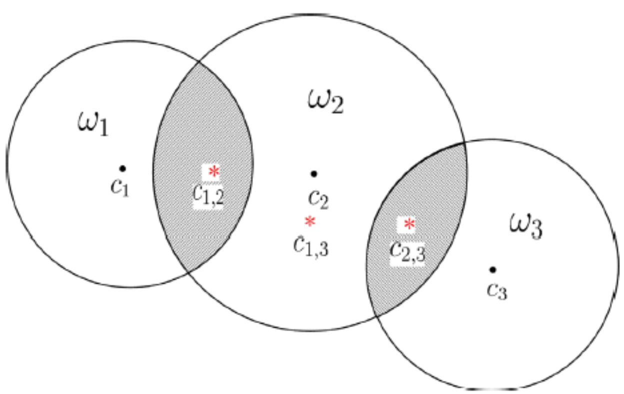
Pattern Classification With Belief Functions

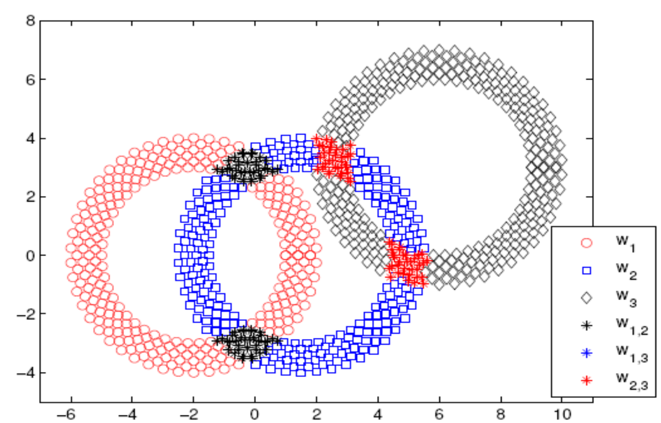
Zhunga Liu
liuzhunga@nwpu.edu.cn
Northwestern Polytechnical University
Xi'an, China

Outline

1 Introduction

- Traditional Pattern Classification methods tend to perform poorly on datasets containing uncertain data. This is unacceptable for some high-risk classification tasks.
- The imprecision can be better than error for the classification in some cases.
- Approaches that combine belief functions theory with pattern classification can improve the ability of classifiers to reduces classification errors and risks.





Outline

1. Introduction

1.1 Basic Knowledge

1.2 Typical evidential classifier

2. Recent development

2.1 Evidential classification of incomplete patterns

2.2 Evidential transfer classification for heterogeneous data

2.3 Combination of Transferable Classification

2.4 Evidential classifier fusion with refined reliability evaluation

2.5 Evidential Combination of classifiers with different frames of discernment

3. Conclusion

1.1 Basic Knowledge

- A frame of discernment Θ : A finite set of mutually exclusive elements in a domain

$$\Theta = \{\theta_1, \ldots, \theta_n\}$$

- The set consisting of all subsets of Θ is called the power-set of Θ, denoted as 2^{Θ}

$$\Theta = \{\theta_1, \theta_2, \theta_3\}$$

$$2^{\Theta} = \{\emptyset, \theta_1, \theta_2, \theta_3, \{\theta_1, \theta_2\}, \{\theta_1, \theta_3\}, \{\theta_2, \theta_3\}, \Theta\}$$

1.1 Basic Knowledge

A mass function (also called basic belief assignment): A mapping $m(\cdot):2^{\Theta}\rightarrow[0,1]$ assigning a mass value to each hypothesis $A\subseteq\Theta$ of the frame of discernment Θ such that

$$\begin{cases}\sum_{A\subseteq\Theta} m(A)=1\\ m(\emptyset)=0\end{cases}$$

If A satisfies $m(A)>0$, then A is called focal element.

If A satisfies $m(A)=\max(m(\cdot))$, then A is called to be main focal element.

1.1 Basic Knowledge

Based on the mass function, one can define belief functions $Bel(\cdot)$ and plausibility functions $Pl(\cdot)$:

$$Bel(A) = \sum_{A,B\in 2^{\Theta};B\subseteq A} m(B)$$

$$Pl(A) = \sum_{A,B\in 2^{\Theta};A\cap B\neq\emptyset} m(B) = 1 - Bel(\bar{A})$$

Belief functions $Bel(\cdot)$ and plausibility functions $Pl(\cdot)$ are commonly used to represent upper and lower bounds on probability.

1.1 Basic Knowledge

Dempster's Combination Rule

$$m(\cdot) = [m_1 \oplus m_2](\cdot)$$

Dempster's rule $\oplus$ is defined as

$$\begin{cases} m(\emptyset) = 0 \\ m_{DS}(A) = \dfrac{\sum_{B\cap C=A} m_1(B)m_2(C)}{1-\sum_{B\cap C=\emptyset} m_1(B)m_2(C)} = \dfrac{\sum_{B\cap C=A} m_1(B)m_2(C)}{\sum_{B\cap C\neq\emptyset} m_1(B)m_2(C)} \end{cases}$$

Conflict information is denoted as

$$k_{12} = \sum_{B\cap C=\emptyset} m_1(B)m_2(C)$$

Outline

1. Introduction

1.1 Basic Knowledge

1.2 Typical evidential classifier

2. Recent development

2.1 Evidential classification of incomplete patterns

2.2 Evidential transfer classification for heterogeneous data

2.3 Combination of Transferable Classification

2.4 Evidential classifier fusion with refined reliability evaluation

2.5 Evidential Combination of classifiers with different frames of discernment

3. Conclusion

1.2 Typical evidential classifier

Evidential K-NN(EK-NN) :

- The Basic Belief Assignment(BBA) of a sample x_i with respect to one of its neighbor x_j belonging to $\omega_s \in \Omega$ is defined as

$$m_j^{x_i}(\omega_s) = \alpha e^{-\gamma_s d^{\beta}}$$
$$m_j^{x_i}(\Omega) = 1 - \alpha e^{-\gamma_s d^{\beta}}$$

- α, β, γ are parameters that need to be regulated, d is the distance between the target and its nearest neighbor in the training data.

1.2 Typical evidential classifier

Evidential K-NN(EK-NN) :

- Each BBA in EK-NN contains only two focal elements ω_s and Ω. The classification result which contains only single and completely unknown classes is obtained by fusing the k BBA results of k nearest neighbors.

- EK-NN integrates the distance information between the target and its neighbors and the uncertainty of the data.

- In practice, when x is far away from the neighbors, x should be considered as noise regardless of value of K. Nevertheless, when K is big, most belief will be committed to a particular class in EK-NN.

1.2 Typical evidential classifier

Belief K-NN(BK-NN) :

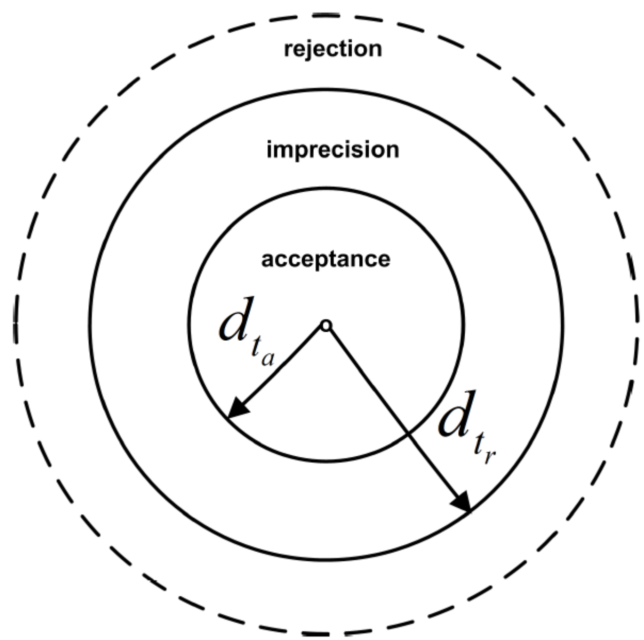
In BK-NN, BBA can partition targets into meta-classes like {w1,w2}, which reveals the imprecision of classification.

For each target x_i, its BBA with respect to its close neighbor x_j belonging to w_s can be given by the fusion of two BBA $m_1(\cdot|d_{ij})$ and $m_2(\cdot|d_{ij})$.

$$f_1(d_{ij}, d_{t_a}^{w_s}) \triangleq \frac{1}{1+e^{\lambda_j(d_{ij}-d_{t_a}^{w_s})}}$$

$$f_2(d_{ij}, d_{t_r}^{w_s}) \triangleq \frac{1}{1+e^{-\lambda_j(d_{ij}-d_{t_r}^{w_s})}}$$

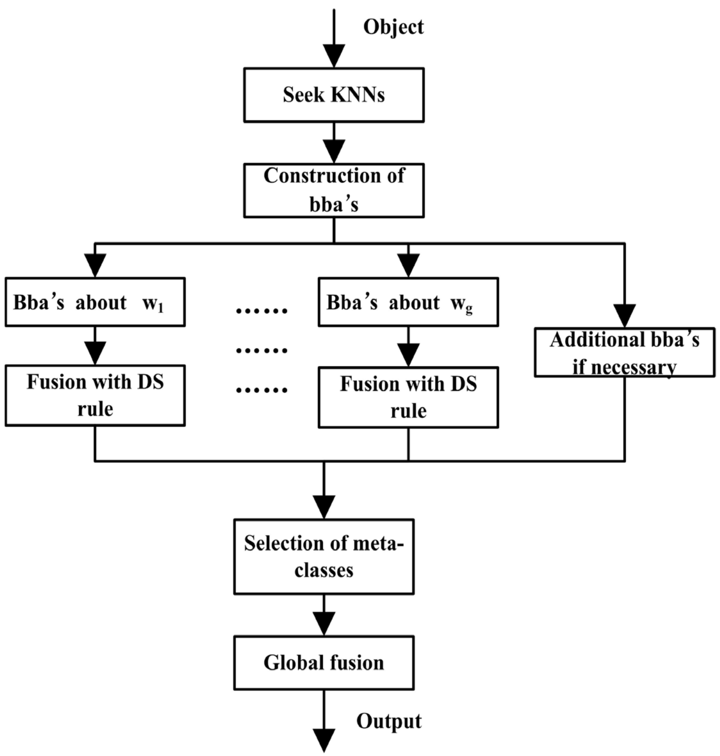
	$m_1(.\vert d_{ij})$	$m_2(.\vert d_{ij})$
w_s	$f_1(d_{ij}, d_{t_a}^{w_s})$	0
$\overline{w_s}$	0	$f_2(d_{ij}, d_{t_r}^{w_s})$
Ω	$1-f_1(d_{ij}, d_{t_a}^{w_s})$	$1-f_2(d_{ij}, d_{t_r}^{w_s})$



1.2 Typical evidential classifier

Belief K-NN(BK-NN) :

BK-NN first performs local fusion on the results of the same class. This can be fused directly using DS rules. The global fusion like DP rule is then performed on the results of the local fusion.



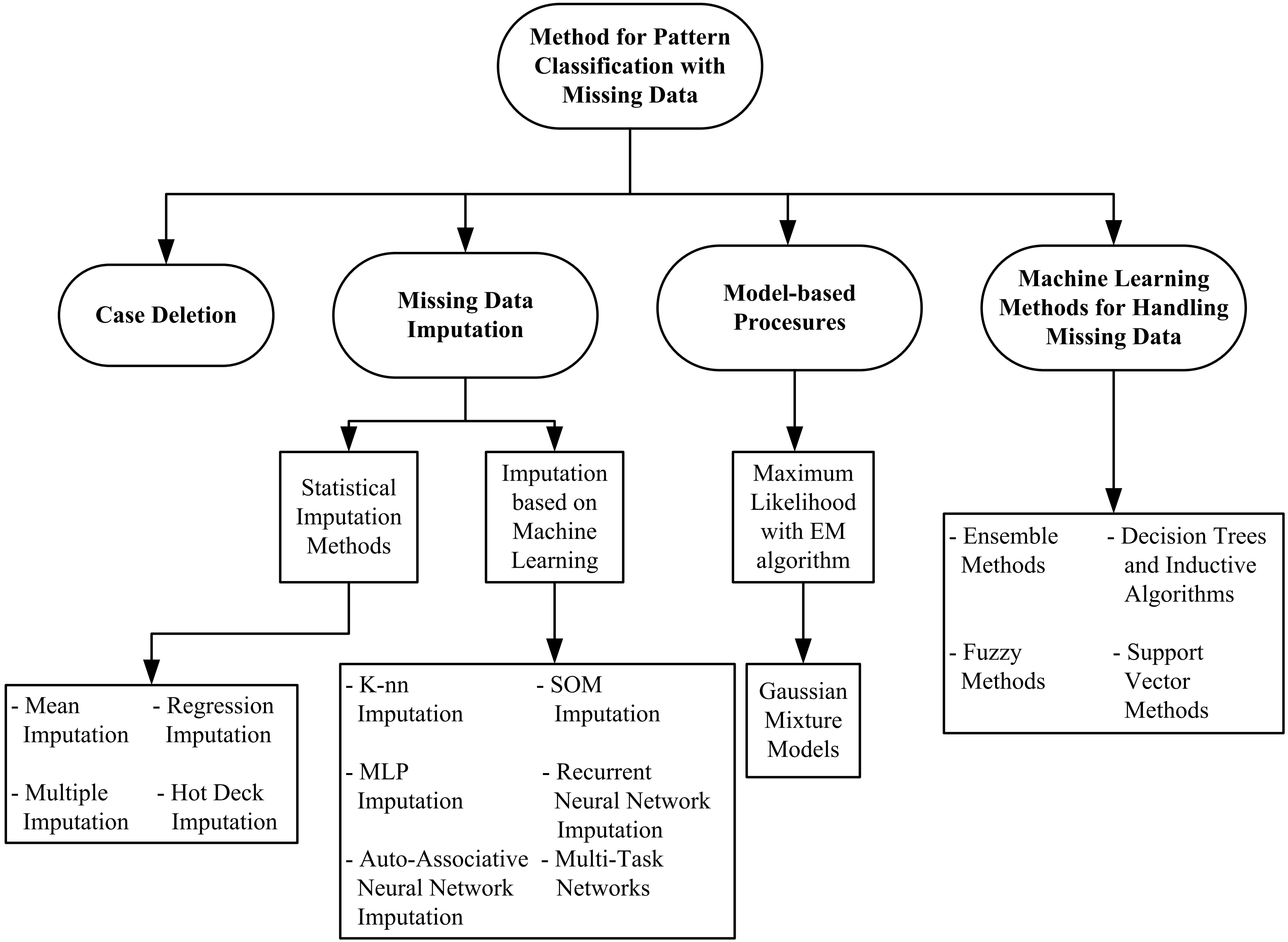
$$m_{1,s}(A) = \begin{cases} \sum_{\substack{B_1 \cap B_2 = A \\ A \neq \emptyset}} m_{1,s-1}(B_1) m_s^{w_s}(B_2) \\ \sum_{\substack{B_1 \cap B_2 = \emptyset \\ B_1 \cup B_2 = A}} m_{1,s}(B_1) m_s^{w_s}(B_2) \end{cases}$$

Outline

2.1. Evidential classification of incomplete patterns

- In practice, it is often encountered that the object attributes are partly missing due to sensor failure, record mistakes, etc.
- The incomplete pattern classification is an interesting topic, and there exist a number of methods to solve it.
- A credal classification method for incomplete pattern with adaptive imputation of missing values is proposed based on evidence theory.

2.1. Evidential classification of incomplete patterns

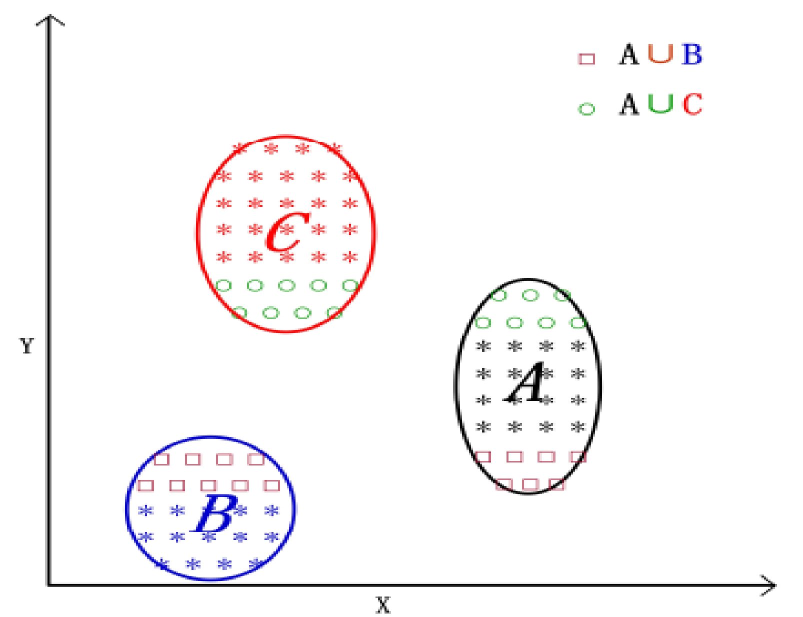


2.1. Evidential classification of incomplete patterns

In classification of incomplete pattern, the missing values can be important, and the classification result of pattern with different estimations may be distinct (e.g. AUB).

However, sometimes the missing values have little influence on the classification (e.g. A).

The influence of missing values mainly depends on the context.



2.1. Evidential classification of incomplete patterns

We present a classification method with the selective imputation of missing values based on belief functions.

- The object is directly classified using the given attributes.
- If it cannot be clearly classified, the missing values will be imputed before classification.

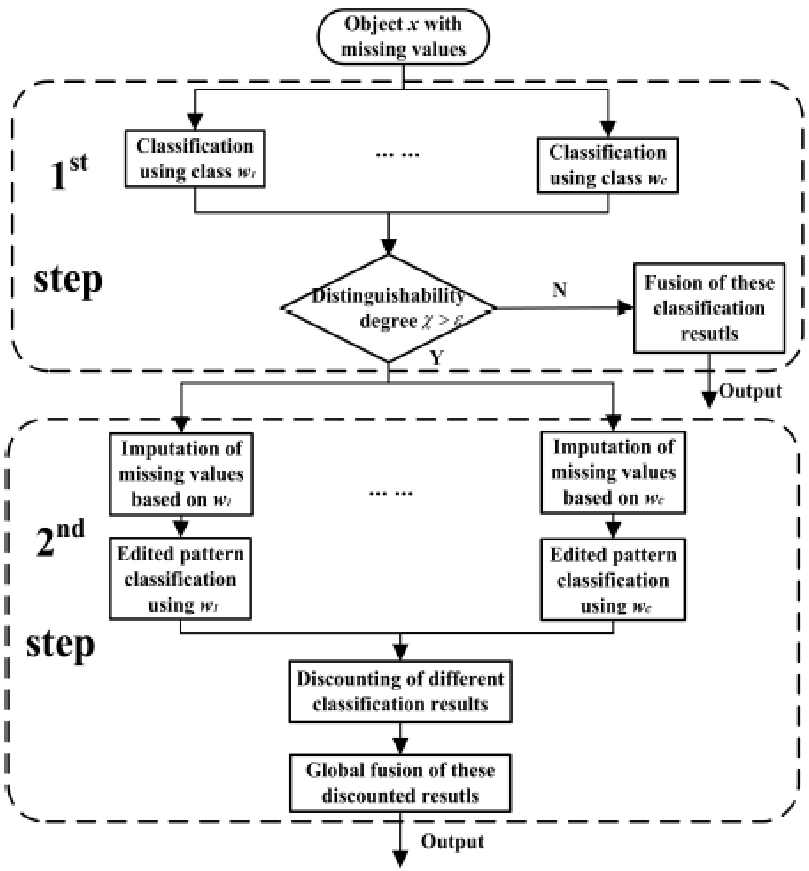
Credal classification allows the objects to belong to specific classes, and the sets of classes (i.e. meta-classes).The object hard to classify will be cautiously committed to meta-class. This can reduce errors, and reveal the imprecision of classification.

2.1. Evidential classification of incomplete patterns

It mainly consists of two important steps:

① Direct classification of incomplete pattern

② Classification with imputation of missing values



2.1.1 Direct classification using available data

A test data set X = {x_1,…, x_N} is classified with training data set Y = {y_1,…, y_H} in frame {ω_1,…, ω_c}.

The prototype of each class i.e. {o_1, . . . , o_c} is defined by

$$O_g = \frac{1}{N_g} \sum_{y_j \in \omega_g} y_j$$

In a *c*-class problem, one can get *c* pieces of simple classification result according to each class:

$$\begin{cases} m_i^{o_g}(\omega_g) = e^{-\eta d_{ig}} \\ m_i^{o_g}(\Omega) = 1 - e^{-\eta d_{ig}} \end{cases}$$

2.1.1 Direct classification using available data

The indistinguishability degree of an object *x* associated with different classes given by

$$\chi_i = \frac{m_i^{o_{2nd}}(\omega_{2nd})}{m_i^{o_{max}}(\omega_{max})}$$

- If it is smaller than a given threshold, the available attributes are considered sufficient for good classification.
- The *c* BBA's are directly combined by DS rule to obtain the final classification results.
- If it is bigger than a given threshold, the missing attributes play a crucial role in the classification.

2.1.2 Classification with imputation of missing values

Multiple estimation of missing values:

- SOM is applied in each training class, and M×N nodes will be obtained. The K nearest nodes can be found.
- The K nodes have different contributions in estimation of missing values. The weight is defined based on distance to the node

$$p_{ik}^{\omega_g} = e^{(-\lambda d_{ik}^{\omega_g})}$$

with

$$\lambda = \frac{cNM(cNM-1)}{2\sum_{i,j} d(\sigma_i, \sigma_j)}$$

2.1.2 Classification with imputation of missing values

- The weighted mean value of the selected K nodes will be used to fill the missing values

$$\hat{y}_i^{\omega_g} = \left(\sum_{k=1}^{K} p_{ik}^{\omega_g} \sigma_k^{\omega_g} \right) / \left(\sum_{k=1}^{K} p_{ik}^{\omega_g} \right)$$

- By doing this, one gets *C* versions of edited pattern according to each training class, and they will be respectively classified based on each training class.

2.1.3 Ensemble classifier

- The weighting factors of these *c* pieces of results are defined by the sum of weights of the K SOM nodes

$$\rho_i^{\omega_g} = \sum_{k=1}^{K} p_{ik}^{\omega_g}$$

- The relative weighting factors are defined by

$$\hat{\alpha}_i^{\omega_g} = \frac{\rho_i^{\omega_g}}{\rho_i^{\omega_{max}}}$$

2.1.3 Ensemble classifier

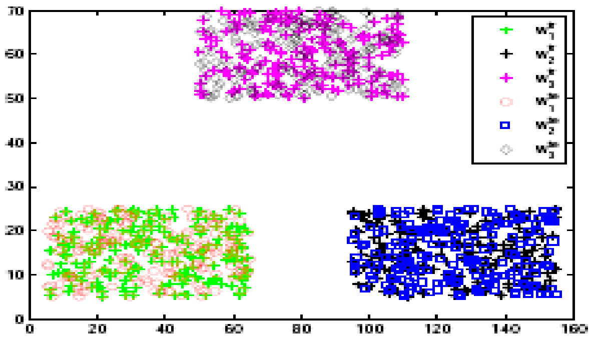
- The c pieces of results are discounted by relative weighting factors

$$\begin{cases} \hat{m}_i^{o_g}(\omega_g) = \alpha_i^{\omega_g} m_i^{o_g}(\omega_g) \\ \hat{m}_i^{o_g}(\Omega) = 1 - \alpha_i^{\omega_g} + \alpha_i^{\omega_g} m_i^{o_g}(\Omega) \end{cases}$$

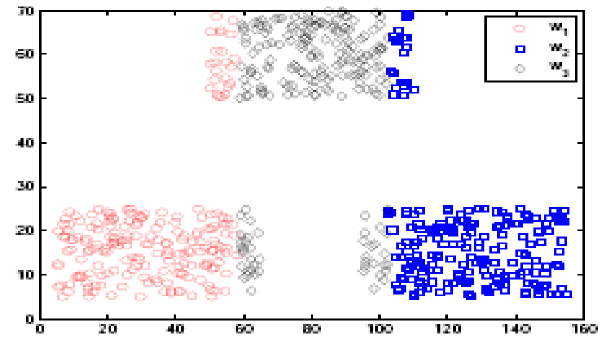
- These discounted BBA's are combined for the classification

$$\begin{cases} m_i(\omega_g) = \hat{m}_i^{o_g}(\omega_g) \prod_{j \neq g} \hat{m}_i^{o_j}(\Omega) \\ m_i(A) = \prod_{\cup_j \omega_j = A} \hat{m}_i^{o_j}(\omega_j) \prod_{k \neq j} \hat{m}_i^{o_k}(\Omega) \end{cases}$$

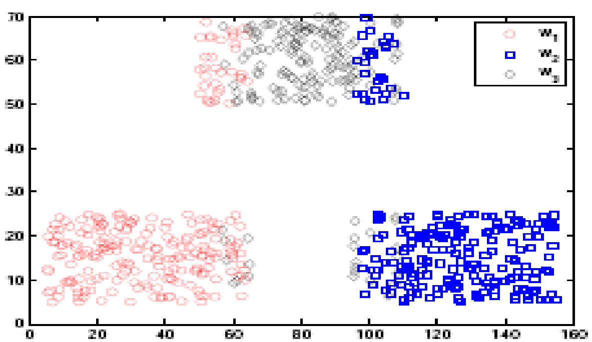
2.1.4 Experiments



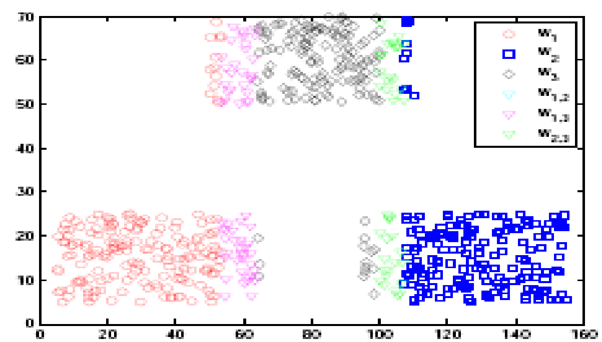
(a). Training data (* symbols), and test data.



(b). Classification result by FCMI ($Re = 14.67, time = 0.0469s$).



(c). Classification result by KNNI ($Re = 14.17, time = 7.9531s$).



(d). Classification result by CCAI ($Re = 5.83, Ri_2 = 16.83, time = 0.0469s$).

CCAI works with relatively low computation burden, and it can capture the imprecision of classification thanks to belief functions.

2.1.4 Experiments

name	classes	attributes	instances
Breast (B)	2	9	699
Iris (I)	3	4	150
Seeds (S)	3	7	210
Wine (W)	3	13	178

CLASSIFICATION RESULTS FOR DIFFERENT REAL DATA SETS (IN %).

data set	n	FCMI Re	KNNI Re	PCC $\{Re, Ri_2\}$	CCAI $\{Re, Ri_2\}$
Breast	3	3.81	3.95	{3.81, 2.34}	{3.66, 0}
	6	7.32	8.20	{5.42,1.32}	{4.83, 1.61}
	7	11.42	11.54	{10.10, 2.64}	{9.00, 0.66}
Iris	1	7.33	4.89	{5.33, 2.67}	{4.00, 1.33}
	2	14.11	11.33	{8.67,4.00}	{8.00, 4.67}
	3	17.33	18.44	{12.67, 9.33}	{11.33, 12.00}
Seeds	2	15.24	11.19	{9.52, 4.76 }	{9.52, 0}
	4	17.14	11.98	{10.48, 4.29}	{10.00, 0.48}
	6	20.95	25.71	{16.19, 14.76}	{16.19, 13.81}
Wine	3	26.97	26.97	{26.97, 1.69}	{6.74, 1.12}
	7	33.24	30.43	{29.78, 2.25}	{7.30, 3.93}
	11	33.43	30.90	{30.34, 2.81}	{12.36, 3.93}

Credal classification is efficient to capture the imprecision and reduce errors. The objects in meta-classes are hard to correctly classify, and they should be cautiously treated.

Outline

1. Introduction

1.1 Tranditional classifier

1.2 Typical evidential classifier

2. Recent development

2.1 Evidential classification of incomplete patterns

2.2 Evidential transfer classification for heterogeneous data

2.3 Combination of Transferable Classification

2.4 Evidential classifier fusion with refined reliability evaluation

2.5 Evidential Combination of classifiers with different frames of discernment

3. Conclusion

2.2 Evidential transfer classification for heterogeneous data

- It remains a challenging problem for data classification without training patterns.
- In many applications, there may exist some labeled data in other related domain (source domain). Such labeled data can help to solve the classification problem in target domain.
- The source domain and target domain are heterogeneous and they represent the distinct feature spaces.
- In previous work, each object usually has only one mapping value in corresponding domain. This cannot well reflect the uncertainty of mapping/transformation.

2.2 Evidential transfer classification for heterogeneous data

A new transfer classification method for heterogeneous data is proposed based on evidence theory. It mainly consists of three important steps:

① Estimation of mapping value scope in source domain

② Determination of Mapping value in Source domain

③ Classification of mapping value based on evidence theory

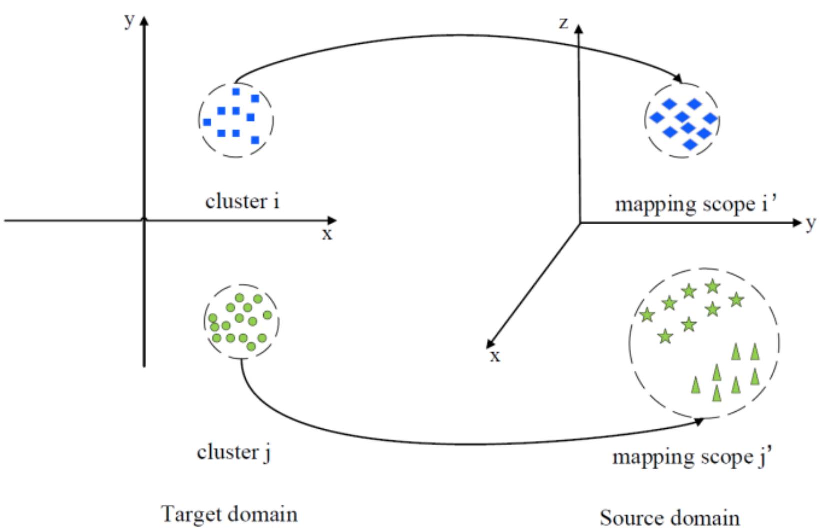
2.2.1 Estimation of mapping value scope in source domain

- For classification, we need to transfer query object $\hat{x}$ to the source domain using these labeled patterns $\tilde{Y} = \{\widetilde{y_1}, \ldots, \widetilde{y_g}\}$
- Due to the heterogeneousness of domains, it is difficult to directly obtain one exact mapping value for $\hat{x}$ in source domain.
- When the patterns in source domain are divided in some clusters by SOM, we find the nearest cluster of $\hat{x}$ which is denoted as $X_c = \{x_{c_1}, \ldots, x_{c_k}\}$.
- The actual mapping values of the patterns in X_c as $\mathrm{Y_c} = \{\mathrm{y_{c_1}}, \ldots, \mathrm{y_{c_k}}\}$ are considered as the possible mapping values of $\hat{x}$ in source domain.

2.2.2 Determination of Mapping value in Source domain

- The mapping between such heterogeneous domains may be very uncertain.
- The estimation uncertainty (dispersion degree) of the possible mapping values can be characterized by the standard deviation (std) of

$$\overline{y}_c = \frac{1}{c_k}\sum_{j=1}^{c_k} y_j, \qquad \delta_c = \sqrt{\frac{1}{c_k}\sum_{j=1}^{c_k} |y_j - \overline{y}_c\,|}$$



The mapping values of cluster j in source domain are very disperse, and this means that the transformation between such heterogeneous domains is very uncertain.

2.2.2 Determination of Mapping value in Source domain

- Based on the obtained standard deviation, some potential mapping values of $\hat{\mathrm{x}}$ are selected as

$$\hat{\mathbf{y}} = \begin{cases} \overline{\boldsymbol{y}}_{c}, & \delta_c \leq t \\ \{\overline{\boldsymbol{y}}_{c_1}, \dots, \overline{\boldsymbol{y}}_{c_K}\}, & \delta_c > t \end{cases}$$

- It can be seen that the query pattern $\hat{\mathrm{x}}$ can have either one or multiple mapping values in source domain depending on the estimation uncertainty.

2.2.3 Classification of mapping value based on evidence theory

- If there is only one mapping value, i.e. $\bar{\boldsymbol{y}}_c$, it can be directly classified according to $\bar{\boldsymbol{y}}_c$.
- When $\hat{\mathrm{x}}$ has multiple possible mapping values as $\{\bar{\boldsymbol{y}}_{c_1}, \ldots, \bar{\boldsymbol{y}}_{c_K}\}$, these mapping values will be respectively classified by the trained classifier in source domain.
- Then, the classification results of these possible mapping values are combined by evidence theory.

$$\mathbf{m} = {}^{w_1}\,\mathbf{m}_1 \oplus {}^{w_2}\,\mathbf{m}_2 \ldots \oplus {}^{w_{c_K}}\,\mathbf{m}_{c_K}$$

2.2.5 Experiments

- Transfer classification of UCI data

BASIC INFORMATION ON USED UCI DATA.

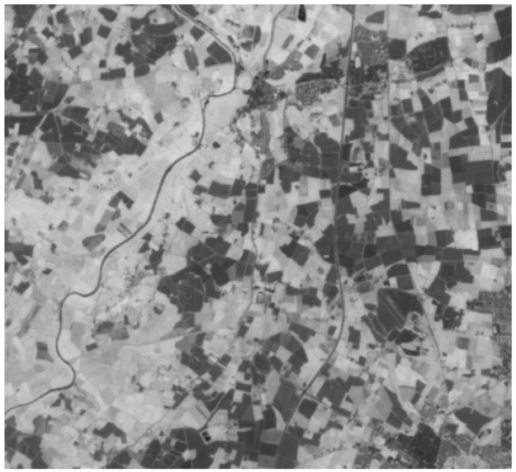
Data	Class	Attribute	Instance	N_s	N_t
Abalone	3	8	4177	5	3
Cmsc	2	18	540	10	8
Magic	2	10	19020	6	4
Page	5	10	5473	6	4
Seeds	3	7	210	4	3
Statlog	6	36	4225	16	20
Vehicle	4	18	846	10	8
Vertebral	3	6	310	4	2
Waveform	3	21	5000	9	12
Satimage	7	36	6435	20	16
Pima	2	8	768	5	3

CLASSIFICATION ACCURACY OF DIFFERENT METHODS (IN%).

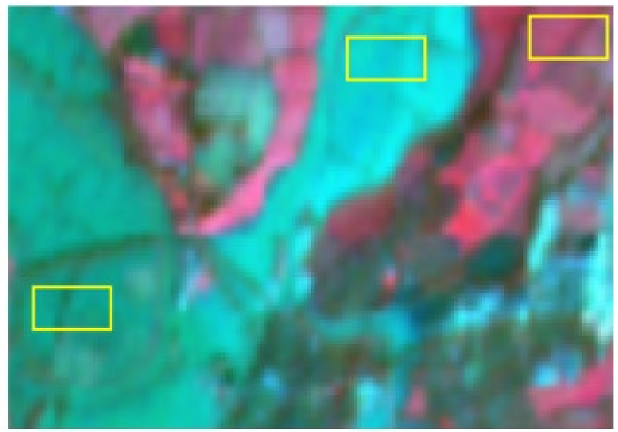
Data	FCM	CUCS	SVMHTC	CUMMV	EHTC
Abalone	48.33	50.30	47.26	49.61	**51.69**
Cmsc	76.52	89.81	85.00	83.15	**91.38**
Magic	63.66	70.04	68.76	**72.16**	71.97
Page	76.98	87.30	80.98	83.62	**89.95**
Seeds	77.62	**83.05**	78.57	79.46	82.38
Statlog	68.14	71.81	70.98	70.29	**76.00**
Vehicle	46.99	51.38	52.60	49.88	**54.02**
Vertebral	51.03	66.13	64.19	71.67	**80.00**
Waveform	61.42	73.84	65.90	62.73	**74.56**
Satimage	50.30	69.73	59.27	60.49	**70.15**
Pima	63.09	65.63	63.54	65.31	**67.06**

2.2.5 Experiments

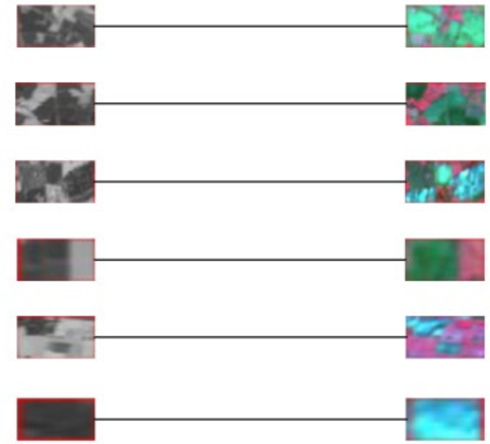
- Transfer classification of heterogeneous remote sensing data



Target



Source



Pairs

PERFORMANCE EVALUATION ON NDVI (TARGET DOMAIN) & SPOT (SOURCE DOMAIN) IMAGES (IN%).

Methods	OA	Kappa
SVMHTC	81.97	71.26
CUCS	84.20	78.61
FCM	70.05	65.17
CUMMV	85.01	78.65
EHTC	**87.02**	**80.73**

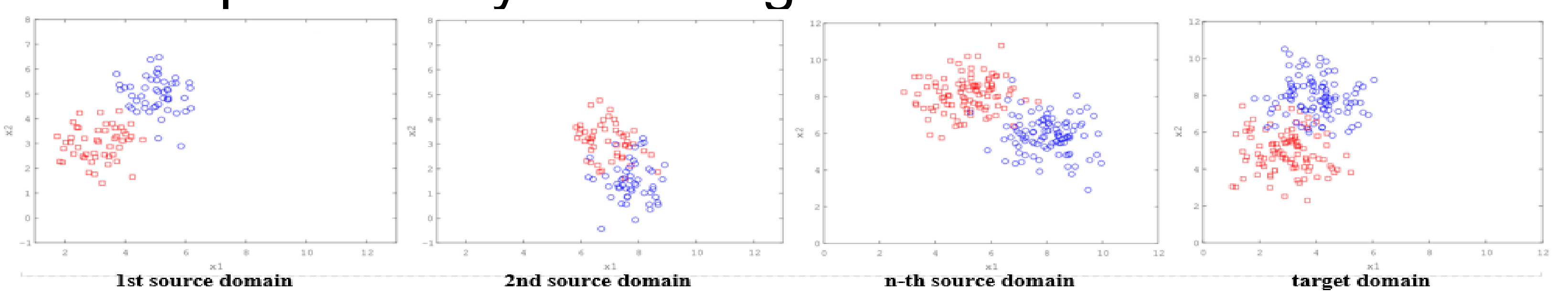
PERFORMANCE EVALUATION ON NDVI (SOURCE DOMAIN) & SPOT (TARGET DOMAIN) IMAGES (IN%).

Method	OA	Kappa
SVMHTC	76.43	66.08
CUCS	80.95	74.51
FCM	68.60	61.99
CUMMV	79.56	72.15
EHTC	**82.39**	**75.28**

Outline

2.3 Combination of Transferable Classification

- Transfer learning is an effective tool to solve the classification problem with few or even no labeled data. It uses the knowledge in source domain to help build classification model in target domain.
- In practice, there may exist multiple source domains, which can provide more or less complementary knowledge for pattern classification in the target domain.
- However, the current methods usually concentrate/merge different source domains, and it is difficult to fully exploit the complementary knowledge of different sources



2.3 Combination of Transferable Classification

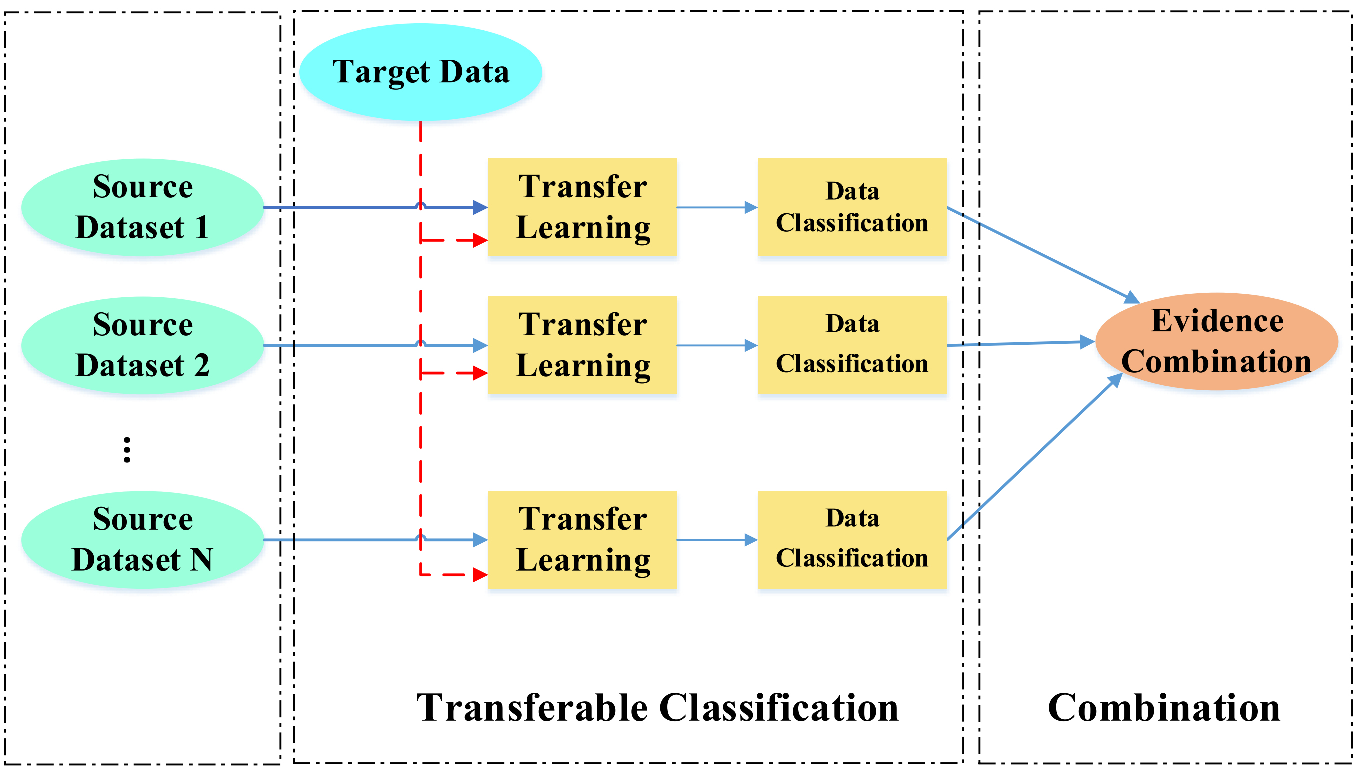
- In order to utilize the complementary knowledge of multiple source domains, a decision-level combination method is proposed for the multisource domain adaptation based on evidential reasoning.
- The classification results obtained from different source domains usually have different reliabilities/weights, which are calculated according to domain consistency.
- Thus, the multiple classification results are discounted by the weights, and then DS rule is employed to combine these discounted results.

2.3 Combination of Transferable Classification

It mainly consists of two important steps:

① Transferable classification

② Evidential Combination



2.3.1 Transferable classification

- The reconstruction error of transformed patterns in the source and target domains by mapping is defined by

$$\left\| \frac{1}{n_s} \sum_{i=1}^{n_s} A^T x_i - \frac{1}{n_t} \sum_{j=n_s+1}^{n_s+n_t} A^T x_j \right\| = tr(A^T X M X^T A)$$

- Then, we can obtain the transferable classification results

$$m_i(A) \qquad i = 1, \ldots, n$$

2.3.2 Evidential Combination

- Estimate the weights of transferable classification results

- Distance before matching

$$\widetilde{d_{S_iT}}\left(D_{S_i}, D_T\right) = 2\left(1 - 2\widetilde{e_i}(C)\right), i = 1, \cdots, n.$$

$$\widetilde{e_i}(C) = \frac{1}{N_i + N_T}\sum_{j=1}^{N_i+N_T}\left|C\left(x_j\right) - \widetilde{y_j}\right|, i = 1, \cdots, n$$

- Distance after matching

$$\widehat{d_{S_iT}}\left(D_{S_i}, D_T\right) = 2\left(1 - 2\widehat{e_i}(C)\right), i = 1, \cdots, n.$$

$$\widehat{e_i}(C) = \frac{1}{N_i + N_T}\sum_{k=1}^{N_i+N_T}\left|C\left(\hat{x}_k\right) - \tilde{y}_k\right|, i = 1, \cdots, n$$

2.3.2 Evidential Combination

- The distance before matching reflects the difference degree of different sources under the original feature representation, and the distance after matching reflects the difference degree of different sources under the new feature representation.

- Final Distance

$$d_j\left(D_{S_j},D_T\right)=\sqrt{\tilde{d}_j\left(D_{S_j},D_T\right)\cdot\hat{d}_j\left(D_{S_j},D_T\right)}$$

2.3.2 Evidential Combination

- Estimated weights

$$\beta_i = \frac{\tilde{\beta}_i}{\max(\tilde{\beta}_1, \cdots, \tilde{\beta}_n)}, i = 1, \cdots, n$$

$$\tilde{\beta}_i = e^{-d_{s_i T}}, i = 1, \cdots, n.$$

- Discounted results

$$\begin{cases} \widetilde{m_i}(A) = \beta_i \cdot m_i(A), A \in 2^{\Omega}, A \neq \Omega \\ \widetilde{m_i}(\Omega) = 1 - \beta_i + \beta_i \cdot m_i(\Omega) \end{cases}$$

- Combination results

$$m = \tilde{m}_1 \oplus \cdots \oplus \tilde{m}_n.$$

2.3.4 Experiments

- Multi-source Transfer classification of benchamrk data

BASIC INFORMATION OF THE BENCHMARK DATA SETS

Data set	Domain	Feature	Sample	Class
Office+Caltech10	Amazon (A)	800	958	10
	Caltech (C)	800	1123	10
	DSLR (D)	800	157	10
	Webcam (W)	800	295	10
PIE	PIE_C05 (PIE1)	1024	3332	68
	PIE_C07 (PIE2)	1024	1629	68
	PIE_C09 (PIE3)	1024	1632	68
	PIE_C27 (PIE4)	1024	3329	68
	PIE_C29 (PIE5)	1024	1632	68
Office-31	Amazon (A)	800	2715	31
	DSLR (D)	800	482	31
	Webcam (W)	800	776	31

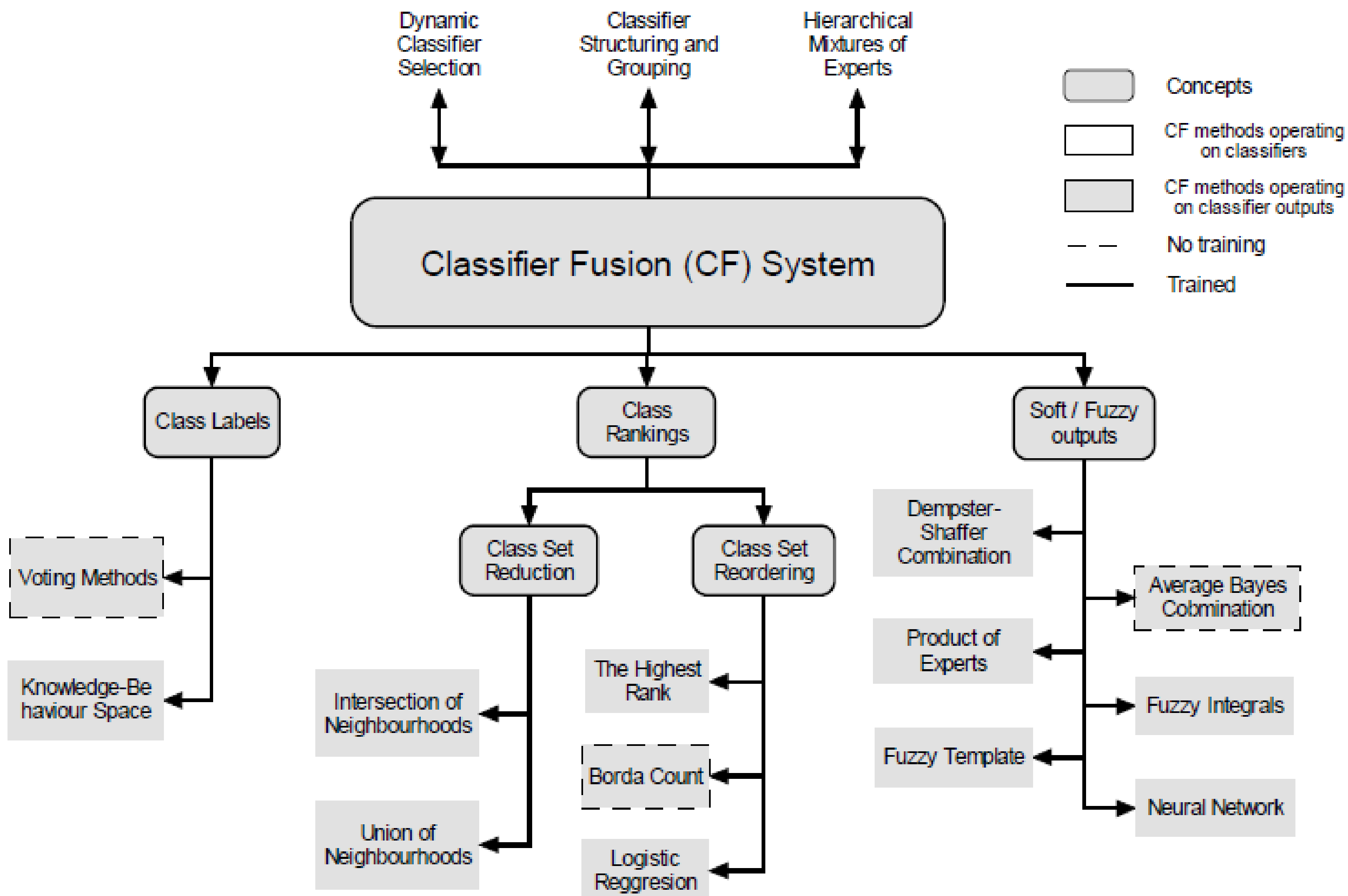
CLASSIFICATION PERFORMANCE OF DIFFERENT FUSION METHODS BASED ON TCA IN THE OFFICE+CALTECH10 DATA SETS

	$\mathcal{R}_{SVM}$	$\mathcal{R}_{k-NN}$	$\mathcal{R}_{CMSD}$	$\mathcal{R}_{MV}$	$\mathcal{R}_{AF}$	$\mathcal{R}_{DS}$	$\mathcal{R}_{GFK}$	$\mathcal{R}_{CORAL}$	$\mathcal{R}_{TCA}$	$\mathcal{R}_{TCA+CMSD}$	$\mathcal{R}_{TCA+MV}$	$\mathcal{R}_{TCA+WMV}$	$\mathcal{R}_{TCA+AF}$	$\mathcal{R}_{TCA+WAF}$	$\mathcal{U}_{TCA+AFC}$	$\mathcal{R}_{TCA+DS}$	$\mathcal{U}_{TCA+DSC}$	$\mathcal{R}_{TCA+WDS}$	$\mathcal{U}_{TCA+WDSC}$
C→A	30.53	22.76					41.02	20.15	**47.91**										
D→A	29.58	26.62	24.32	26.10	29.12	28.91	32.05	30.69	20.88	46.35	33.51	41.23	36.95	40.92	*40.75±0.27*	38.00	*39.15±0.03*	45.20	***48.72±0.12***
W→A	29.54	22.65					31.84	26.20	33.30										
A→C	44.08	24.04					40.25	23.06	41.41										
D→C	28.50	26.09	24.04	26.63	26.09	22.80	30.10	32.06	27.16	43.19	37.13	40.61	41.23	42.83	*43.60±0.72*	42.65	*43.33±0.01*	**43.46**	***45.89±0.03***
W→C	26.51	18.08					30.72	25.73	31.79										
A→D	40.13	23.57					36.61	30.75	36.94										
C→D	45.22	24.84	37.58	26.11	36.94	33.76	41.40	26.75	49.68	70.06	59.87	73.89	77.07	80.89	*77.18±1.84*	73.89	*75.44±0.14*	79.62	***82.96±0.31***
W→D	62.82	44.59					77.90	73.25	**81.53**										
A→W	39.66	27.12					40.00	26.10	40.00										
C→W	42.37	26.10	29.49	32.20	39.32	35.25	40.68	19.66	45.08	57.29	52.88	64.41	63.39	65.76	*64.29±0.97*	61.69	*64.40±0.06*	**72.20**	***74.26±0.18***
D→W	65.42	43.73					64.41	63.56	67.46										
Average	40.39	27.52	28.85	27.76	32.87	30.18	40.25	33.16	43.57	54.22	45.84	55.04	54.66	57.60	*56.55*	54.06	*55.42*	**60.21**	***62.75***

Outline

2.4. Classifier fusion with refined reliability evaluation

- Classifier fusion is an efficient strategy to improve the classification accuracy. The classification produced by individuals may be uncertain, and different classifiers can provide complementary knowledge.
- The complementarity (diversity) among classifiers can be achieved by extracting different features, by employing different classifiers.
- There exist a number of classifiers fusion methods.


Dynamic Classifier Selection
Classifier Structuring and Grouping
Hierarchical Mixtures of Experts
Concepts
CF methods operating on classifiers
CF methods operating on classifier outputs
No training
Trained
Classifier Fusion (CF) System
Class Labels
Class Rankings
Soft / Fuzzy outputs
Voting Methods
Knowledge-Be haviour Space
Class Set Reduction
Class Set Reordering
Intersection of Neighbourhoods
Union of Neighbourhoods
The Highest Rank
Borda Count
Logistic Reggresion
Dempster-Shaffer Combination
Average Bayes Cobmination
Product of Experts
Fuzzy Integrals
Fuzzy Template
Neural Network

2.4. Classifier fusion with refined reliability evaluation

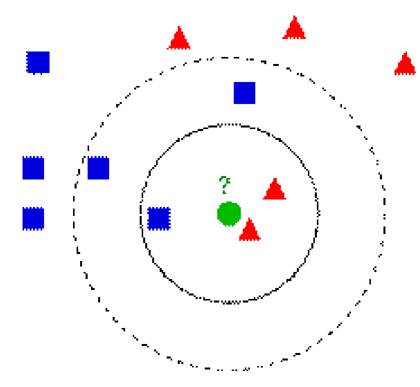
- Classifiers to combine generally have different reliabilities, and the proper reliability evaluation is helpful to further improve the accuracy.
- The weight of classifier is often determined by the overall classification performance in training set.
- This cannot efficiently characterize the difference of each pattern.

2.4. Classifier fusion with refined reliability evaluation

- The reliabilities of classification results are related with the objects to classify.
- Different elements in soft classification result may also have different reliabilities, since the difference of the output value and the expected value usually is not the same.
- We want to develop a method for revising the classifier output by a refined reliability evaluation. By doing this, one expects to make the output closer to truth.

2.4.1 Basic principle

- Refined reliability matrix representing the conditional probability of object belonging to class ω_i but classified to ω_j.
- The given classifier is expected to produce the similar performance on the object and on its close neighbors.
- The K nearest neighbors of object are found in each attribute space, and they are handled by the same classifier.
- We can estimate the reliability matrix according to the classification result of these neighbors.



2.4.2 Refined reliability evaluation

In the estimation of conditional probability, the distance between the object y and the neighbor x_k, is considered

If y is far from x_k, x_k is considered with a small influence on the estimation. Thus, the bigger distance, the smaller weight of the neighbor. The weighted sums of the conditional probabilities of x_k, belonging to class ω_i but classified to ω_j is computed by

$$\beta_{ij} = \sum_{X_k} P(\hat{c}(X) = \omega_j | c(X_k) = \omega_i) \cdot \delta_k$$

$$= \sum_{X_k | c(X_k) = \omega_i} P_k(j) \cdot \delta_k$$

$$\delta_k = e^{-\gamma \cdot d_k}$$ and $$d_k \triangleq \frac{d(y, X_k)}{\min_{k \in [1,K]} d(y, X_k)}$$

2.4.2 Refined reliability evaluation

The conditional probability $\mathrm{P}\left(\hat{c}(\mathrm{y}) = w_j | c(y) = w_i\right)$ should be proportional to β_{ij} as $\mathrm{P}\left(\hat{c}(\mathrm{y}) = w_j | c(y) = w_i\right) = \rho . \beta_{ij}$. The reliability matrix $\mathrm{R} = \left[r_{ji}\right]$ can be derived according to Bayes rule

$$r_{ji} = P\left(c(y) = \omega_i \middle| \hat{c}(y) = \omega_j\right)$$
$$= \frac{P(\hat{c}\left((y) = \omega_j \middle| c(y) = \omega_i\right) P(c(y) = \omega_i)}{\sum_{l=1}^{c} P\left(\hat{c}(y) = \omega_j \middle| c(y) = \omega_l\right) P(c(y) = \omega_l)}$$

The priori probability is usually assumed uniformly distributed

$$r_{ji} = \frac{P(c(y = \omega_j | \hat{c}(y) = \omega_i)}{\sum_{l=1}^{c} P(c(y) = \omega_j | \hat{c}(y) = \omega_l)} = \frac{\rho \beta_{ij}}{\rho \sum_{l=1}^{c} \beta_{lj}} = \frac{\beta_{ij}}{\sum_{l=1}^{c} \beta_{lj}}$$

2.4.3 Cautious discounting

- Once reliability matrix R is determined, we will modify the classification result to make it closer to the truth.
- The matrix is estimated by the neighborhoods, which are more or less different from the object. So we must not be completely confident about it.
- We propose a cautious discounting rule to transfer the classification knowledge to the associated partial ignorance

$$m_n(\omega_i \cup \omega_j) = m_{n1}(\omega_i \cup \omega_j) + m_{n2}(\omega_j \cup \omega_i)$$
$$= P\left(c(y) = \omega_i \middle| \hat{c}(y) = \omega_j\right) \cdot \mu_n(j)$$
$$+P\left(c(y) = \omega_j \middle| \hat{c}(y) = \omega_i\right) \cdot \mu_n(i), \qquad if \;\; i \neq j$$

$$m_n(\omega_i) = P(c(y) = \omega_i | \hat{c}(y) = \omega_i) \cdot \mu_n(i), \qquad if \;\; j = i$$

2.4.4 Classifier fusion

- DS rule is employed here to combine the discounted classification results from different classifiers

$$m(A) = m_1 \oplus m_2 = \begin{cases} \dfrac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - \sum_{B \cap C = \emptyset} m_1(B) m_2(C)}, \forall A \neq \emptyset \in 2^{\Omega}, B, C \in 2^{\Omega} \\ 0, A = \emptyset \end{cases}$$

- In the final fusion results, the plausibility functions Pl(.) is used here for decision making support

$$\omega_g = \arg\max_j Pl(\omega_j)$$

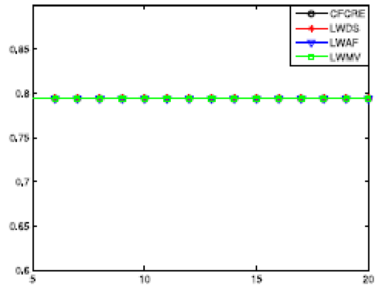
2.4.5 Experiment

Data	Classes	Attributes	Instances
Texture (Te)	11	40	5500
Vehicle (Ve)	4	18	946
Movement-libras (ML)	15	90	360
Sonar (So)	2	60	208
Segment (Se)	7	19	2310

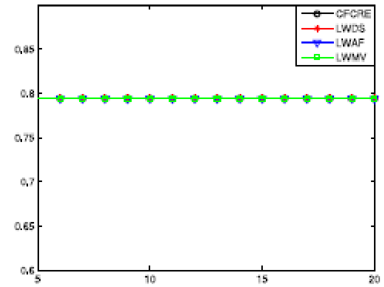
Data	N	$[AC_l, AC_u]$	WMV	WAF	WDS	NEW
Te	4	[59.18, 65.35]	81.62	80.87	83.13	95.38
Te	8	[53.00, 68.36]	84.59	82.59	85.28	94.44
Ve	2	[38.18, 49.53]	52.25	51.57	51.56	65.17
Ve	6	[38.53, 49.29]	55.16	52.40	55.78	64.93
ML	9	[26.67, 44.17]	55.38	54.36	61.55	76.68
ML	15	[24.44, 38.06]	56.56	49.03	62.10	71.65
So	6	[53.37, 73.08]	74.37	73.29	77.31	81.97
So	20	[53.37, 74.04]	72.57	71.36	75.81	78.85
Se	7	[32.73, 67.10]	76.60	80.28	80.61	90.70
Se	2	[63.72, 69.87]	82.34	81.63	82.36	91.79

The proposed method can efficiently improve the accuracy thank to the refined reliability evaluation and the cautious discounting technique.

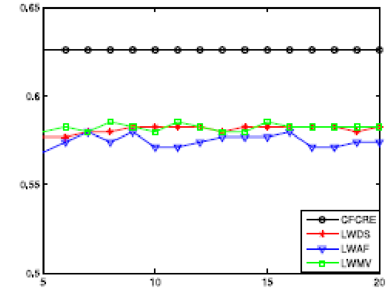
Comparisons with related methods



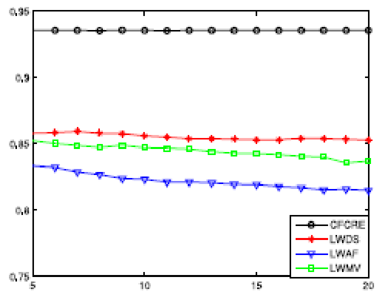
(a-1).SPECTF Heart with 4 classifiers.



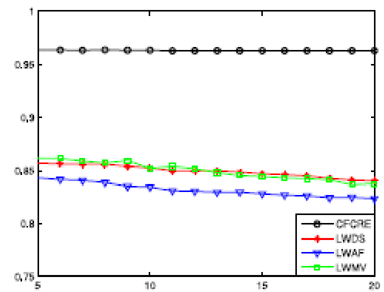
(a-2).SPECTF Heart with 11 classifiers.



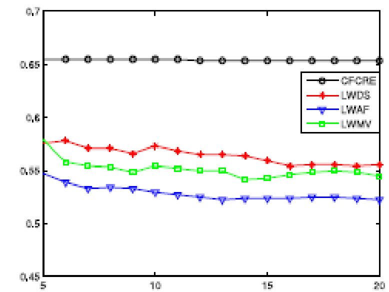
(b). Bupa with 3 classifiers.



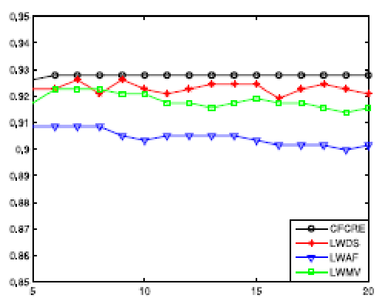
(d-1). Texture with 10 classifiers.



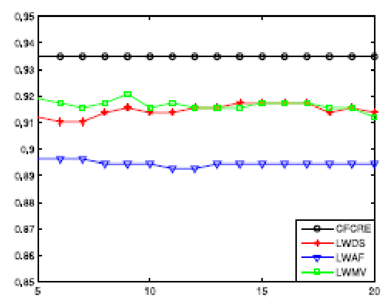
(d-2). Texture with 5 classifiers.



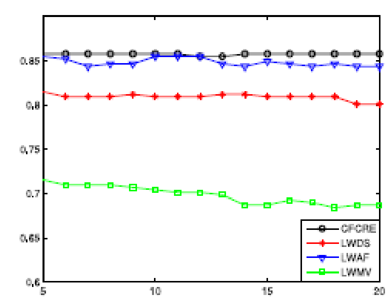
(e-1).Vehicle with 6 classifiers.



(f-1). Wbdc with 10 classifiers.

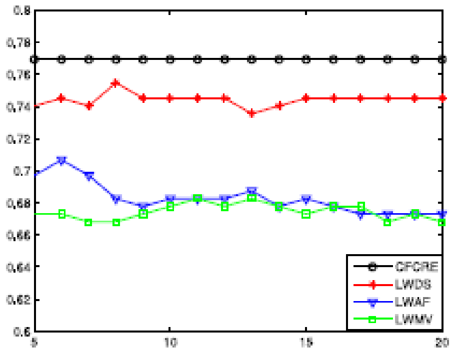


(f-2). Wbdc with 6 classifiers.

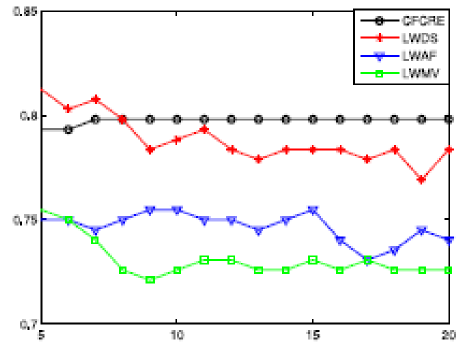


(g-1). Ionosphere with 4 classifiers.

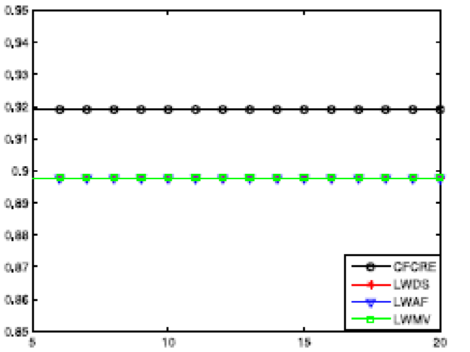
Comparisons with related methods



(j-1).Sonar with 30 classifiers.



(j-2).Sonar with 10 classifiers.



(k-1). Page-blocks with 2 classifiers.

- The new method produces the highest accuracy because of the consideration of local knowledge.
- The new method is very robust to the K value, which is convenient for the real applications.

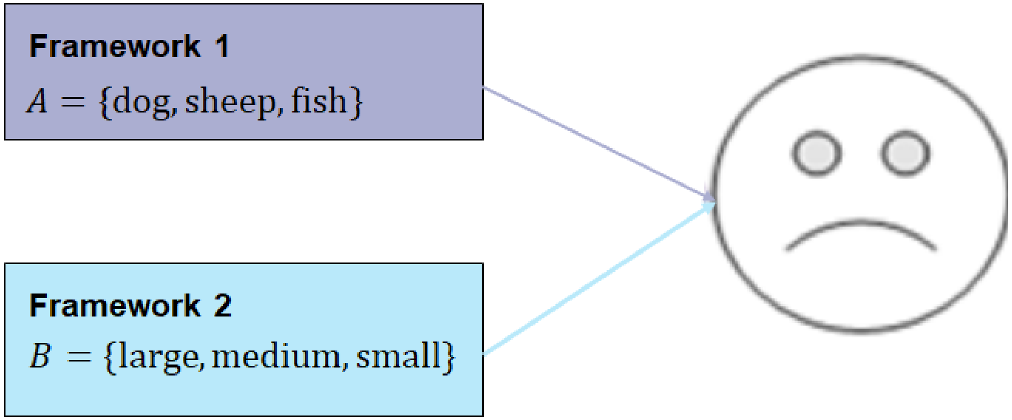
Outline

2.5 Evidential combination of classifiers with different frames of discernment

- Classifier fusion remains an effective method to improve classification performance.
- In applications, the classifiers learnt using different attributes may work with various frames of discernment (FoD) of classification. There generally exist more or less complementary knowledge among these classifiers.
- However, how to efficiently combine such classifiers under different FoD is a challenging problem.

2.5 Evidential combination of classifiers with different frames of discernment

- We propose a new method for classifier fusion with different FoD based on the belief functions (BF), which allow to well represent and deal with uncertain information.
- The credal transformation rules are developed to map the various FoD into a common one.
- It allows to transfer the belief of one class in the given FoD not only to several singleton classes but also to the meta-classes and the ignorance in other chosen.



2.5.1 Determination of credal transformation rules

- We want to obtain the classification result of one object *y* over a chosen FoD Ψ by the combination of these given classification results.
- The classifiers working with different FoD are not easy to combine directly, and we must transform the classification results represented by BBA m_n in Θ_N to the FoD Ψ before applying combination procedure.

2.5.1 Determination of credal transformation rules

- The two sources of evidence m^{Θ}, m^{Ψ} defined over the FoD $\Theta = \{\theta_1, \dots, \theta_p\}$ and $\Psi = \{\psi_1, \dots, \psi_p\}$.
- The transformation matrix $\Gamma_{p\times(q+2)}$ for mapping the evidence from Θ to Ψ is given by

$$\Gamma_{p\times(q+2)} = \begin{bmatrix} \gamma_{1,1} & \gamma_{1,2} & \cdots & \gamma_{1,q} & \gamma_{1,(q+1)} & \gamma_{1,(q+2)} \\ \gamma_{2,1} & \gamma_{2,2} & \cdots & \gamma_{2,q} & \gamma_{2,(q+1)} & \gamma_{2,(q+2)} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \gamma_{p,1} & \gamma_{p,2} & \cdots & \gamma_{p,q} & \gamma_{p,(q+1)} & \gamma_{p,(q+2)} \end{bmatrix}$$

subject to

$$\sum_{j=1}^{q+2} \gamma_{i,j} = 1$$

2.5.2 Estimation of transformation matrices

- The belief redistributed to the singleton class ψ_k is defined by

$$m^{\Psi}(\psi_k) = \sum_{i=1}^{p} \gamma_{i,k} m^{\Theta}(\theta_i)$$

- The belief discounted to the ignorant element Ψ is defined by

$$m^{\Psi}(\Psi) = \sum_{i=1}^{p} \gamma_{q+2} m^{\Theta}(\theta_i)$$

- The belief committed to the imprecise element A is defined by

$$m^{\Psi}(A) = \sum_{i=1}^{p} \delta_{\vartheta_i}, A\gamma_{i,(q+1)} m^{\Theta}(\vartheta_i), A \subset \Psi$$

with

$$\delta_{\vartheta_i,A} = \begin{cases} 1 & \vartheta_i \in A \\ 0 & \text{otherwise} \end{cases}$$

2.5.2 Estimation of transformation matrices

- For one pattern x_k, the evidence defined over different FoD Θ_n is transformed to the objective FoD Ψ in a similar way, and the transformed BBA $m_{n,k}^{\Psi}$ are combined in Ψ.

- The combination result of the N BBA $m_{n,k}^{\Psi}$ is converted into pignistic probability BetP(.)

$$\left\| \mathrm{Bet}P_{m_k^{\Psi}} - t_k \right\| = \epsilon$$

with

$$m_k^{\Psi} = \bigoplus_{n=1}^{N} m_k^{\Psi}$$

2.5.4 Experiments

Data	n	SMM	SSM	DTM	DRC-DT	DRC-SVM	CCDF
Iri	3	73.60±6.21	69.06±7.82	69.86±6.50	72.17±6.61	**83.46±5.38**	75.73±6.96
Win	3	81.23±4.76	79.77±4.36	78.87±4.45	74.14±5.15	72.13±6.34	**85.73±4.73**
Win	5	80.11±5.94	79.32±5.38	78.98±5.13	82.37±5.38	74.23±5.22	**84.38±5.46**
Vow5	3	71.73±5.11	71.91±4.83	71.86±5.10	68.45±7.34	69.67±5.47	**75.02±5.99**
Vow5	5	51.37±3.73	51.91±3.57	49.24±3.36	46.37±5.99	47.34±5.24	**54.06±5.89**
Sat	3	79.46±1.16	79.48±1.07	79.16±0.84	80.18±0.75	78.14±0.67	**80.72±0.86**
Rob	3	84.69±0.93	83.89±0.70	83.77±0.70	**89.45±0.82**	83.01±0.34	87.00±0.87
Veh	3	62.62±1.41	62.50±1.02	62.24±1.10	60.25±2.98	62.13±1.24	**63.49±2.88**
Veh	5	56.97±2.56	56.47±3.93	56.80±1.32	58.49±1.38	48.78±2.34	**60.37±1.26**
Ver	4	55.35±5.70	54.12±5.99	53.09±6.16	58.06±5.91	57.35±5.57	**59.16±5.97**
Ver	5	50.06±5.84	51.03±5.03	50.45±5.23	53.29±5.49	50.23±5.45	**55.16±5.91**
Bre4	3	73.07±7.47	73.33±6.96	73.07±9.30	67.13±4.26	71.24±6.37	**76.15±4.99**
Bre4	5	40.00±1.32	73.84±2.16	73.84±2.16	69.73±3.59	67.34±1.56	**77.43±1.35**
Gla	3	59.06±4.61	60.37±4.54	59.81±4.33	54.2±4.99	54.20±3.23	**61.40±4.13**
Gla5	5	60.39±1.46	59.90±2.84	60.39±2.81	58.94±3.28	56.23±1.45	**61.18±0.83**
Thy	3	84.90±5.60	85.09±4.48	84.44±4.59	85.36±5.22	**86.66±5.22**	84.81±6.05
Thy	5	83.61±5.05	81.48±4.93	77.87±5.17	82.54±4.87	74.34±2.15	**85.37±4.20**
Spe8	3	35.88±1.49	88.08±0.78	35.78±1.52	92.28±0.98	91.56±0.45	**92.79±0.79**
Wqr	3	50.62±0.77	50.01±0.66	50.50±0.69	**54.34±0.13**	49.98±0.37	53.25±1.08
Bre	3	64.71±0.91	63.39±1.82	63.77±1.48	64.13±1.25	64.12±0.49	**68.49±0.91**

The proposed method can efficiently improve the classification accuracy thanks to the transfer imprecision, which has been well characterized by estimating the transformation matrices.

Outline

1. Introduction

1.1 Tranditional classifier

1.2 Typical evidential classifier

2. Recent development

2.1 Evidential classification of incomplete patterns

2.2 Evidential transfer classification for heterogeneous data

2.3 Combination of Transferable Classification

2.4 Evidential classifier fusion with refined reliability evaluation

2.5 Evidential Combination of classifiers with different frames of discernment

3. Conclusion

3.Conclusion

- We present a classification method with the selective imputation of missing values based on belief functions.
- We develop an evidence-based weighted fusion method to combine transfer classification for heterogeneous data.
- We propose a method called combination of transferable classification to integrate the complementary information in multiple source domains using belief functions.
- We develop a method for revising the classifier output by a refined reliability evaluation to improve fusion robustness.
- We design a new method for classifier fusion with different FoD based on the belief functions to well represent and deal with uncertain information.

3.Conclusion

In the future,

- We will combine belief functions and deep learning to improve the ability of deep models to characterize uncertainty and imprecision.
- We will extend belief functions to multimodal data to improve the comprehensive fusion performance of multimodal data.
- We will carry out the applications of belief functions in unknown target open set recognition to extend the applicable scenarios of belief functions.

Thank you！

Questions？