Ensemble Enhanced Evidential *k*-NN classifier through random subspaces.

Asma Trabelsi[1,2], Zied Elouedi[1], Eric Lefevre[2]

[1] Université de Tunis, Institut Supérieur de Gestion de Tunis, Larodec
[2]Université d'Artois, Laboratoire de Génie Informatique et d'Automatique de l'Artois (LGI2A)

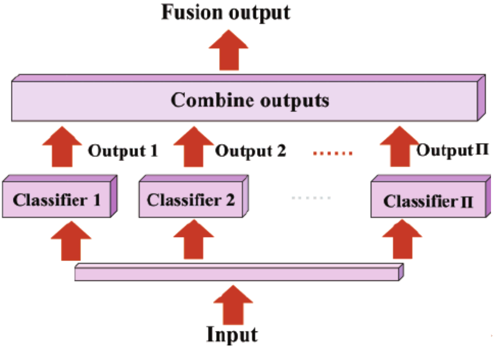
11 July 2017

Outline

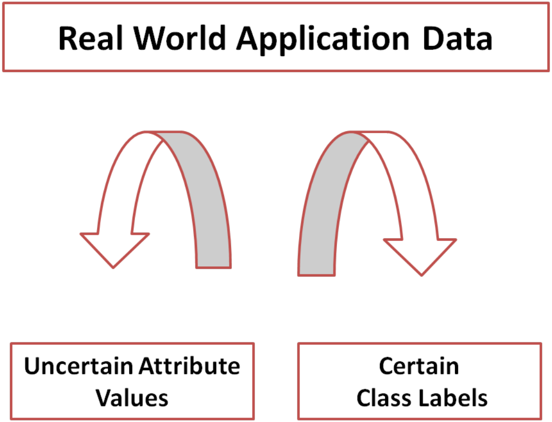
Outline

Introduction

Classifier fusion is regarded as an effective solution for solving several real world classification problems.



Problematic



Problematic

How to deal with uncertain attribute values for solving pattern recognition problems

Problematic

Enhanced Evidential *k*-Nearest Neighbors

Motivation

Ensemble Enhanced Evidential k-Nearest Neighbors

Outline

Evidence theory (1/2)

Evidence theory (1/2)

Frame of discernment

$$\Theta = \{\theta_1, \theta_2.....\theta_N\}$$

$$2^{\Theta} = \{A, A \subseteq \Theta\}$$

Evidence theory (1/2)

Frame of discernment

$$\Theta = \{\theta_1, \theta_2 .....\theta_N\}$$

$$2^\Theta = \{A, A \subseteq \Theta\}$$

Basic Belief Assignment (bba)

$$m : 2^\Theta \rightarrow [0, 1]$$

$$\sum_{A \subseteq \Theta} m(A) = 1$$

Evidence theory (1/2)

Frame of discernment

$$\Theta = \{\theta_1, \theta_2 \ldots\ldots \theta_N\}$$

$$2^\Theta = \{A, A \subseteq \Theta\}$$

Basic Belief Assignment (bba)

$$m : 2^\Theta \rightarrow [0, 1]$$

$$\sum_{A \subseteq \Theta} m(A) = 1$$

Combination rule

The Dempster rule allows to combine bbas provided by distinct pieces of evidence. It is set as $\forall\ A \subseteq \Theta$:

$$m_1 \oplus m_2(A) = \frac{1}{1-K} \sum_{B \cap C = A} m_1(B) m_2(C),$$

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

Evidence theory (1/2)

Frame of discernment

$$\Theta = \{\theta_1, \theta_2.....\theta_N\}$$

$$2^\Theta = \{A, A \subseteq \Theta\}$$

Basic Belief Assignment (bba)

$$m : 2^\Theta \rightarrow [0, 1]$$

$$\sum_{A \subseteq \Theta} m(A) = 1$$

Combination rule

The Dempster rule allows to combine bbas provided by distinct pieces of evidence. It is set as $\forall\ A \subseteq \Theta$:

$$m_1 \oplus m_2(A) = \frac{1}{1-K} \sum_{B \cap C = A} m_1(B) m_2(C),$$

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

Decision making

The TBM framework, which consists on two main levels (Credal level, Pignistic level), allows to make decision:

$$BetP(A) = \sum_{B \cap A = \emptyset} \frac{|A \cap B|}{|B|} m(B), \quad \forall A \in \Theta$$

Evidence theory (2/2)

Evidence theory (2/2)

Dissimilarity between bbas

The Jousselme distance between two pieces of evidence m_1 and m_2 is found as follows:

$$d(m_1, m_2) = \sqrt{\frac{1}{2}(\overrightarrow{m_1} - \overrightarrow{m_2})^T D(\overrightarrow{m_1} - \overrightarrow{m_2})}$$

Evidence theory (2/2)

Dissimilarity between bbas

The Jousselme distance between two pieces of evidence m_1 and m_2 is found as follows:

$$d(m_1, m_2) = \sqrt{\frac{1}{2}(\overrightarrow{m_1} - \overrightarrow{m_2})^T D(\overrightarrow{m_1} - \overrightarrow{m_2})}$$

- $\overrightarrow{m_1}$ and $\overrightarrow{m_2}$ are vector representations of m_1 and m_2

Evidence theory (2/2)

Dissimilarity between bbas

The Jousselme distance between two pieces of evidence m_1 and m_2 is found as follows:

$$d(m_1, m_2) = \sqrt{\frac{1}{2}(\overrightarrow{m_1} - \overrightarrow{m_2})^T D(\overrightarrow{m_1} - \overrightarrow{m_2})}$$

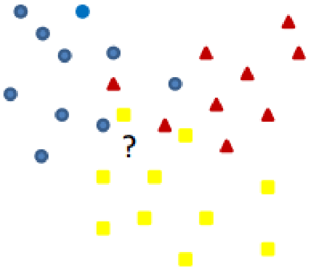
- $\overrightarrow{m_1}$ and $\overrightarrow{m_2}$ are vector representations of m_1 and m_2
- D is the Jaccard similarity measure defined by:

$$D(A, B) = \begin{cases} 1 & \text{if } A{=}B{=}\emptyset \\ \frac{|A \cap B|}{|A \cup B|} & \forall\, A,B \in 2^{\Theta} \end{cases}$$

Outline

Enhanced Evidential k Nearest Neighbors classifier

- Let $\Omega = \{w_1, \ldots, w_c\}$ denotes the set of classes.
- Each instance is described by:
  - Uncertain attribute values $x \in R^N$ represented within the belief function framework;
  - A certain class label $y \in \Omega$.
- Objective: given a learning set $L = \{(x_1, y_1), \ldots, (x_n, y_n)\}$, predict the class label of a new instance described by uncertain attribute values x using the k- Nearest Neighbors classifier.



Enhanced Evidential *k* Nearest Neighbors classifier

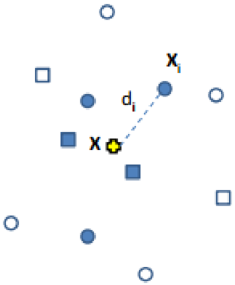
Example

Assume that our data are composed with five instances characterized by three uncertain attributes $x = \{Hair, Eye, Height\}$ and a certain class y with possible values $\{w_1,\ w_2\}$. The basic belief assignments, which are affected to the attribute values, will be defined on the frame of discernments $\Theta_{Hair} = \{Blond, Dark\}$, $\Theta_{Eye} = \{Brown, Blue\}$ and $\Theta_{Height} = \{Short, Middle, Tall\}$.

	Hair	*Eye*	*Height*	*d*
O_1	$m_1^{\Theta_{Hair}}(\{Dark\})$=0.5 $m_1^{\Theta_{Hair}}(\Theta_{Hair})$=0.5	$m_1^{\Theta_{Eye}}(\{Brown\})$=1 $m_1^{\Theta_{Eye}}(\Theta_{Eye})$=0	$m_1^{\Theta_{Height}}(\{Middle\})$=0.95 $m_1^{\Theta_{Height}}(\Theta_{Height})$=0.05	Ω_1
O_2	$m_2^{\Theta_{Hair}}(\{Blond\})$=0.1 $m_2^{\Theta_{Hair}}(\Theta_{Hair}) = 0.9$	$m_2^{\Theta_{Eye}}(\{Blue\})$=0.82 $m_2^{\Theta_{Eye}}(\Theta_{Eye})$=0.18	$m_2^{\Theta_{Height}}(\{Middle\})$=1 $m_2^{\Theta_{Height}}(\Theta_{Height})$=0	Ω_1
O_3	$m_3^{\Theta_{Hair}}(\{Blond\})$=0.6 $m_3^{\Theta_{Hair}}(\Theta_{Hair}) = 0.4$	$m_3^{\Theta_{Eye}}(\{Brown\})$=0.2 $m_3^{\Theta_{Eye}}(\Theta_{Eye})$=0.8	$m_3^{\Theta_{Height}}(\{Tall\})$=0.55 $m_3^{\Theta_{Height}}(\Theta_{Height})$=0.45	Ω_2
O_4	$m_4^{\Theta_{Hair}}(\{Dark\})$=0.7 $m_4^{\Theta_{Hair}}(\Theta_{Hair}) = 0.3$	$m_4^{\Theta_{Eye}}(\{Brwon\})$=0 $m_4^{\Theta_{Eye}}(\Theta_{Eye})$=1	$m_4^{\Theta_{Height}}(\{Short\})$=1 $m_4^{\Theta_{Height}}(\Theta_{Height})$=0	Ω_1
O_5	$m_5^{\Theta_{Hair}}(\{Blond\})$=1 $m_5^{\Theta_{Hair}}(\Theta_{Hair}) = 0$	$m_5^{\Theta_{Eye}}(\{Blue\})$=0.18 $m_5^{\Theta_{Eye}}(\Theta_{Eye})$=0.82	$m_5^{\Theta_{Height}}(\{Middle\})$=0.15 $m_5^{\Theta_{Height}}(\Theta_{Height})$=0.85	Ω_2

Enhanced Evidential k Nearest Neighbors classifier

- Let $N_k(x) \subset L$ denotes the set of the k nearest neighbors of x in L, based on the Jousselme distance measure.
- Each $x_i \in N_k(x)$ can be considered as a piece of evidence regarding the class of x.
- The strength of this evidence decreases with the distance d_i between x and x_i.



Enhanced Evidential k Nearest Neighbors classifier

Enhanced Evidential k Nearest Neighbors classifier

If $y_i = w_k$, the evidence of (x_i, y_i) can be represented by the simple mass function:

$$
\begin{aligned}
m_i(\{w_k\}) &= \varphi_k(d_i) \\
m_i(\{w_l\}) &= 0 \; \forall \; l \neq k \\
m_i(\Omega) &= 1 - \varphi_k(d_i)
\end{aligned}
$$

Enhanced Evidential k Nearest Neighbors classifier

If $y_i = w_k$, the evidence of (x_i, y_i) can be represented by the simple mass function:

$$m_i(\{w_k\}) = \varphi_k(d_i)$$
$$m_i(\{w_l\}) = 0 \ \forall \ l \neq k$$
$$m_i(\Omega) = 1 - \varphi_k(d_i)$$

d_i is calculated as the sum of Jousselme distances between the uncertain attribute values

Enhanced Evidential k Nearest Neighbors classifier

If $y_i = w_k$, the evidence of (x_i, y_i) can be represented by the simple mass function:

$$m_i(\{w_k\}) = \varphi_k(d_i)$$
$$m_i(\{w_l\}) = 0 \ \forall \ l \neq k$$
$$m_i(\Omega) = 1 - \varphi_k(d_i)$$

d_i is calculated as the sum of Jousselme distances between the uncertain attribute values

φ_k is a decreasing function from $[0, +\infty)$ to $[0, 1]$ such that $lim_{d \to +\infty}$ $\varphi_k(d) = 0$.

Enhanced Evidential *k* Nearest Neighbors classifier

If $y_i = w_k$, the evidence of (x_i, y_i) can be represented by the simple mass function:

$$m_i(\{w_k\}) = \varphi_k(d_i)$$
$$m_i(\{w_l\}) = 0 \; \forall \; l \neq k$$
$$m_i(\Omega) = 1 - \varphi_k(d_i)$$

d_i is calculated as the sum of Jousselme distances between the uncertain attribute values

φ_k is a decreasing function from $[0, +\infty)$ to $[0, 1]$ such that $lim_{d \to +\infty}$ $\varphi_k(d) = 0$.

Enhanced Evidential *k* Nearest Neighbors classifier

Enhanced Evidential k Nearest Neighbors classifier

Choice of the function φ_k: φ_k(d) = α $exp(-\gamma_k d^2)$.

Enhanced Evidential k Nearest Neighbors classifier

Choice of the function φ_k: $\varphi_k(\mathrm{d}) = \alpha\, exp(-\gamma_k d^2)$.

Parameters $\gamma_1, \ldots, \gamma_c$ can be optimized using:

- Exact method relying on a gradient search procedure for medium sized databases.
- A linearization method for large training sets.

Enhanced Evidential k Nearest Neighbors classifier

Choice of the function φ_k: φ_k(d) = α $exp(-\gamma_k d^2)$.

Parameters $\gamma_1, \ldots, \gamma_c$ can be optimized using:

- Exact method relying on a gradient search procedure for medium sized databases.
- A linearization method for large training sets.

α is a parameter such that $0 < \alpha < 1$.

Enhanced Evidential k Nearest Neighbors classifier

Choice of the function φ_k: $\varphi_k(\mathrm{d}) = \alpha \; exp(-\gamma_k d^2)$.

Parameters $\gamma_1, \ldots, \gamma_c$ can be optimized using:

- Exact method relying on a gradient search procedure for medium sized databases.
- A linearization method for large training sets.

α is a parameter such that $0 < \alpha < 1$.

The evidence of the k nearest neighbors of x is pooled using Dempster's rule of combination:

$$m = \oplus_{x_i \in N_k(x)} m_i$$

Outline

Ensemble Enhanced Evidential k Nearest Neighbors classifier

Ensemble Enhanced Evidential *k* Nearest Neighbors classifier

Diversity between classifiers is a substantial factor for achieving a good ensemble.

Ensemble Enhanced Evidential k Nearest Neighbors classifier

Diversity between classifiers is a substantial factor for achieving a good ensemble.

Diversity may be achieved by diversifying the input features

Ensemble Enhanced Evidential k Nearest Neighbors classifier

Diversity between classifiers is a substantial factor for achieving a good ensemble.

Diversity may be achieved by diversifying the input features

Ensemble Enhanced Evidential k-NN classifier through feature subspaces.

Ensemble Enhanced Evidential k Nearest Neighbors classifier

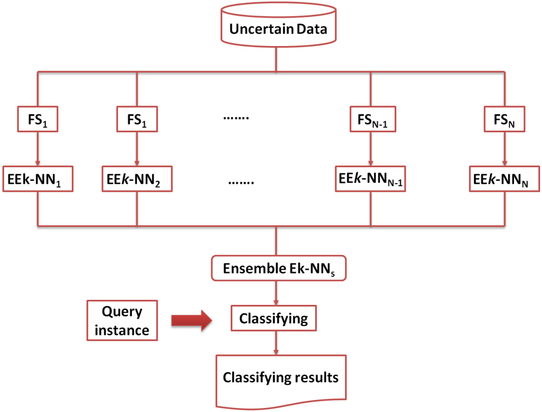
Diversity between classifiers is a substantial factor for achieving a good ensemble.

Diversity may be achieved by diversifying the input features

Ensemble Enhanced Evidential k-NN classifier through feature subspaces.

Generate feature subspaces using the Random Subspace Method.

Ensemble Enhanced Evidential k Nearest Neighbors classifier



Ensemble Enhanced Evidential k Nearest Neighbors classifier

Ensemble Enhanced Evidential k Nearest Neighbors classifier

Number of created classifiers

25 EEk-NNs classifiers are sufficient for reducing the error rate and for improving performance.

Ensemble Enhanced Evidential k Nearest Neighbors classifier

Number of created classifiers

25 EEk-NNs classifiers are sufficient for reducing the error rate and for improving performance.

Size of feature subsets

Randomly select the subspace size, relative to each individual EEk-NN classifier, in the range $[n/3;2n/3]$

Outline

1. Introduction
2. Evidence Theory
3. Enhanced Evidential k Nearest Neighbors classifier
4. Ensemble Enhanced Evidential k Nearest Neighbors classifier
5. **Experimentation**
6. Conclusions & Future works

Experimentation setups

Experimentation setups

Databases	Instances	Attributes	Classes
Voting records	435	16	2
Heart	267	22	2
Monks	195	23	2
Lymphography	148	18	4
Audiology	226	69	24

Experimentation setups

Databases	Instances	Attributes	Classes
Voting records	435	16	2
Heart	267	22	2
Monks	195	23	2
Lymphography	148	18	4
Audiology	226	69	24

All these databases do not contain uncertain condition attributes represented within the belief function framework.

Experimentation setups

Databases	Instances	Attributes	Classes
Voting records	435	16	2
Heart	267	22	2
Monks	195	23	2
Lymphography	148	18	4
Audiology	226	69	24

All these databases do not contain uncertain condition attributes represented within the belief function framework.

Generate synthetic databases

Generate synthetic databases by taking into account the original databases and a degree of uncertainty P to transform actual condition attribute value v_{A^k} of each object u_i, where $A^k \in A$, into a basic belief assignment:

$$m_i^{\Theta_k}(\{v_{A^k}\}) = 1 - P$$

$$m_i^{\Theta_k}(\Theta_k) = P$$

Experimentation setups

Databases	Instances	Attributes	Classes
Voting records	435	16	2
Heart	267	22	2
Monks	195	23	2
Lymphography	148	18	4
Audiology	226	69	24

All these databases do not contain uncertain condition attributes represented within the belief function framework.

Generate synthetic databases

Generate synthetic databases by taking into account the original databases and a degree of uncertainty P to transform actual condition attribute value v_{A^k} of each object u_i, where $A^k \in A$, into a basic belief assignment:

$$m_i^{\Theta_k}(\{v_{A^k}\}) = 1 - P$$

$$m_i^{\Theta_k}(\Theta_k) = P$$

The degree of uncertainty P takes value in the interval [0,1]:

- Certain Case (P=0)
- Low Uncertainty ($0 \leq P < 0.4$)
- Middle Uncertainty ($0.4 \leq P < 0.7$)
- High Uncertainty ($0.7 \leq P \leq 1$)

Experimentation results

Results for Heart database (%)

	$k = 1$		$k = 3$		$k = 5$		$k = 7$		$k = 9$	
	EEk-NN	Ensemble EEk-NN	EEk-NN	Ensemble EE$k-$NN	EEk-NN	Ensemble EEk-NN	EEk-NN	Ensemble EE$k-$NN	EE$k-$NN	Ensemble EE$k-$NN
No	61.15	**67.30**	63.84	**70.38**	67.30	**68.07**	70	**70.03**	71.15	**71.23**
Low	58.46	**68.84**	64.23	**66.15**	66.92	**69.23**	68.07	**68.07**	**79.03**	78.24
Middle	60	**69.23**	63.07	**65.38**	66.15	**67.69**	**69.61**	67.30	**68.07**	67.69
High	63.84	**68.46**	63.07	**65.76**	66.36	**66.53**	70.76	**71.13**	69.61	**70.03**

Results for Vote Records database (%)

	$k = 1$		$k = 3$		$k = 5$		$k = 7$		$k = 9$	
	EEk-NN	Ensemble EEk-NN	EEk-NN	Ensemble EE$k-$NN	EEk-NN	Ensemble EEk-NN	EEk-NN	Ensemble EE$k-$NN	EE$k-$NN	Ensemble EE$k-$NN
No	92.79	**92.05**	92.32	**92.65**	**93.02**	92.32	93.72	**94.01**	**93.72**	92.81
Low	92.09	**93.14**	93.02	**93.65**	92.55	**93.24**	93.25	**94.25**	93.25	**94.78**
Middle	91.62	**92.79**	91.39	**92.56**	91.39	**93.12**	91.86	**92.94**	92.32	**94.16**
High	84.18	**87.20**	87.67	**88.60**	88.60	**89.30**	**89.30**	86.97	89.76	**91.86**

Outline

Conclusions

Conclusions

- An ensemble EEk-NN classifier through random subspaces.

Conclusions

- An ensemble EE*k*-NN classifier through random subspaces.

- An ensemble EE*k*-NN classifier has outperformed the E*k*-NN that is learned in the full feature space.

Future works

Future works

- Solutions allowing to pick out the best feature subsets.

Future works

- Solutions allowing to pick out the best feature subsets.

- Compare an ensemble EE*k*-NN classifier through random subspaces with ensemble EE*k*-NN classifier learned through other feature subpace methods.

THANK YOU
FOR
your
ATTENTION!
ANY QUESTIONS?