	0.9733
	<b>FRF-FHG model</b>	<b>0.7500</b>	<b>1.0000</b>	<b>0.8571</b>	<b>0.9940</b>	<b>0.9767</b>
	Statistical model	0.8586	0.7876	0.8164	0.9317	0.9200
	Hypergraph neural network model	0.8710	0.8261	0.8480	0.9423	0.8967
credit	RF	0.9080	0.8587	0.8827	0.9604	0.8986
	FRF	0.9000	0.8804	0.8901	0.9601	0.9034
	<b>FRF-FHG model</b>	<b>0.8854</b>	<b>0.9239</b>	<b>0.9043</b>	<b>0.9605</b>	<b>0.9130</b>
	Statistical model	0.8396	0.9674	0.8990	0.9629	0.9034
	Hypergraph neural network model	0.8140	0.8500	0.8317	0.8923	0.8310
Titanic Survival Prediction	RF	0.7273	0.6990	0.7129	0.8386	0.7836
	FRF	0.8049	0.6408	0.7135	0.8589	0.8022
	FRF-FHG model	0.8272	0.6505	0.7283	0.8652	0.8134
	Statistical model	0.9091	0.3883	0.5442	0.8059	0.7500
	<b>Hypergraph neural network model</b>	<b>0.7654</b>	<b>0.7321</b>	<b>0.7484</b>	<b>0.8567</b>	<b>0.7989</b>

Overall, in the breast cancer dataset, the FRF-FHG model is tied for the best with FRF, meeting the medical requirements of dual low risk. In the shopping basket dataset, the FRF-FHG model performs the best, achieving zero missed detections and controllable invalid recommendations. In the credit assessment task, the FRF-FHG model remains optimal and can balance bad debt risk and business returns. In the Titanic survival prediction mission, the hypergraph neural network model was the best, adapted to the rescue priority logic. Although the FRF-FHG model was suboptimal, it had no fatal flaws. Although the Recall was 65.05%, slightly lower than the former's 73.21%, the Precision was 82.72%, better than the former's 76.54%. From the dimension of resource consumption, The FRF-FHG model has significant advantages and can be applied to special scenarios where there is an extreme fear of resource waste. The horizontal cross-task comparison shows that the FRF-FHG model achieves optimal or co-optimal performance in three tasks and suboptimal performance in one task, with no invalid scenarios. This fully demonstrates its outstanding cross-scenario adaptability and performance stability.

In conclusion, the FRF-FHG model can serve as an overall optimal model, providing high-value decision support for complex classification tasks in multiple domains.

#### 4. Conclusions

To address diverse and complex classification challenges, this study introduces the FRF-FHG model, which constructs a fuzzy hypergraph using branch paths from each fuzzy decision tree

in the FRF model for optimization. Comparative experiments with four classification models, four cross-domain datasets, and five core evaluation metrics demonstrate the FRF-FHG model's superior overall performance, validating its adaptability and stability in complex classification scenarios. By effectively integrating the rule learning capabilities of the FRF model with the complex relationship representation of the fuzzy hypergraph, the FRF-FHG model achieves outstanding cross-scenario adaptability and performance stability, while also offering high structural flexibility and practical application value. This makes it a robust solution for decision-making needs in multi-domain complex classification tasks.

#### References

- [1] Zhang Z, Mei X, Jiang H, et al. Dynamical analysis of Hyper-SIR rumor spreading model[J]. Applied Mathematics and Computation, 2023;446:127887.
- [2] Fang Y, Zheng Y. Metric learning based on attribute hypergraph[C]//2017 IEEE International Conference on Image Processing (ICIP). IEEE, 2017;3440-3444.
- [3] Yin T, Ding W, Ju H, et al. The fuzzy hypergraph neural network model based on sparse k-nearest neighborhood granule[J]. Applied Soft Computing, 2025;112721.
- [4] Wei J X, Long C, Fu H, et al. Malicious Domain Name Detection Method Based on Enhanced Embedded Feature Hypergraph Learning[J]. Journal of Computer Research and Development, 2024;61(09):2334-2346.
- [5] Bianconi G, Dorogovtsev N S. Theory of percolation on hypergraphs[J]. Physical

- Review E, 2024;109(1-1):014306.
- [6] Shi J. A Research for Neighborhood Hypergraph Based Classification Algorithm for Incomplete Information System[D]. Chongqing University of Posts and Telecommunications, 2016.
- [7] Wang X, Yang Z, Han Y, et al. Method of sequential intention inference for a space target based on meta-fuzzy decision tree[J]. Advances in Space Research, 2024;74(8):4050-4067.
- [8] Qiu S, Wang Q, Xiong M, et al. Adaptive extraction of rotating Doppler shift based on random forest regression[J]. Physics Open, 2025;100347.
- [9] Han R. Research on Automatic Classification of Intangible Cultural Heritage Products Based on Statistical Models[J]. Industrial Applications and Communications, 2022;41(08):158-160.
- [10] Su Y J. Construction and Analysis of Hypergraph Neural Network Model by Learning Line Graph Expansion and its Classification in Alzheimer's Disease[D]. Taiyuan University of Technology, 2023.