Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 14, No. 2, 2025

Multimodal Data Fusion Methods for Mixed Sets of Fuzzy and Precise Numbers

Yucheng Zhang[1, 2, 3, 4], Shibin Zhang[1, 3, 4, *], Lili Yan[2, 3, 4], Yan Chang[2, 3, 4], Zhi Qin[2, 3, 4], Min Yang[2, 3, 4]

1College of Artificial Intelligence (Xin Gu Industrial College), Chengdu University of Information Technology, Chengdu, 610225, China
2School of Cybersecurity (Xin Gu Industrial College), Chengdu University of Information Technology, Chengdu, 610225, China
3Advanced Cryptography and System Security Key Laboratory of Sichuan Province, Chengdu, 610225, China
4SUGON Industrial Control and Security Center, Chengdu, 610225, China
[*]Corresponding author

Abstract: This paper presents a novel data fusion framework to address the challenges of extracting valuable information from big data characterized by its "5V" attributes. The framework integrates Dempster-Shafer Evidence Theory (DSET) for handling fuzzy data and Information Theory for optimizing the fusion process, resulting in a comprehensive and robust approach. A new algorithm for data preprocessing and classification is proposed, utilizing K-means clustering to differentiate between fuzzy and precise data. Fuzzy data is processed using DSET, while the Bhattacharyya Jensen-Shannon Divergence (BJS divergence) is employed to quantify uncertainty and information loss more accurately than traditional methods. The framework includes an enhanced credibility evidence fusion algorithm that assigns higher credibility to well-supported evidence, ensuring reliable decision-making. Evaluated using a medical dataset, the framework demonstrates significant improvements in performance metrics such as accuracy and F1-score. The study concludes that the proposed method effectively bridges the gap between fuzzy labels and precise classification, offering a universal solution for data fusion that outperforms existing approaches and sets a new benchmark for future research in data fusion technology.

Keywords: Data Fusion; Dempster-Shafer Evidence Theory; Bhattacharyya Jensen-Shannon Divergence; Information Theory.

1. Introduction

In the era of big data, the sources of data are becoming more and more extensive, and the types of data are becoming more and more diverse. Researchers are facing unprecedented challenges of massive data. However, due to the "5V" characteristics of big data (Volume, Variety, Value, Veracity, Velocity), it is difficult for people to fully mine the implicit information in big data. Therefore, people urgently need effective means to integrate massive data and mine valuable information for use, thus transcending the limitations of a single data source. Through data fusion technology, a deeper understanding and analysis of the data can be realized.

Data fusion technology, as an important data processing method, has a wide range of applications, such as target detection [1-3], battlefield assessment [4], medical diagnosis [5-7], remote sensing mapping [8-10], fault diagnosis [11-14], intelligent manufacturing [15-16], etc. In the field of intelligent manufacturing, data fusion technology has effectively improved people's processing ability and utilization efficiency of industrial big data. Through data fusion technology, massive, high-dimensional, multi-source heterogeneous noisy industrial data are cleaned, denoised, integrated, modeled, and classified on multiple scales, providing reliable data resources for subsequent correlation analysis, performance prediction, and optimization decision-making. Therefore, data fusion technology has strongly promoted the development of application fields.

Data association is a key step in data fusion, involving determining the correspondence between a set of observed data and a specific target. Common data association techniques include the nearest neighbor method (NN), *K*-means clustering (*K*-Means), probabilistic data association (PDA), joint probabilistic data association (JPDA), and multiple hypothesis testing (MHT), etc.[17]. In addition, the data fusion system needs to make decisions. Decision fusion technology usually uses symbolic information and considers uncertainty and constraints during the fusion process. Common decision fusion methods include Bayesian methods, Dempster-Shafer reasoning, deductive reasoning, and semantic methods, etc.[17].

Inspired by the McGurk effect[18] in the audio-visual speech recognition task, scholars have begun to explore the theory and methods of multimodal data fusion. With the development of deep learning technology, multimodal data fusion has ushered in new opportunities. For example, Dai and others' research was inspired by the McGurk effect, which demonstrates the importance of audio-visual information fusion for human perception. Although traditional machine learning technology has limitations in multimodal data fusion, the development of deep learning technology has brought new opportunities to this field. Dai and others' research "Analysis of multimodal data fusion from an information theory perspective" deeply discusses the theoretical basis of multimodal data fusion, especially the relationship between data information and model performance from the perspective of information theory. They pointed out that although deep learning technology provides new possibilities for processing complex multimodal data, there is currently a lack of theoretical analysis methods that link data information with model performance. To this end, they proposed a set of basic concepts and principles, including defining the model accuracy limit of multimodal tasks, and proved that when additional modal data provides more effective information, the performance of multimodal models can be improved in theory. In addition, they were the first to quantify the information loss in the representation space of deep learning[19], providing a new perspective for understanding and improving the information fusion process in deep learning models.

Although fuzzy logic has performed well in processing multimodal data, how to effectively handle fuzzy information from different modalities in multimodal data fusion is a challenge. It is necessary to study how to integrate various types of fuzzy information to obtain a more comprehensive data view. More effective algorithms need to be developed to integrate fuzzy logic and deep learning to improve the performance of data fusion. At the same time, how to effectively represent and reason about fuzzy information is an area that few researchers have ventured into. Current methods may not fully capture all the nuances of fuzzy information, and more advanced models are needed to improve the expression and processing capabilities of fuzzy information.

This study proposes an innovative data fusion framework aimed at overcoming existing problems and providing more comprehensive, robust, and efficient solutions through the following means:

1) Innovative combination of fusion methods: By combining DS evidence theory and information theory, a framework capable of handling fuzzy data is constructed to improve the framework's ability to process and fuse fuzzy information, thereby extracting richer information.

2) Application of BJS divergence: The introduction of BJS divergence quantifies uncertainty and information loss, solving the limitations of using KL divergence in data fusion problems from an information theory perspective when dealing with fuzzy data. It can better handle uncertainty and incompleteness in probability distributions. When fusing fuzzy information, BJS divergence can provide a more accurate similarity measure. In addition, when two probability distributions do not overlap at all, KL divergence may become meaningless or lead to the disappearance of gradients, and BJS divergence effectively solves this problem. BJS divergence also has boundedness, and its value is always between 0 and 1, making it more intuitive and easy to interpret in data fusion models.

3) Algorithm design and performance improvement: An effective label set coverage algorithm and a classifier construction method based on DS evidence theory are proposed, and comprehensive understanding of data is considered, taking into account the complementarity between modalities to reduce computational complexity and improve fusion performance.

Through the above design, this study not only promotes the theoretical development of data fusion technology but also provides new tools and methods for data processing in practical applications.

2. Related Work

1) Dempster-Shafer Evidence Theory (DSET):

DSET is a mathematical framework for representing and processing uncertain information. It quantifies uncertainty in the sample space through basic probability assignment (BPA). Evidence theory is a general extension of Bayesian theory. Evidence theory provides an effective method for fusing evidence from different sources. The main advantage of evidence theory over Bayesian probability theory is that it can not only assign mass to hypotheses corresponding to individual classes but also explicitly represent the union of these classes as unknown. The trust function provides a more flexible model for representing uncertainty than probability distribution. In Bayesian probability theory, all prior probabilities must be given before the fusion process, while evidence theory can start from complete ignorance. The prior information required by evidence theory is easier to obtain than that of Bayesian probability theory.

In evidence theory, the mass function, also known as basic probability assignment, abbreviated as BPA. If $m(A)$ >0, then A is a focal element. The union of focal elements is called the core of the mass function. A piece of evidence's BPA reflects the degree of support for proposition A in the recognition framework. Assuming there are n BPAs on the recognition framework, denoted as:

$$m(A)=\begin{cases}\frac{1}{1-K}\sum_{\cap A_i=A,A_i\in\theta}\prod_{i=1}^{k}m_i(A_i),\ A\neq\emptyset\\ 0,\ A=\emptyset\end{cases} \tag{1}$$

$$K=\sum_{\cap A_i=A,A_i\in\theta}\prod_{i=1}^{k}m_i(A_i) \tag{2}$$

2) Information Theory Basics

Information entropy represents the uncertainty of information, $H(x)=-\sum_{i=1}^{n}p(x_i)logp(x_i)$, where $p(x_i)$ represents the probability of event x_i occurring. Conditional entropy represents the uncertainty of random variable X under the condition that random variable Y is known. $H(X|Y)=-\sum_{x,y}p(x,y)\log(p(x|y))$. When an observation sample c_i of an entity ei has the same features and labels as another observation cj, it is called a consistent sample. Consistent samples can increase the statistical belief of the model and reduce the model's belief entropy. When an observation sample c_i of an entity ei has different features but the same label as another observation cj, it is called a contradictory sample. Contradictory samples can reduce the statistical belief of the model and increase the model's belief entropy. Contradictory samples may come from noisy data and partially observable data. To address the uncertainty of contradictory samples, a direct method can be adopted to extend the observation dimension, including adding more observable modal channels to reduce model belief entropy and minimum model accuracy entropy.

Bhattacharyya Jensen-Shannon Divergence is a method for measuring the difference between two probability distributions. It combines the characteristics of Jensen-Shannon divergence and Bhattacharyya distance, which can better reflect the similarity and difference between probability distributions.

$$BJS(P||Q)=BJS(Q||P) \tag{3}$$

$$[BJS(P||Q)=frac\{1\}\{2\}(H(P,Q)+H(Q,P))] \tag{4}$$

Where $H(P,Q)$ represents the joint entropy.

Comparison of BJS Divergence and KL Divergence: KL divergence is another method for measuring the difference between two probability distributions, but it does not satisfy symmetry. BJS divergence, as a symmetric divergence, can more fairly reflect the difference between two probability distributions. Especially in data fusion from an information theory perspective, BJS divergence provides an effective means to balance accuracy and fuzziness.

$$[D\_\{KL\}(P||Q)=sum\_\{i=1\}\hat{}\{n\}p(x\_i)log(frac\{p(x\_i)\}\{q(x\_i)\})] \tag{5}$$

where $[p(x\_i)]$ and $[q(x\_i)]$ represent the true distribution of the data and the data distribution predicted by the model, respectively.

3. Proposed Method

3.1. Overall Process of Data Fusion

For data to be fused, first perform data preprocessing to transform the data into a matrix form convenient for computer processing. For all data, if it is precise data, it is not processed first; if it is fuzzy data, it is processed based on DS evidence theory for data fusion, which can effectively clean the fuzzy data into precise data. Then, the cleaned fuzzy data and the original precise data are fused together from the perspective of information theory, and the final fusion result is obtained.

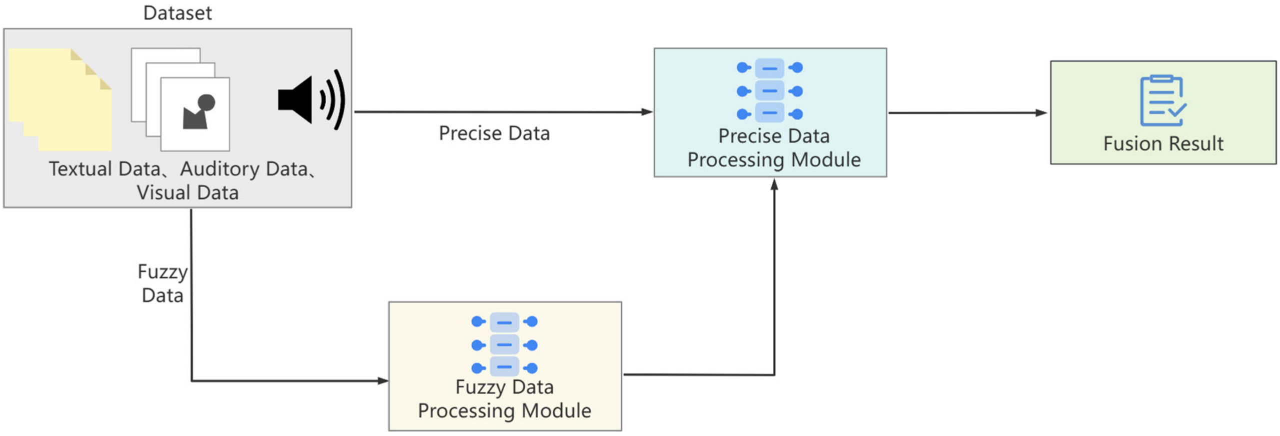

Figure 1. Overall Process of Data Fusion

Data Preprocessing: Data preprocessing is the foundation of data analysis, involving steps such as data cleaning, feature extraction, and data standardization. The purpose is to transform the original data into a matrix containing useful information, laying the foundation for subsequent data fusion.

Data Classification: Data classification is achieved through unsupervised learning methods, where the K-means clustering algorithm is used to distinguish between fuzzy data and precise data. The K-means algorithm assigns data points to K clusters through an iterative process until convergence. We chose the K-means algorithm because it is simple, fast, and easy to understand. In addition, this paper also considers the quality of clustering, using the silhouette coefficient to evaluate the clustering effect.

Algorithm 1. Data Preprocessing and Classification

```
Input: Raw data D
Output: Classified data set C, including vague and precise data
Begin:
    Cleanse D from noise and handle missing alues.
    Detect and remove outliers from D.
    Normalize the features in D.
    Apply K-means clustering to D to obtain lusters.
    Classify clusters with a high degree of overlap s vague data.
    Classify distinct clusters as precise data.
    Return the classified data set C.
End.
```

3.2. Fuzzy Data Processing Process

Fuzzy Data Processing: In the fuzzy data processing stage, this paper uses DS evidence theory to process fuzzy label datasets. The goal of this stage is to transform fuzzy data into a more easily processed precise data form for subsequent analysis and decision-making. The following is a detailed processing process:

Data Set Partitioning: First, we use the effective label set coverage algorithm to divide the fuzzy label dataset into multiple small data groups. Each data group represents an effective label set, which can cover the entire category space, ensuring the diversity and representativeness of the data.

Data Fusion Based on DS Evidence Theory: For each data subset, we design and train a classifier based on DS evidence theory. These classifiers will use fuzzy labels as evidence of uncertainty and assign probabilities to each category through BPA. The training of the classifier uses advanced machine learning techniques, such as logistic regression, support vector machines, or neural networks.

Model Training and Testing: Use machine learning methods to train our fusion model. With a large amount of training data, the model will learn how to make the best decisions based on weights and information content. In addition, we will also use an independent test set to evaluate the model's generalization ability and performance.

Decision Making: According to the BPA after fusion, we use the pignistic probability transformation to convert the probabilities of multiple hypotheses into the probability of a single category and select the category with the highest probability as the final classification result.

The specific process is as follows:

First, use logistic regression and convolutional neural networks as basic classifiers to process the synthetic dataset. For the synthetic dataset, we generated data with 5 categories following a Gaussian distribution and introduced random noise and interdependencies among features to simulate real-world complexity. Selected six categories with precise category labels and associated them with ambiguous hyper-category labels.

Given as the existing label set. To implement k different classifiers, it is necessary to find k effective label set covers, denoted as, and join the unselected elements of the fuzzy label to select multiple hypotheses until the conditions for effective label set covers are met. After deduplication and simplification operations, an effective label set cover can be obtained, represented as the set of effective label set covers $U = \{c_1, c_2, \dots, c_k\}$.

Algorithm 2. Synthetic Data Classification

```
Input: Raw synthetic dataset S, Number of classifiers k, Existing label set $Y$
Output: Set of effective label cover sets $U = \{C_1, C_2, \dots, C_k\}$
Begin:
        Data Generation
        Generate synthetic dataset S with 5 categories following Gaussian distribution.
            Introduce random noise and interdependencies among features to simulate.
        Label Association
            Associate six categories with precise labels and ambiguous hyper-category labels.
        Base Classifier Selection
            Select Logistic Regression and Convolutional Neural Network as base classifiers.
        Label Cover Set Identification
            Initialize an empty set U for storing effective label cover sets.
        Deduplication and Simplification
        For each label cover set $C_i$ in U, perform deduplication to remove duplicate labels.
                Simplify each $C_i$ by removing redundant labels that do not contribute to the classification diversity.
        Effective Label Cover Set Validation
        Validate each $C_1$in U to ensure they meet the criteria for being effective label cover sets.
        Return Results
                Return the set of effective label cover sets U, which now contains k distinct and effective label cover sets.
End.
```

Based on the number of elements in $\hat{y}_i$, the existing label set $\hat{y}$ is divided into $|S|$ subsets R_j. Secondly, all initial label set covers in the set U are identified and retained. Thirdly, the valid label set covers in U are simplified. Finally, the set of valid label set covers U is deduplicated.

For an m-class classification task, the set of all possible classes is denoted as $S = \{s_1, s_2, \dots, s_m\}$. Given (n) data samples $X \in R^{n \times d}$ and labels $X = \{x_1, x_2, \dots, x_n\}, Y = \{y_1, y_2, \dots, y_n\}$. Each ($x_i$, in {R}^d) is a (d)-dimensional vector, and y_i is a set containing one or more elements, with ($|y_i|$ leq m) (meaning the maximum number of elements in (y_i) is (m)).

Let $S = \{s_1, s_2, \dots, s_m\}$be the set of crisp labels, and y_i be a fuzzy label. Classifying data with fuzzy labels involves learning a classifier from X and Y, where some of the y_i may be the fuzzy labels defined above. The classifier should be capable of categorizing new samples into crisp classes, even if there is no relevant data for such classes in X.

When a sample is labeled, it is sometimes impossible to assign a precise label to the sample. In such cases, more reasonable fuzzy labels are used to avoid incorrect labeling.

Algorithm 3. Fuzzy Label Classification

```
Input: Data samples $X \in R^{(n \times d)}$, Labels Y, Number of classes m.
Output: De-duplicated and simplified set of effective label cover sets $U$
Begin:
    Define Class Set
    Let $S = \{s_1, s_2, \dots, s_m\}$ be the set of all possible classes.
    Initialize Data and Labels
    $X = \{x_1, x_2, \dots, x_n\}$ where each $x_i \in R^d$ is a d-dimensional vector.
    $Y = \{y_1, y_2, \dots, y_n\}$ where each $y_i$ is a set with $|y_i| \leq m$.
    Split Label Set
    For each $y_t$.in Ŷ, split the label set Ŷ into |S| subsets $R\_j$ based on the number of elements in $y_t$.
    Find Initial Label Cover Sets
    Initialize an empty set U for storing label cover sets.
    For each subset $R_j$, find all initial label cover sets and add them to U.
    Simplify Label Cover Sets
        For each label cover set in U, perform simplification to remove redundant labels that do not contribute to the classification.
    Validate Effective Label Cover Sets
    For each label cover set in U, validate to ensure it is effective for classification, considering the ability to classify new samples to precise classes even if not present in X.
    De-duplicate Label Cover Sets
    Perform de-duplication on the set U to remove any duplicate label cover sets.
    Handle Fuzzy Labels
        For samples that cannot be assigned a precise label, use reasonable fuzzy labels to avoid incorrect labeling.
    Return Results
    Return the de-duplicated and simplified set of effective label cover sets U.
End.
```

Assuming considering m mutually exclusive categories, the recognition framework H is constructed as: $\theta = \{H_1, H_2, \dots, H_m\}$, each element represents a specific category, corresponding to a true label. Then, the power set is constructed:

$$2\hat{}\theta = \{\emptyset, \{H_1\}, \{H_2\}, \dots, \{H_m\}, \{H_1, H_2\} \dots\dots \{H_1, H_2 \dots\dots H_m\}\} \quad (6)$$

If there is a high conflict between pieces of evidence, the synthesis rules of evidence fusion may lead to counterintuitive results during the normalization process. Therefore, to avoid this phenomenon, this paper adopts the following strategies: First, if a BPA is supported by other BPAs to a greater extent, its credibility is higher, and the evidence has a greater impact on the final fusion result, which can avoid over-reliance on a single piece of evidence. Then, if there are n base classifiers, in order to increase the confidence of the advantageous subset and provide convergence after weighted averaging, the weighted average evidence is combined with itself n-1 times. There is no risk of potential counterintuitive combination results due to high conflict. It can also provide convergence for subsequent decision-making.

If there is only a single value in the fusion result, you can simply choose the category with the highest probability. However, if there are ambiguous labels in the outputs of two different classifiers, there are multiple hypotheses in the fusion result. This paper proposes a solution to distribute the probabilities of multiple hypotheses according to the base average, called probability conversion. It is a transformation method similar to maximum entropy. The Pignistic probability transformation is defined in the Transferable Belief Model (TBM). The model has a two-layer structure, including the Belief layer and the Pignistic layer. These two layers are usually indistinguishable, and the probability function is used to quantify the confidence on both layers. Confidence is transmitted between pieces of evidence in the Belief layer and is transformed into Pignistic probability using the Pignistic probability transformation. After the Pignistic probability transformation, the fusion result is transformed into m_{da} .

Algorithm 4. Enhanced Credibility Evidence Fusion

```
Input: Belief degree assignments (BPAs) from n base classifiers, Evidence set E.
Output: De-conflicted and converged fusion result $m\_da$.
Begin:
    Assess Belief Degree Support
        For each BPA in E, determine the support it receives from other BPAs.
        Assign higher credibility to BPAs that receive greater support.
    Weighted Average of Evidence
        Calculate the weighted average evidence $m\_wa$, giving more weight to BPAs with higher credibility.
    Enhance Sub-Set Confidence and Convergence
        Combine the weighted average evidence $m\_wa$ with itself n-1 times to increase the confidence of the advantageous subset and ensure convergence.
    Handle Single Value in Fusion Result
        If $m\_wa$ contains only a single value, select the category with the highest probability.
    Handle Multiple Hypotheses in Fusion Result
        If there are ambiguous labels from different classifiers in m_wa, apply probability conversion to distribute the probabilities among the hypotheses.
    Probability Conversion
        Implement probability conversion, similar to the maximum entropy method, to convert the probabilities in $m\_wa$ to Pignistic probabilities.
    Pignistic Probability Transformation
        Use the Transferable Belief Model (TBM) with a two-layer structure: the Belief layer and the Pignistic layer.
        Quantify the confidence in the Belief layer and transform it to Pignistic probability using the Pignistic probability transformation.
    Transform $m\_wa$ to $m\_da$
        After Pignistic probability transformation, convert the fusion result $m\_wa$ to $m\_da$ to avoid counterintuitive combinations due to high conflict.
    Return the Final Fusion Result
        Return the de-conflicted and converged fusion result $m\_da$ for subsequent decision-making.
End.
```

After the above processing, the fuzzy data in the dataset has been transformed into precise data. Next, an additional fully connected layer is used for information fusion to generate the final prediction result. When an observation sample c_i of an entity e_i has the same features and labels as another observation c_j, it is called a consistent sample. Consistent samples can increase the model's statistical belief and reduce the model's belief entropy. When an observation sample c_i of an entity e_i has different features but the same label as another observation c_j, it is called a contradictory sample. Contradictory samples can reduce the model's statistical belief and increase the model's belief entropy. Contradictory samples may come from noisy data and partially observable data. To address the uncertainty of contradictory samples, a direct method can be adopted to extend the observation dimension, including adding more observable modal channels to reduce model belief entropy and minimum model accuracy entropy.

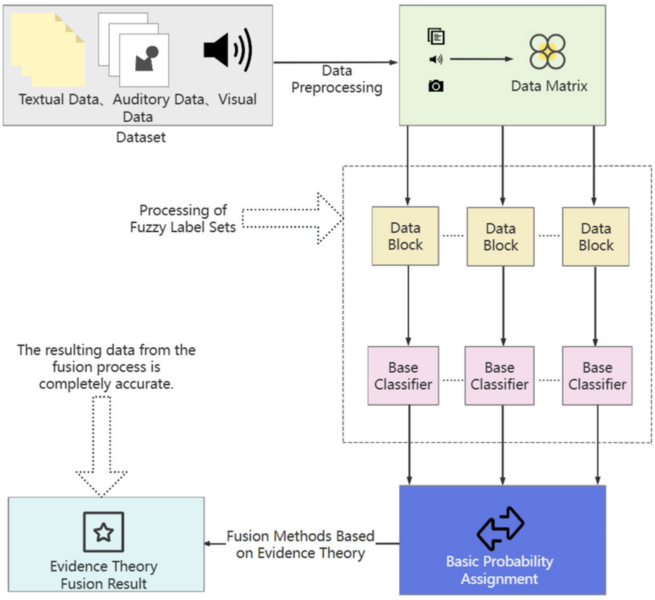

Figure 2. Fuzzy Data Processing Process

Algorithm 5. Precise Processing of Fuzzy Data

```
Input: Vague data groups V.
Output: Precise labels for each data group.
Begin:
    For each vague data group v in V:
        a. Train a base classifier on v.
        b. Obtain the probability distribution output as evidence.
    Apply the evidence theory to combine evidence from all classifiers.
    Use the combined evidence to assign a precise label to each data group.
    Return the precise labels.
End.
```

Precise Data Processing: In the precise data processing stage, we start from the perspective of information theory to optimize the data fusion process. The following is a detailed processing process:

Feature extraction and preprocessing: Necessary preprocessing is carried out on precise data to eliminate noise and irrelevant information, and then features that help decision-making are extracted.

Calculation of information entropy and mutual information: Information entropy is calculated for each data feature to assess its uncertainty. $H(x) = -\sum(i = 1)^\wedge n + p(x_i)logp(x_i)$, where $p(x_i)$ represents the probability of event x_i occurring. At the same time, mutual information between different data features is calculated to determine the degree of interdependence between them. Information entropy represents the uncertainty of information. By calculating the difference between conditional entropy and information entropy without data, the amount of information provided by a single modality or modality combination can be quantified. This helps to assess the contribution of different modalities to the model's predictive performance.

Weight allocation and optimization: Based on the results of information entropy and mutual information, initial weights are assigned to each data feature. These weights reflect the importance of each feature in the fusion process. Then, the weights are further adjusted through optimization algorithms to minimize the difference between the fusion result and the true situation.

Application of BJS divergence: In the data fusion process, we use BJS divergence to measure the uncertainty and information loss of the data. BJS divergence is a symmetric divergence, $BJS(P||Q) = BJS(Q||P)[\,BJS(P||Q) = frac\{1\}\{2\}(H(P,Q) + H(Q,P))\,]$,where $H(P,Q)$ represents the joint entropy. It combines the advantages of Jensen-Shannon divergence and Bhattacharyya distance, enhancing the sensitivity to uncertainty in the data fusion process.

3.3. Precise Data Processing Process

Through the above process, we will achieve the transformation from fuzzy data to precise data, and further improve the performance and decision accuracy of data fusion through the data fusion method from the perspective of information theory.

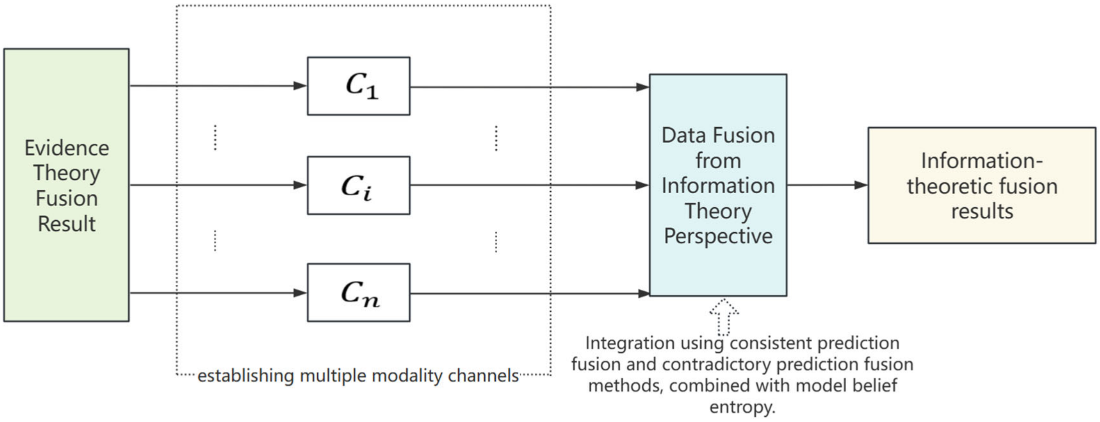

Figure 3. Precise Data Processing Process

Algorithm 6. Data Fusion Strategy

```
Input: Precise data $P$, precise labels from Algorithm 5.
Output: Final fused data set $F$.
Begin:
    Initialize an empty data set $F$.
    For each precise data point p in $P$:
        a. Encode p into a latent representation.
        b. Add the encoded representation to $F$.
    For each precise label l in $L$:
        a. Map l to the corresponding data points in $F$ using the BJS distance.
    Fuse the data in F using a deep learning model, considering the BJS distance.
    Return the fused data set F.
End.
```

4. Experimental Purpose and Dataset

The purpose of this experiment is to verify the performance of the proposed data fusion framework in processing fuzzy data and precise data, and to provide a basis for the discovery of important data. To this end, we used a medical system dataset, which includes fuzzy information and precise information.

This paper conducts experiments on the following medical datasets. Since each sample in the dataset is accompanied by a precise label, it is necessary to randomly extract a portion of the samples in the dataset and construct fuzzy labels.

Medical dataset: The medical dataset consists of 5 categories, denoted as A, B, C, D, and E, with each category composed of a Gaussian cluster. An example is used to plot 5 independent features for the cluster in a 5-dimensional space.

Then, interdependencies between these features are introduced. The covariance is multiplied by a random matrix and added, where the random numbers are uniformly distributed between -1 and 1. Then, further noise is added to the features. This is done by adding random noise to the features. These 5 strata jointly constitute the recognition framework. A total of 3000 data points were generated through the above settings, including 2400 training points and 600 test points. To simulate real-world scenarios, it is necessary to generate fuzzy labels. This study constructed these fuzzy labels to replace the original precise labels. Specifically, one-third of the training samples with labels A and B were replaced with{A, B}; one-third of the training instances with labels B and C were replaced with {B, C} and {D, E}.

This paper uses accuracy, $F1$-score, ROC-AUC, and PR-AUC as performance evaluation metrics. These metrics can comprehensively evaluate the performance of the classifier in processing fuzzy data.

Table 1. Performance Metrics

	PRECISION	F1-SCORE	ROC-AUC↑	PR-AUC↑
V-IGNORING	62.23%	53.35%	86.18%	68.23%
V-MAPPING	53.78%	47.45%	84.52%	58.41%
OUR METHOD	**72.47%**	**67.46%**	**92.45%**	**78.43%**

The results show that when processing fuzzy labels, our fusion method has improved the F1 value by at least 19.87% compared to other comparison methods, and the accuracy has been significantly improved. It was found that setting k=3 can achieve better results. The reason may be that the evidence fusion of three output BPAs better utilizes the uncertain information in fuzzy labels. When the number of BPAs is increased, the performance of the method tends to stabilize. The reason may be that as the number of BPAs increases, the uncertainty information in the fuzzy labels is better utilized, and the performance of the method tends to stabilize. The V-Ignoring method performed poorly because this direct discard method ignored the valuable uncertain information between fuzzy labels. After discarding the fuzzy labels, there may even be cases where only a few samples of certain categories exist, leading to the classifier losing the ability to recognize some specific categories. The reason for the poor performance of V-Mapping may be that unnecessary noise was introduced when the uncertain information was transformed into a certain form. When the fuzzy labels are disambiguated, a large number of false positive labels are introduced at the same time, resulting in a decrease in the reliability of the converted labels. Therefore, this comparative method also performed poorly.

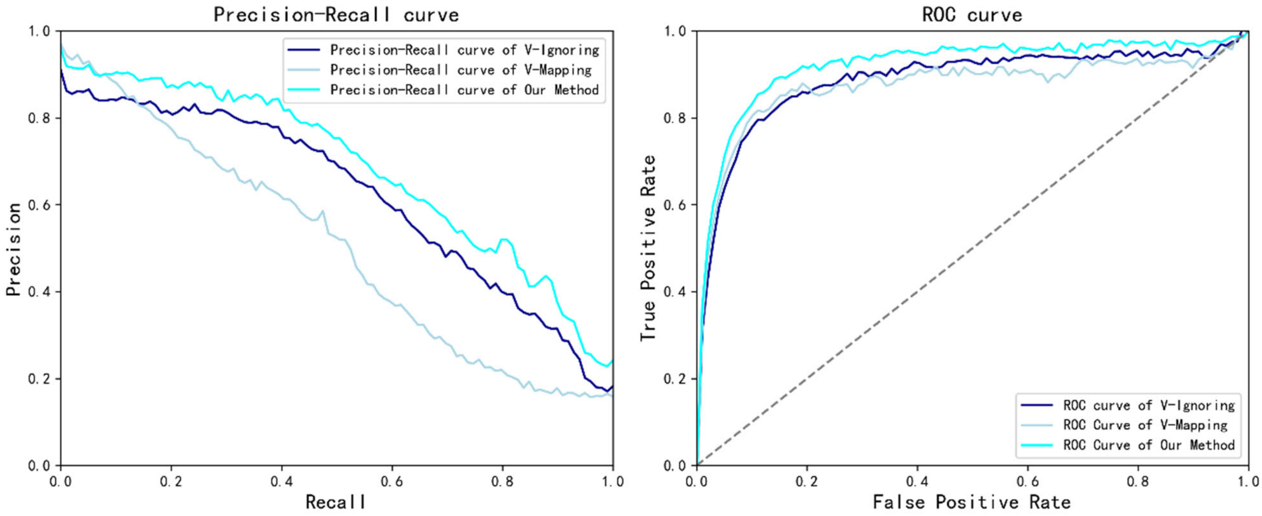

Figure 4. Comparison Curve of Performance

The results show that our proposed fusion method has good fusion performance on the synthetic medical dataset with fuzzy labels. Compared with the fusion of precise data, although the data used has fuzzy labels, the performance level of the proposed method is relatively close. From past experience, the fusion performance of precisely labeled data is the target (upper limit) for methods dealing with fuzzy labels. In other words, researchers expect the evidence fusion method proposed in this paper to perform closer to the results of precise label fusion when dealing with data with fuzzy labels. Compared with other comparative methods, our proposed fusion method has significantly improved performance when dealing with data with fuzzy labels. On the synthetic medical dataset, the gap between F1-score and the upper limit has been reduced from 28.60% and 35.16% to 12.59% and 14.78%. Similarly, for the other two indicators ROC-AUC and PR-AUC, the gap with the upper limit has also been greatly reduced.

In addition, this study deployed a DNN model to fit the fusion results to verify the effect of data fusion. The DNN, through its deep structure, can automatically extract useful features from the data and has strong nonlinear processing capabilities. It can be seen from the figure that the decision boundary of this fusion model well separates the samples of the original dataset, indicating that the fusion model is effective.

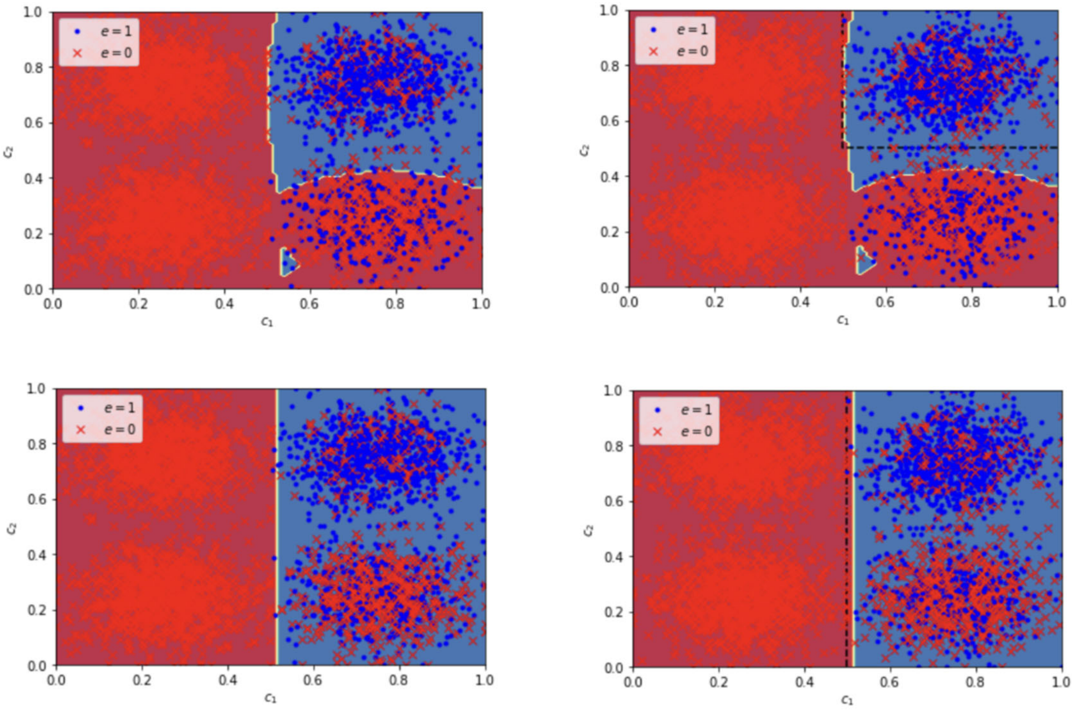

Figure 5. DNN Model Fitting of Fusion Results

5. Conclusion

This paper proposes a new method for classifying vaguely labeled data based on evidence fusion. The problem of classifying data with vague labels is formally defined. A new effective label set coverage allocation algorithm is proposed to group the vaguely labeled data into several small data groups, thereby reducing the number of combinations. Each small data group with a base classifier is regarded as a piece of evidence, and evidence theory is applied to the classification of vaguely labeled data. By converting the evidence theory and the transformation of the credibility function, the gap between vague labels and precise classification results is bridged. This method is universal because it can be equipped with any classifier that can give a probability distribution result. Experiments show that this method can effectively combine the complementary information of vaguely labeled data.

After processing the fuzzy data, the multimodal data fusion process is analyzed from the perspective of information theory. Essentially, the process of multimodal data fusion is to fuse all the information provided by each modal channel to reduce the model's uncertainty and improve the model's predictive accuracy. This paper defines the basic concepts of multimodal data fusion analysis, including how to measure the maximum model accuracy in multimodal data fusion tasks, how to measure model uncertainty, and how to distinguish different situations of multimodal data fusion. It is ultimately proven that if the modal data contains unique information that improves model accuracy, then the predictive model with additional modal channels cannot reduce the upper limit of model accuracy, and the modal channels are indispensable to achieve the upper limit of model accuracy. From an information perspective, this analysis can be applied to all possible cases of multimodal data fusion that may contain incomplete data, inconsistent data, and conflicting data. In future work, the issue of the exponential growth of fuzzy label indices can be further studied. To provide a data basis for the discovery of massive important data in the future.

Acknowledgements

This work is supported by the National Key Research and Development Plan of China, Key Project of Cyberspace Security Governance (No.2022YFB3103103), the Key Research and Development Project of Chengdu(No. 2023-XT00-00002-GX), the Key Research and Development Support Program Project of Chengdu(No.2024-YF05-01227-SN).

Declarations

Conflict of Interest None of the authors present any conflicts of interest whatsoever. The material presented in this article is not subjected to any copyright of any class or nature.

References

[1] Wei, Y. , Cheng, Z. , Zhu, B. , Zhai, X. , & Zhang, H. . (2019). Infrared and radar fusion detection method based on heterogeneous data preprocessing. Optical and Quantum Electronics, 51(10). https://doi.org /10.1007/s11082-019-2053-z

[2] Zhang, P. , Liu, W. , Lei, Y. , & Lu, H. . (2019). Hyperfusion-net: hyper-densely reflective feature fusion for salient object detection. Pattern Recognition, 93(4). https://doi.org/ 10.1016/j.patcog.2019.05.012

[3] A, F. J. , A, X. W. , B, J. G. , A, Q. L. , C, J. Z. A. , & A, H. L. , et al. (2020). Bi-connect net for salient object detection. Neurocomputing, 384, 142-155. https://doi.org/ 10.1016/j.neucom.2019.12.020

[4] Khaleghi, B. , Khamis, A. , Karray, F. O. , & Razavi, S. N. . (2013). Multisensor data fusion: a review of the state-of-the-art. Information Fusion, 14(1), 28-44. https://doi.org/ 10.1016/j.inffus.2011.08.001

[5] Zheng, H. , Cai, A. , Zhou, Q. , Xu, P. , Zhao, L. , & Li, C. , et al. (2017). Optimal preprocessing of serum and urine metabolomic data fusion for staging prostate cancer through

design of experiment. Analytica Chimica Acta, 68-75. https://doi.org/10.1016/j.aca.2017.09.019

[6] Dolly, D. R. J. , Peter, J. D. , Bala, G. J. , & Jagannath, D. J. (2020). Image fusion for stabilized medical video sequence using multimodal parametric registration. Pattern Recognition Letters, 135. https://doi.org/10.1016/j.patrec.2020.04.001

[7] Zhang, H. , Ananda, R. , Fu, X. , Sun, Z. , Wang, X. , & Chen, K. , et al. (2023). Multi-channel sensor network construction, data fusion and challenges for smart home.

[8] Maimaitijiang, M. , Sagan, V. , Sidike, P. , Hartling, S. , & Fritschi, F. B. . (2019). Soybean yield prediction from uav using multimodal data fusion and deep learning. Remote Sensing of Environment, 237. https://doi.org/10.1016/j.rse.2019.111599

[9] Liu, X. , Liu, Q. , & Wang, Y. . (2019). Remote sensing image fusion based on two-stream fusion network. Information Fusion, 55. https://doi.org/10.1016/j.inffus.2019.07.010

[10] Rajah, P. , Odindi, J. , & Mutanga, O. . (2018). Feature level image fusion of optical imagery and synthetic aperture radar (sar) for invasive alien plant species detection and mapping. Remote Sensing Applications Society and Environment, 10, 198–208.

[11] Huang, M. , Liu, Z. , & Tao, Y. . (2019). Mechanical fault diagnosis and prediction in iot based on multi-source sensing data fusion. Simulation Modelling Practice and Theory, 102：101981. https://doi.org/10.1016/j.simpat.2019.101981

[12] Gong, X. , Ding, L. , Du, W. , & Wang, H. . (2017). Gear fault diagnosis using dual channel data fusion and eemd method. Procedia Engineering, 174(Complete), 918-926. https://doi.org/10.1016/j.proeng.2017.01.242

[13] Safizadeh, M. S. , & Latifi, S. K. . (2014). Using multi-sensor data fusion for vibration fault diagnosis of rolling element bearings by accelerometer and load cell. Information Fusion, 18, 1-8. https://doi.org/10.1016/j.inffus.2013.10.002

[14] Yan, J. , Hu, Y. , & Guo, C. . (2019). Rotor unbalance fault diagnosis using dbn based on multi-source heterogeneous information fusion. Procedia Manufacturing, 35, 1184-1189. https://doi.org/10.1016/j.promfg.2019.06.075

[15] Vita, F. D. , Bruneo, D. , & Das, S. K. . (2020). On the use of a full stack hardware/software infrastructure for sensor data fusion and fault prediction in industry 4.0. Pattern Recognition Letters,138 : 30-37. https://doi.org/10.1016/j.patrec.2020.06.028

[16] Rato, T. J. ,& Reis, M. S. . (2019). Optimal fusion of industrial data streams with different granularities. Computers & Chemical Engineering, 130(Nov.2), 106564.1-106564.16. https://doi.org/10.1016/j.compchemeng.2019.106564

[17] Alofi, A. , Alghamdi, A. , Alahmadi, R. , Aljuaid, N. , & Hemalatha, M. . (2017). A review of data fusion techniques. International Journal of Computer Applications, 167(7), 37-41.

[18] Mcgurk, H. , & Macdonald, J. . (1976). Hearing lips and seeing voices. Nature, 264(5588), 746-748. https://doi.org/10.1038/264746a0

[19] Dai, Y. , Yan, Z. , Cheng, J. , Duan, X. , & Wang, G. . (2023). Analysis of multimodal data fusion from an information theory perspective. Information Sciences, 623, 164-183. https://doi.org/10.1016/j.ins.2022.12.014

[20] Song, M. , Sun, C. , Cai, D. , Hong, S. , & Li, H. . (2022). Classifying vaguely labeled data based on evidential fusion. Information Sciences: An International Journal(583-), 583, 159–173. https://doi.org/ 10.1016/j.ins.2021.11.005