Proceedings of the 42nd Chinese Control Conference
July 24-26, 2023, Tianjin, China

Hierarchical Belief K-Nearest Neighbors for Human Activity Recognition

Yilin Dong and Yong Zhou

***Abstract*—In Wearable Body Sensor Networks (WBSNs), the multi-sensor fusion strategy is widely utilized in Human Activity Recognition (HAR) problems. As the classical decision-level fusion strategy, the Belief Functions (BF) theory often uses a flat structure to fuse the outputs of basic classifiers. However, it is very challenge and difficult to accurately identify confusing activities such as walking upstairs and downstairs based on such flat structure. In this paper, a novel Hierarchical Belief K-Nearest Neighbors approach (Hie-BKNN) for activity recognition is proposed. First, the data-driven hierarchical tree is constructed by using the spectral clustering strategy. Then, inspired by BF theory and classical K-nearest neighbors, the pair-aware belief masses are calculated for each node and the belief hierarchical tree can be generated. Afterwards, we adopt a novel hierarchical fusion strategy to fuse all involved belief hierarchical trees. Finally, by selecting the probability maximization path on the entire tree, the label of a test sample can be effectively predicted. The experimental results show that our proposed Hie-BKNN approach can precisely recognize daily activities of two different UCI public data sets, i.e., Smartphone and WISDM datasets.**

***Index Terms*—Multi-Sensor Fusion, Activity Recognition, Hierarchical Classification, Belief Functions Theory, Wearable Body Sensor Networks.**

I. Introduction

A. Background and Research Motivation

In recent years, with the development of Internet of Things (IoT) technology, Human Activity Recognition (HAR) in Body Sensor Networks (BSNs) is widely used in monitoring environmental assisted living [1], sports injury detection [2], health management [3], medical diagnosis [4], and elderly care [5]. Generally, traditional activity recognition models [6] usually rely on machine learning models to identify activities by using the signals sensed from a single wearable sensor. However, many recent works [7–10] have been already proved that HAR model coupled with multi-sensor data fusion techniques is an effective solution for high-precision activity recognition from imperfect sensor samples such as noise signals, data loss or inconsistency. As the classical decision-level fusion strategy, Belief Functions (BF) theory [11], also called Dempster-Shafer evidence theory (DST), is an important reasoning framework for uncertainty modeling, which has obvious mathematical advantages in fusion of imperfect information [12]. Thus, BF theory has been widely used in HAR fields because of its ability in describing and eliminating the uncertainty in the activity recognition problems. For example, in our previous works [7], the selected kernel density estimation models for each specific activity were first built and combined with Dezert-Smarandache theory (DSmT) in the decision-level fusion stage to predict the label of unknown activities. Following this line, we further extended the type of classical belief structure in the theory of BF and proposed the hesitant fuzzy belief structure to deal with HAR problems in [13]. Besides, Noor et.al [14] integrated the ontological reasoning mechanism with DST to provide support for handling uncertainty in the activity recognition. Although the performance of the BF-based activity recognition models has been improved, it is often difficult to recognize those fine-grained and confusing activities, such as going up and down stairs, opening and closing doors, and so on. This is mainly because that the classical flat structure is often applied in the fusion stages of BF theory which could easily neglect the inter-class taxonomic relationships. Fortunately, there often exists the potential hierarchy structure in daily activities, which helps to efficiently organize raw data with various categories. Because of such hierarchical structure between activities, the more difficult and confusing activity classification tasks can be divided into several simpler subtasks, so they can be solved more effectively. In the literature [15] and [16], authors fully compared and discussed the advantages and disadvantages of these two strategies: hierarchical classification and flat classification, and then gave their conclusion that hierarchical classification is more robust in classification tasks than flat classification.

Yilin Dong is with College of Information Engineering, Shanghai Maritime University. Emails:yldong@shmtu.edu.cn. Yilin Dong is the corresponding author.

Yong Zhou is with College of Information Engineering, Shanghai Maritime University. Emails:2440459694@qq.com.

B. Challenges

Although under the framework of BF theory, the activity recognition model based on the hierarchical structure has a predictable classification effect, it also faces many challenges:

- **Construction of hierarchical structure**: for the BF-based hierarchical activity recognition, the key problem is to construct a good hierarchical label structure. However, in real scenarios, the prior knowledge of how to organize data hierarchically is usually limited. And we usually only know some rough relationships between categories which are not enough for constructing a complete hierarchical structure;
- **Calculation of belief alignment**: In the theory of BF, the belief assignments need to be calculated before the process of decision-level fusion. Thus, how to construct the belief masses of all involved categories reasonably and accurately in the hierarchical structure is particularly

important. However, such problem is still one of the difficulties in the application of BF theory [17]; in addition, the traditional combination rules in BF theory are only applicable to combine Sources of Evidence (SoEs) with the flat structure, it is very necessary to propose new fusion rules for fusing SoEs with the hierarchical structure;
- **Implementation of hierarchical reasoning**: performing hierarchical reasoning based on a hierarchical structure is also a challenge [18]. That is to say, in the process of hierarchical reasoning, it is very important to study how to prevent the errors propagation from the root node to the leaf node.

C. Main Contributions

In this paper, we propose a novel Hierarchical Belief K-Nearest Neighbors classifier (Hie-BKNN) for human activity recognition. The main contributions of this work are summarized as follows:

- We propose a data-driven hierarchical structure learning method. First, Dynamic Time Warping (DTW) is used as the similarity metric to compute the similarity degree between activities. Then, the spectral clustering is recursively applied to construct the hierarchical tree;
- Through the nearest neighbors search, the belief mass of the unlabeled sample for each node is calculated to generate the belief hierarchical tree. Then, a new hierarchical fusion rule is also proposed to combine all involved belief hierarchical trees;
- Finally, according to the interpretable maximum probability decision rule, we can classify unlabeled samples following the top-down optimal path in the fused belief hierarchical tree. Comprehensive experiments on two public activity recognition datasets demonstrate the superiority of the proposed Hie-BKNN.

The rest of this article is organized as follows. Section II reviews the basics of BF theory. Section III presents the proposed Hie-BKNN. Section IV reports the experimental results. Finally, Section V concludes this article.

II. Basics of Belief Function theory

In DST framework, the Frame of Discernment (FoD)[1] $\Theta \triangleq \{\theta_1, \ldots, \theta_n\}$ $(n \geq 2)$ is a set of exhaustive and exclusive elements (hypotheses) which represent the possible solutions of the problem under consideration. Shafer's model assumes $\theta \cap \theta' = \emptyset$ for $\theta \neq \theta'$ in Θ. A BBA $m(\cdot)$ is defined by the mapping: $2^\Theta \mapsto [0,1]$, verifying $m(\emptyset) = 0$ and $\sum_{\theta \in 2^\Theta} m(\theta) = 1$. The belief (or credibility) and plausibility functions are respectively defined by $Bel(\theta) \triangleq \sum_{\theta' \in 2^\Theta | \theta' \subseteq \theta} m(\theta')$ and $Pl(\theta) \triangleq \sum_{\theta' \in 2^\Theta | \theta' \cap \theta \neq \emptyset} m(\theta')$. $BI(\theta) \triangleq [Bel(\theta), Pl(\theta)]$ is called the belief interval of θ.

In order to combine two distinct SoEs, the classical DS rule in [11]) was proposed and defined by $m_{DS}(\emptyset) = 0$ and $\forall \theta \in 2^\Theta \setminus \{\emptyset\}$:

$$m_{DS}(\theta) = \frac{\sum_{\theta', \theta'' \in 2^\Theta | \theta' \cap \theta'' = \theta} m_1(\theta') m_2(\theta'')}{1 - \sum_{\theta', \theta'' \in 2^\Theta | \theta' \cap \theta'' = \emptyset} m_1(\theta') m_2(\theta'')}. \quad (1)$$

[1]Here, we use the symbol $\triangleq$ to mean *equals by definition*.

III. Hierarchical Belief K-Nearest Neighbors for HAR

In this section, we first introduce the principle of hierarchical tree construction, then describe the specific steps of Hie-BKNN classifier in details.

A. Hierarchical Tree Construction

As shown in previous works [18], it is reasonable to classify highly similar sequential classes into a super class. Generally, two main necessary steps are included: 1) affinity matrix is computed which aims to measure the inter-class similarity, and 2) the spectral clustering is utilized for hierarchical tree construction.

Algorithm 1: DTW algorithm

Input: $X = \{x_1, x_2, \cdots, x_{V_1}\}$ and $X' = \{x'_1, x'_2, \cdots, x'_{V_2}\}$: two time series.
Output: The optimal warping path; The distance of X and X'.

```
Initialize: DTW(1,1) = d_{1,1} = sqrt((x_1 - x'_1)^2);
for i = 2, ..., V_1 do
    for j = 2, ..., V_2 do
        d_{i,j} = d(x_i, x'_j) = sqrt((x_i - x'_j)^2);
        D_1 = d_{i,j} + DTW(i-1, j);
        D_2 = d_{i,j} + DTW(i-1, j-1);
        D_3 = d_{i,j} + DTW(i, j-1);
        DTW = min(D_1, D_2, D_3);
        The optimal
          path_{i,j}=min_index((i-1, j), (i-1, j-1), (i, j-1)).
    end
end
```

1) Calculating Affinity Matrix: In order to obtain an affinity matrix, we first need to select a robust time-series distance metric. Until now, it is still an unsolved issue to choose the robust distance metric to measure the similarity degree between daily activity-based time series data [19]. In general, the Euclidean distance [20] is widely used to calculate the sum of the distances between corresponding points among the time series. However, Euclidean distance is difficult to accurately measure distorted and shifting time-series points. In this article, we utilize DTW [21] to effectively measure the difference between time series. And the pseudo-code of the DTW is presented in Algorithm 1. Then, we can get the element of affinity matrix, which is calculated as follows:

$$a_{i'j'} = \exp\left(-\frac{DTW\left(X_{i'}, X'_{j'}\right)}{\delta}\right). \quad (2)$$

where δ is the pre-defined parameter and $i', j' \in [1, c]$. By using DTW, we can construct the affinity matrix, which is mathematically denoted as

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1c} \\ \vdots & \ddots & \vdots \\ a_{c1} & \cdots & a_{cc} \end{bmatrix} \quad (3)$$

Then the spectral clustering algorithm [22, 23] is utilized to generate the hierarchical tree.

2) Hierarchical Tree Construction: Here, the construction of the hierarchical tree adopts a top-down construction strategy. In the process of the actual construction, the tree nodes contained in each layer of the hierarchical tree will be recursively divided. First, considering all the activities categories contained in the root node, the spectral clustering is then performed on the entire affinity matrix for the root node. In this paper, we just focus our discussions on hierarchical binary trees. Thus, all non-leaf nodes in the hierarchical tree are clustered into two groups (left child node and right child node). At the same time, considering that each child node may contains some categories (the number of categories is greater than 1), then the spectral clustering will be also conducted on the affinity sub-matrix corresponding to the categories contained in the child node. By analogy, all nodes are clustered through recursive spectral clustering until the following rule is met: the current node is a leaf node, that is, this node contains only one category.

Note that the set of classes contained in the node ν is denoted by $C(\nu)$. The ith children node of the node ν contains the class set C_ν^i. For each node ν, the union of the class sets contained in the node ν, that is, $\bigcup_{i \in C(\nu)} C_\nu^i = C(\nu)$. Moreover, any pairwise sibling nodes do not overlap, which satisfies $C_\nu^i \cap C_\nu^j = \emptyset, i \neq j$. Thanks to the construction of such hierarchical tree, the elements and their combinations in the target class set can effectively correspond to the nodes in the derived tree structure: the root node represents the completely unknown class (in BF theory, this concepts is mathematically denoted by Θ); those middle nodes corresponds to the coarse-grained category which represents the union of several different classes (these nodes can be denoted by disjunctive focal elements in BF theory) and the leaf nodes can be regarded as the most fine-grained classes (singletons in BF theory).

B. Hierarchical Belief K-Nearest Neighbors Classifier

In this part, we describe the main idea of our proposed HieBKNN method based on the derived hierarchical tree-based structure. There are two points need to mention:

- Firstly, we do not use the classical Euclidean distance to measure the distances between time series. Instead, the nearest neighbors are selected based on the distances calculated by DTW;
- Secondly, once all nearest neighbors of the testing sample for each node in hierarchical tree are chosen, these selected samples will be used to calculate the belief masses and then determine the label of the unknown testing sample after hierarchical fusion;
- Finally, the best decision-making path in hierarchical tree is determined by the maximum value of belief masses.

1) Calculating the local belief masses by using the nearest neighbors: Let $\{(X_i, Y_i)|i = 1, \cdots, N\}$ be a collection of N training pairs, in which $X_i = [x_1, \cdots, x_V]^T$ is the ith training sample with V features and $Y_i \in \{y_1, \cdots, y_c\}$ is the corresponding class label. Considering that in the derived hierarchical tree, all involved training samples have already clustered into Q groups, the k nearest neighbors of the testing sample should also be selected in these Q groups, respectively. Given a query instance X^t, its class membership can be determined through the following steps:

- For each X^t, finds k nearest neighbors of X^t from $\{(X_{i'}^\nu, Y_{i'}^\nu)|i' = 1, \cdots, N'\}$ in ν node using DTW and then puts them into $KNN\_X_\nu^t$ set. And the set of classes contained in this node ν is denoted by $C(\nu) = Y_i^\nu$, which is the subset of the original class label $\{y_1, y_2, \cdots, y_c\}$;
- Each nearest neighbor of X^t in $KNN\_X_\nu^t$ is considered as an item of evidence that supports certain hypotheses regarding the class membership of X^t. Let X_k^ν be one of its k nearest neighbors with class label $Y_k^\nu = y_q \in Y_i^\nu$. The mass function induced by X_k^ν, which supports the assertion that X^t also belongs to y_q is

$$m_{t,k}^\nu(\{y_q\}) = \alpha exp(-\gamma \cdot d_{t,k}^2). \tag{4}$$

  where $d_{t,k}$ is the distance between X_k^ν and X^t calculated by DTW, while α and γ are pre-defined parameters and in this paper, we just set the values of α and γ equal to 1 and 0.05.
- Then, we can obtain the belief tree-based structure for the testing sample X^t based on its corresponding nearest neighbors. The principle of construction is quite simple: through the nearest neighbor samples in each node ν of hierarchical tree, the belief assignment is calculated based on (4) and then this value is regarded as the weight linking the corresponding node in the tree structure.

Based on the above steps, we can obtain the k belief hierarchical trees, which is determined by the k selected nearest neighbors. In order to improve the final classification accuracy, we need to fuse the all belief hierarchical trees. However, the traditional fusion rules in BF theory can not be directly used for fusing SoEs under such hierarchical structure. Therefore, we need to propose the new hierarchical fusion rule.

2) Combining all neighbors' knowledge with hierarchical fusion strategy: As we discussed before, we model the multi-granularity predefined activities according to the concept of FEs in BFs theory. That is to say, for the leaf nodes in tree, the most fine-grained activities are modeled by singletons and the related internal nodes are described by disjunctive FEs. Through the belief KNN in each layer of tree-based hierarchical structure, the connection weights of each node can be calculated and these connection weights are regarded as the belief masses of FEs in tree structure.

In order to reduce the risk of error propagation caused by nearest neighbor noise, we propose a novel hierarchical fusion strategy to fuse tree-based belief mass function based on the selected nearest neighbors. Specifically, the two-dimensional BBAs in the same tree-layer are first fused with Dempster's rule (1) and then the derived fusion BBAs are regarded as the connected weights of the corresponding nodes. Then, the final belief mass of each leaf node is successively multiplied by the connection weights from the root node to the leaf node. Finally, we can predict the type of the unknown activity based on the principle of probability maximization.

The principle of our proposed tree-based combination rule for nearest neighbors is illustrated in Fig.1. As we can observe in Fig.1, k belief based hierarhical trees are derived from

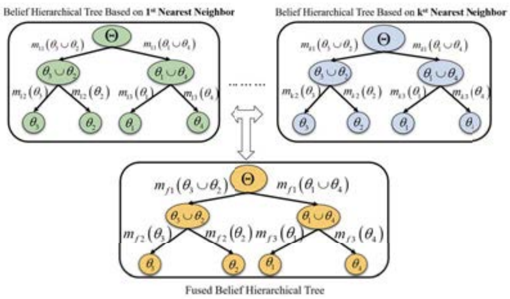

Fig. 1: **Belief Hierarchical Tree Combination Rule.**

nearest neighbors trained by belief KNN and our task here is to combine all these involved belief hierarchical trees. In order to briefly show the principle of our fusion strategy for combining belief hierarchical trees, we set the value of k equals to 2 and three BBAs from the belief hierarchical tree based on 1st nearest neighbor are given for instance:

$$\mathbf{m}_{11} = \{m_{11}(\theta_3 \cup \theta_2), m_{11}(\theta_1 \cup \theta_4)\}; \tag{5}$$
$$\mathbf{m}_{12} = \{m_{12}(\theta_3), m_{12}(\theta_2)\}; \tag{6}$$
$$\mathbf{m}_{13} = \{m_{13}(\theta_1), m_{13}(\theta_4)\}. \tag{7}$$

Similarly, for the belief hierarchical tree based on 2st nearest neighbor, three involved BBAs are also included:

$$\mathbf{m}_{21} = \{m_{21}(\theta_3 \cup \theta_2), m_{21}(\theta_1 \cup \theta_4)\}; \tag{8}$$
$$\mathbf{m}_{22} = \{m_{22}(\theta_3), m_{22}(\theta_2)\}; \tag{9}$$
$$\mathbf{m}_{23} = \{m_{23}(\theta_1), m_{23}(\theta_4)\}. \tag{10}$$

Before the fusion step, we need to ensure that each piece of BBAs meets the normalization condition. Then, in our fused belief hierarchical tree, the corresponding BBAs in the same layer are combined by the classical $DS(\cdot)$ (1):

$$\mathbf{m}_{f1} = DS(\mathbf{m}_{11}, \mathbf{m}_{21}) = \{m_{f1}(\theta_3 \cup \theta_2), m_{21}(\theta_1 \cup \theta_4)\}; \tag{11}$$
$$\mathbf{m}_{f2} = DS(\mathbf{m}_{12}, \mathbf{m}_{22}) = \{m_{f2}(\theta_3), m_{f2}(\theta_2)\}; \tag{12}$$
$$\mathbf{m}_{f3} = DS(\mathbf{m}_{13}, \mathbf{m}_{23}) = \{m_{f3}(\theta_1), m_{f3}(\theta_4)\}. \tag{13}$$

Once the learned hierarchical tree is obtained, the label of the test sample needs to be predicted. And the principle of decision making here is very simple and the selection of the best path (from the root node to the leaf node) is completely based on the maximum value of the corresponding belief mass.

IV. Experiments and Discussions

In this part, two famous UCI datasets: Smartphone[2][24] and WISDM dataset[3][25] are applied to verify the proposed Hie-BKNN model. More descriptions of these two involved datasets Smartphone and WISDM can be found in [24] and [25].

1) Results on UCI Smartphone Dataset:

a) Hierarchical tree construction: According to the algorithm described in Section III-A, the hierarchical tree of six daily activities involved in UCI Smartphone are constructed by using DTW distance and spectral clustering strategy. The specific steps are as follows: first, the original Smartphone dataset is divided into training (70%) and testing (30%) sets. And then, we further divide the obtained training sets into six sub-training sets based on the six daily activities (Sitting, Standing, Laying, Walking, $Walking_{up}$, $Walking_{dn}$ which are denoted by $\theta_1, \theta_2, \theta_3, \theta_4$, θ_5 and θ_6, respectively). Afterwards, the classical DTW measure is applied to calculate the similarity degree between daily activities and the difference matrix is obtained in Table I. It can be easily observed that there exists two potential clusters: the static activities (Sitting, Standing and Laying-red mark in Table I) and the dynamic activities (Walking, $Walking_{up}$, $Walking_{dn}$-green mark in Table I). It can be found that the differences between static and dynamic classes are large, but the differences within classes (static or dynamic) are relatively small. After obtaining the difference matrix, we use (2) to convert the difference matrix into the affinity matrix, which is given in Table II and the parameter δ is set to 1000. As can be seen in Table II, the smaller the difference degree between activities, the higher the value in affinity matrix. Based on the derived affinity matrix, the spectral clustering strategy is used to cluster six groups of activities iteratively. Here, in our following experiments, we directly apply the built-in spectral clustering function in MATLAB software to realize clustering (the parameter settings of this function are as follows):

$$Spectralcluster(AffinityMatrix, NumberofCluster, \\ 'Distance','precomputed','LaplacianNormalization', \\ 'symmetric').$$

Considering that the hierarchical tree structure obtained by spectral clustering is affected by the parameter $NumberofCluster$, and we need to make the final leaf node in the tree structure contain only one activity class. Thus, the value of $NumberofCluster$ is set to six so as to guarantee the number of classes in $C(leafnode)$ equals to one.

TABLE I: Difference Matrix Between Six Activities on UCI Smartphone With DTW.

	Sitting	Standing	Laying	Walking	$Walk_{up}$	$Walk_{dn}$
Sitting	0	757.2	758.5	1140.2	1065.9	1250.9
Standing	757.2	0	802.4	1134	1064.2	1240.8
Laying	758.5	802.4	0	1256.3	1187.8	1355.9
Walking	1140.2	1134	1256.3	0	639.5	820.5
$Walk_{up}$	1065.9	1064.2	1187.8	639.5	0	817.5
$Walk_{dn}$	1250.9	1240.8	1355.9	820.5	817.5	0

TABLE II: Affinity Matrix Between Six Activities on UCI Smartphone.

	Sitting	Standing	Laying	Walking	$Walk_{up}$	$Walk_{dn}$
Sitting	1	0.4690	0.4683	0.3197	0.3432	0.2863
Standing	0.4690	1	0.4482	0.3218	0.3450	0.2892
Laying	0.4683	0.4482	1	0.2847	0.3049	0.2577
Walking	0.3197	0.3218	0.2847	1	0.5275	0.4402
$Walk_{up}$	0.3432	0.3450	0.3049	0.5275	1	0.4415
$Walk_{dn}$	0.2863	0.2892	0.2577	0.4402	0.4415	1

[2]http://archive.ics.uci.edu/ml/datasets/Smartphone-Based+Recognition+of +Human+Activities+and+Postural+Transitions.

[3]http://archive.ics.uci.edu/ml/datasets/WISDM+Smartphone+and+Smartwatch +Activity+and+Biometrics+Dataset+.

b) Belief hierarchical tree construction: Based on the hierarchical tree structure derived from spectral clustering, we use belief k-NN described in III-B to construct the belief hierarchical tree: first, the value of the nearest neighbor parameter k is set to 7. It means that, for the testing sample, seven nearest training time series are chosen based on DTW measure to calculate the belief masses through (4). For example, in order to illustrate the principle of hierarchical fusion, two obtained belief hierarchical trees are shown in Fig.2.

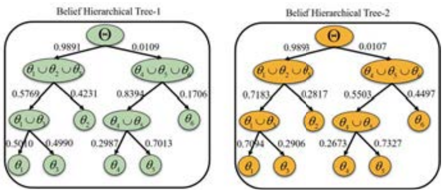

Fig. 2: **Two Belief Hierarchical Trees in UCI Smartphone Dataset.**

c) Hierarchical Fusion: Then, these two belief hierarchical tree can be combined by using our proposed hierarchical fusion rule. First, there exists five BBAs $(\mathbf{m}_{11}, \mathbf{m}_{12}, \mathbf{m}_{13}, \mathbf{m}_{14}, \mathbf{m}_{15})$ in belief hierarchical tree-1:

$$m_{11}(\theta_1 \cup \theta_2 \cup \theta_3) = 0.9891, m_{11}(\theta_4 \cup \theta_5 \cup \theta_6) = 0.0109; \tag{14}$$

$$m_{12}(\theta_1 \cup \theta_3) = 0.5769, m_{12}(\theta_2) = 0.4231; \tag{15}$$

$$m_{13}(\theta_4 \cup \theta_5) = 0.8394, m_{13}(\theta_6) = 0.1706; \tag{16}$$

$$m_{14}(\theta_1) = 0.5010, m_{14}(\theta_3) = 0.4990; \tag{17}$$

$$m_{15}(\theta_4) = 0.2987, m_{15}(\theta_5) = 0.7013. \tag{18}$$

Similarly, there exists another five BBAs $(\mathbf{m}_{21}, \mathbf{m}_{22}, \mathbf{m}_{23}, \mathbf{m}_{24}, \mathbf{m}_{25})$ in belief hierarchical tree-2:

$$m_{21}(\theta_1 \cup \theta_2 \cup \theta_3) = 0.9893, m_{21}(\theta_4 \cup \theta_5 \cup \theta_6) = 0.0107; \tag{19}$$

$$m_{22}(\theta_1 \cup \theta_3) = 0.7183, m_{22}(\theta_2) = 0.2817; \tag{20}$$

$$m_{23}(\theta_4 \cup \theta_5) = 0.5503, m_{23}(\theta_6) = 0.4497; \tag{21}$$

$$m_{24}(\theta_1) = 0.7094, m_{24}(\theta_3) = 0.2906; \tag{22}$$

$$m_{25}(\theta_4) = 0.2673, m_{25}(\theta_5) = 0.7327. \tag{23}$$

Then, we can combine these BBAs in the corresponding layer by using $DS(\cdot)$ and get the final fused five BBAs:

$$\mathbf{m}_{f1} = DS(\mathbf{m}_{11}, \mathbf{m}_{21}) = (0.9999, 0.0001); \tag{24}$$

$$\mathbf{m}_{f2} = DS(\mathbf{m}_{12}, \mathbf{m}_{22}) = (0.7766, 0.2234); \tag{25}$$

$$\mathbf{m}_{f3} = DS(\mathbf{m}_{13}, \mathbf{m}_{23}) = (0.8561, 0.1439); \tag{26}$$

$$\mathbf{m}_{f4} = DS(\mathbf{m}_{14}, \mathbf{m}_{24}) = (0.7102, 0.2898); \tag{27}$$

$$\mathbf{m}_{f5} = DS(\mathbf{m}_{15}, \mathbf{m}_{25}) = (0.1345, 0.8655). \tag{28}$$

d) Decision making: Finally, the final best decision-making path in hierarchical tree can be found based on the principle of maximum belief assignment which is denoted in Fig.3 (the red line is the final decision-making path). And in this case the label of the testing sample belongs to θ_1.

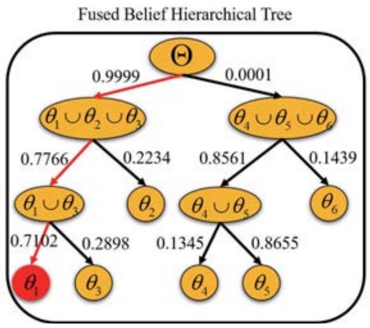

Fig. 3: **The decision-making path in Belief Hierarchical Trees.**

e) Compared to several classical algorithms: As discussed earlier, based on the UCI smartphone data set in the experiment, we compared the proposed method with the classic algorithms. The comparison results are shown in Table III. In addition, the confusion matrix of Hie-BKNN is further given (Fig.4). Benefiting from the hierarchical tree structure, the performance of our proposed Hie-BKNN is better than all the algorithms mentioned, and Hie-BKNN achieved the best performance (90.30%) in all the recognition tasks of activity recognition.

TABLE III: The Performance Comparisons on Smartphone.

Method	Accuracy
Decision Tree	87.4%
Naive Bayes	88.9%
K-Means	52.1%
AdaBoost	40.8%
Gaussian Mixture	49.8%
Hyperbox Neural Network [27]	87.4%
FW K-Nearest Neighbor [28]	87.8%
FW Naive Bayes [28]	88.6%
Ours Hie-BKNN	**90.30%**

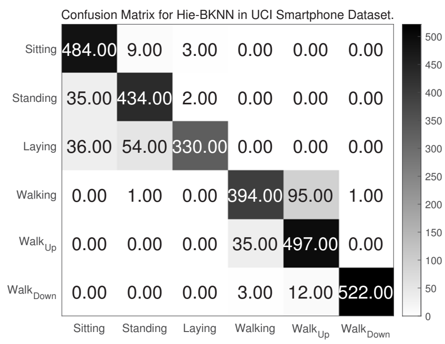

Fig. 4: **Confusion Matrix of Hie-BKNN for Smartphone.**

2) Results on UCI WISDM Dataset:

a) Compared to the state-of-the-art methods: Similar to the experiments in the Smartphone dataset, we compare and analyze the performance of the proposed method and the classic algorithm on the WISDM dataset. The experimental results are shown in Table IV. It can be clearly seen that, compared to all these mentioned algorithms, the Hie-BKNN model has the best performance (93.71%). Compared with the classic method mentioned, the biggest advantage of this method is that, in this paper, we mainly rely on the hierarchical modeling capabilities with our new hierarchical decision level fusion strategy, which help to achieve higher accuracy of activity recognition.

TABLE IV: The Performance Comparisons on WISDM.

Method	Accuracy
ANN	91.52%
ELM	92.06%
RF	91.77%
Deep LSTM	91.42%
Hyperbox Neural Network [27]	81.31%
FW K-Nearest Neighbor [28]	91.12%
FW Naive Bayes [28]	90.73%
DSmT-Based Kernel Density Estimation [7]	91.50%
Our Proposed Method (Hie-BKNN)	**93.71%**

V. Conclusion

In this paper, we propose a novel hierarchical belief K-Nearest Neighbors classifier (Hie-BKNN) for activity recognition in body sensor networks. Compared to existing works, Hie-BKNN first learn the hierarchical tree structure from the training datasets. Then, in the testing stage, the corresponding belief hierarchical tree is derived based on the belief K-Nearest Neighbors by using DTW distance. Moreover, the novel hierarchical fusion rule is proposed to combine all belief hierarchical trees to make the activity recognitions. Finally, we conducted extensive experiments on two publicly available activity classification datasets: Smartphone and WISDM. The results demonstrate the superiority of Hie-BKNN against other state-of-the-art traditional methods.

Acknowledgment

The authors thank the reviewers and editors for giving valuable comments, which are very helpful for improving this manuscript.

References

[1] M. Z. Uddin, M. M. Hassan, A. Alsanad, and C. Savaglio, "A body sensor data fusion and deep recurrent neural network-based behavior recognition approach for robust healthcare," *Information Fusion*, vol. 55, pp. 105–115, 2020.

[2] C. F. Martindale, V. Christlein, P. Klumpp, and B. M. Eskofier, "Wearables-based multi-task gait and activity segmentation using recurrent neural networks," *Neurocomputing*, vol. 432, pp. 250–261, 2021.

[3] D. Chen, Y. Cai, X. Qian, R. Ansari, W. Xu, K.-C. Chu, and M.-C. Huang, "Bring gait lab to everyday life: Gait analysis in terms of activities of daily living," *IEEE Internet of Things Journal*, vol. 7, no. 2, pp. 1298–1312, 2020.

[4] H. F. Nweke, Y. W. Teh, M. A. Al-garadi, and U. R. Alo, "Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges," *Expert Systems with Applications*, vol. 105, pp. 233–261, 2018.

[5] K. Viard, M. P. Fanti, G. Faraut, and J.-J. Lesage, "Human activity discovery and recognition using probabilistic finite-state automata," *IEEE Transactions on Automation Science and Engineering*, vol. 17, no. 4, pp. 2085–2096, 2020.

[6] A. Mannini and S. S. Intille, "Classifier personalization for activity recognition using wrist accelerometers," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 4, pp. 1585–1594, 2019.

[7] Y. Dong, X. Li, J. Dezert, M. O. Khyam, M. Noor-A-Rahim, and S. S. Ge, "Dezert-smarandache theory-based fusion for human activity recognition in body sensor networks," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 11, pp. 7138–7149, 2020.

[8] M. Wang, X. Wang, L. T. Yang, X. Deng, and L. Yi, "Multi-sensor fusion based intelligent sensor relocation for health and safety monitoring in bsns," *Information Fusion*, vol. 54, pp. 61–71, 2020.

[9] J. Li, Z. Wang, S. Qiu, H. Zhao, J. Wang, X. Shi, B. Liang, and G. Fortino, "Multi-body sensor data fusion to evaluate the hippotherapy for motor ability improvement in children with cerebral palsy," *Information Fusion*, vol. 70, pp. 115–128, 2021.

[10] S. U. Yunas and K. B. Ozanyan, "Gait activity classification using multi-modality sensor fusion: A deep learning approach," *IEEE Sensors Journal*, vol. 21, no. 15, pp. 16 870–16 879, 2021.

[11] G. Shafer, *A mathematical theory of evidence*. Princeton university press, 1976.

[12] M. Tabassian, R. Ghaderi, and R. Ebrahimpour, "Combination of multiple diverse classifiers using belief functions for handling data with imperfect labels," *Expert Systems with Applications*, vol. 39, no. 2, pp. 1698–1707, 2012.

[13] Y. Dong, X. Li, J. Dezert, R. Zhou, C. Zhu, L. Wei, and S. S. Ge, "Evidential reasoning with hesitant fuzzy belief structures for human activity recognition," *IEEE Transactions on Fuzzy Systems*, pp. 1–1, 2021.

[14] M. H. M. Noor, Z. Salcic, and K. I.-K. Wang, "Enhancing ontological reasoning with uncertainty handling for activity recognition," *Knowledge-Based Systems*, vol. 114, pp. 47–60, 2016.

[15] R. Babbar, I. Partalas, E. Gaussier, and M.-R. Amini, "On flat versus hierarchical classification in large-scale taxonomies," in *27th Annual Conference on Neural Information Processing Systems (NIPS 26)*, 2013, pp. 1824–1832.

[16] R. Babbar, I. Partalas, E. Gaussier, M.-R. Amini, and C. Amblard, "Learning taxonomy adaptation in large-scale classification," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 3350–3386, 2016.

[17] Z. Luo and Y. Deng, "A matrix method of basic belief assignment's negation in dempster–shafer theory," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 9, pp. 2270–2276, 2020.

[18] Y. Qu, L. Lin, F. Shen, C. Lu, Y. Wu, Y. Xie, and D. Tao, "Joint hierarchical category structure learning and large-scale image classification," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4331–4346, 2017.

[19] Z. Geler, V. Kurbalija, M. Ivanović, and M. Radovanović, "Weighted kNN and constrained elastic distances for time-series classification," *Expert Systems with Applications*, vol. 162, pp. 113–133, 2020.

[20] C. Faloutsos, M. Ranganathan, and Y. Manolopoulos, "Fast subsequence matching in time-series databases," in *International Conference on Data Engineering*, 2001.

[21] H. Li, J. Liu, Z. Yang, R. W. Liu, K. Wu, and Y. Wan, "Adaptively constrained dynamic time warping for time series classification and clustering," *Information Sciences*, vol. 534, pp. 97–116, 2020.

[22] D. Huang, C.-D. Wang, J.-S. Wu, J.-H. Lai, and C.-K. Kwoh, "Ultra-scalable spectral clustering and ensemble clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 6, pp. 1212–1226, 2020.

[23] Y. Pang, J. Xie, F. Nie, and X. Li, "Spectral clustering by joint spectral embedding and spectral rotation," *IEEE Transactions on Cybernetics*, vol. 50, no. 1, pp. 247–258, 2020.

[24] L. X. D.Anguita, A.Ghio and J.L.Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones," *Information Fusion*, vol. 46, pp. 147–170, 2019.

[25] G. M. Weiss, K. Yoneda, and T. Hayajneh, "Smartphone and smartwatch-based biometrics using activities of daily living," *IEEE Access*, vol. 7, pp. 133 190 – 133 202, 2019.

[26] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of machine learning research*, vol. 9, no. 11, 2008.

[27] M. Eastwood and C. Jayne, "Evaluation of hyperbox neural network learning for classification," *Neurocomputing*, vol. 133, pp. 249–257, 2014.

[28] A. Wang, G. Chen, J. Yang, S. Zhao, and C.-Y. Chang, "A comparative study on human activity recognition using inertial sensors in a smartphone," *IEEE Sensors Journal*, vol. 16, no. 11, pp. 4566–4578, 2016.