Belief Functions and Rough Set Theory

Davide Ciucci

davide.ciucci@unimib.it
Dipartimento di Informatica, Sistemistica e Comunicazione
Università di Milano Bicocca

BFTA 2019

Plan

Part 1: introduction to rough set theory

- knowledge representation
- machine learning

Plan

Part 1: introduction to rough set theory

- knowledge representation
- machine learning

Part 2: belief functions and rough set theory

- knowledge representation
- machine learning

Outline

Outline

From approximations to Bel and Pl

- Let L, U be the lower and upper approximations induced by an equivalence relation on a universe X

From approximations to Bel and Pl

- Let L, U be the lower and upper approximations induced by an equivalence relation on a universe X
- For all $H \subseteq X$,

$$Bel(H) = \frac{|L(H)|}{|X|} \qquad Pl(H) = \frac{|U(H)|}{|X|}$$

are a pair of dual belief and plausibility functions

From approximations to Bel and Pl

- Let L, U be the lower and upper approximations induced by an equivalence relation on a universe X
- For all $H \subseteq X$,

$$Bel(H) = \frac{|L(H)|}{|X|} \qquad Pl(H) = \frac{|U(H)|}{|X|}$$

  are a pair of dual belief and plausibility functions
- The corresponding basic probability assignment is

$$m(H) = \begin{cases} \frac{|H|}{|X|} & H \text{ is an equivalence class} \\ 0 & \text{otherwise} \end{cases}$$

... and back

- Let *Bel* and *Pl* be a belief and plausibility functions satisfying the following two conditions:
  1. the set of focal elements forms a partition
  2. $m(H) = \frac{|H|}{|X|}$ for every focal set H

... and back

- Let Bel and Pl be a belief and plausibility functions satisfying the following two conditions:
  1. the set of focal elements forms a partition
  2. $m(H) = \frac{|H|}{|X|}$ for every focal set H
- Then, there exists an equivalence relation R such that the derived lower and upper approximations generate Bel and Pl

Focal Sets $\rightarrow$ Partition $\rightarrow$ Equivalence relation R

Interpretation

About the interpretation of this link [1]

> *Rough set theory is objective – for a given information table, qualities of corresponding approximations are computed. On the other hand, the Dempster-Shafer theory is subjective – it is assumed that values of belief (or plausibility) are given by an expert.*

[1] Pawlak, Z., Grzymala-Busse, J.W., Slowinski, R., Ziarko, W., Rough sets. Communication of the ACM 38, 88–95, 1995

Outline

- Knowledge representation
  - Pawlak rough sets
  - Generalized Relations
  - Approximation spaces
  - Uncertain Decision Tables
- Machine Learning
  - Clustering
  - Uncertainty in the conditions
  - Uncertainty in the decision

From relations to Bel and Pl

To any serial relation there corresponds a belief assignment such that the mass can assume only rational values

From relations to Bel and Pl

To any serial relation there corresponds a belief assignment such that the mass can assume only rational values

Proposition

Let R be at least a serial relation Then

- A pair of dual Belief and Plausibility functions can be defined as before: $Bel(H) = \frac{|L(H)|}{|X|}$ $\quad Pl(H) = \frac{|U(H)|}{|X|}$
- The basic probability assignment is defined as

$$m(H) = \frac{|\{x \in X : R(x) = H\}|}{|X|}$$

Y. Y. Yao, Pawan Lingras: Interpretation of Belief Functions in The Theory of Rough Sets. Inf. Sci. 104(1-2): 81-106 (1998)

... and back

- Let *Bel* and *Pl* be a belief and plausibility such that for all subsets $H \subseteq X$, $m(H)$ is a rational number with denominator $|X|$
- Then, there exists a serial relation R such that the derived lower and upper approximations generate *Bel* and *Pl*

... and back

- Let *Bel* and *Pl* be a belief and plausibility such that for all subsets $H \subseteq X$, $m(H)$ is a rational number with denominator $|X|$
- Then, there exists a serial relation R such that the derived lower and upper approximations generate *Bel* and *Pl*
- In order to consider any rational number, we have to take into account *interval rough sets*.

Outline

Some definitions

Let $(X, G(X))$ be an approximation space
Given an element $a \in X$, we define

- the approximated set of a: $\mathcal{G}(a) = \{A \subseteq X : a \in L(A)\}$

Some definitions

Let $(X, G(X))$ be an approximation space
Given an element $a \in X$, we define

- the approximated set of a: $\mathcal{G}(a) = \{A \subseteq X : a \in L(A)\}$
- the minimal approximated set of a:
  $\mathcal{M}_G(x) = \{A \in \mathcal{G}(a) : \forall B \in \mathcal{G}(a),\ B \subseteq A \Rightarrow B = A\}$

Some definitions

Let $(X, G(X))$ be an approximation space
Given an element $a \in X$, we define

- the approximated set of a: $\mathcal{G}(a) = \{A \subseteq X : a \in L(A)\}$
- the minimal approximated set of a:
  $\mathcal{M}_G(x) = \{A \in \mathcal{G}(a) : \forall B \in \mathcal{G}(a),\ B \subseteq A \Rightarrow B = A\}$
- Let $j(H) = \{x : H \in M_G(x)\}$

Result

Given a lower-upper approximation pair on (X, G), then we can

1. define a basic belief assignement on the universe X as:

$$m(H) = \begin{cases} 0 & j(H) = \emptyset \\ \frac{1}{|X|} \sum_{x \in j(H)} \frac{1}{|M_G(x)|} & \text{otherwise} \end{cases}$$

Result

Given a lower-upper approximation pair on (X, G), then we can

1. define a basic belief assignement on the universe X as:

$$m(H) = \begin{cases} 0 & j(H) = \emptyset \\ \frac{1}{|X|} \sum_{x \in j(H)} \frac{1}{|M_G(x)|} & \text{otherwise} \end{cases}$$

2. $\forall x$, $|M_G(x)| = 1$ iff $\forall H \subseteq X$, $Bel(H) = \frac{|L(H)|}{|X|}$ and $Pl(H) = \frac{|U(H)|}{|X|}$

Anhui Tan, Wei-Zhi Wu, Yuzhi Tao: A unified framework for characterizing rough sets with evidence theory in various approximation spaces. Inf. Sci. 454-455: 144-160 (2018)

Outline

Example

U	Hair	Eyes	Height	ud
o_1	Dark	Brown	Short	$m_1(\{ud_2\}) = 0.5, m_1(\Theta) = 0.5$
o_2	Blond	Blue	Middle	$m_2(\{ud_2\}) = 1$
o_3	Blond	Brown	Short	$m_3(\{ud_1\}) = 0.7, m_3(\Theta) = 0.3$
o_4	Blond	Brown	Tall	$m_4(\{ud_1\}) = 0.95, m_4(\{ud_2\}) = 0.05$
o_5	Dark	Brown	Short	$m_5(\{ud_2\}) = 1$
o_6	Blond	Blue	Middle	$m_6(\{ud_2\}) = 0.95, m_6(\Theta) = 0.05$
o_7	Dark	Brown	Tall	$m_7(\{ud_1\}) = 1$
o_8	Dark	Brown	Middle	$m_8(\{ud_1\}) = 0.975, m_8(\Theta) = 0.025$

Trabelsi, Elouedi, Lingras, "Classification systems based on rough sets under the belief function framework", IJAR 52, 1409–1432, 2011

Definition

An uncertain decision table (UDT) is a structure $(X, C \cup \{d\}, Val, F, m)$, where

- $(X, C \cup \{d\}, Val, F)$ is a decision table
- $m : Val_d \mapsto [0, 1]$ is a basic belief assignment on the decision values

Decision Classes - Example

U	Hair	Eyes	Height	ud
o_1	Dark	Brown	Short	$m_1(\{ud_2\}) = 0.5, m_1(\Theta) = 0.5$
o_2	Blond	Blue	Middle	$m_2(\{ud_2\}) = 1$
o_3	Blond	Brown	Short	$m_3(\{ud_1\}) = 0.7, m_3(\Theta) = 0.3$
o_4	Blond	Brown	Tall	$m_4(\{ud_1\}) = 0.95, m_4(\{ud_2\}) = 0.05$
o_5	Dark	Brown	Short	$m_5(\{ud_2\}) = 1$
o_6	Blond	Blue	Middle	$m_6(\{ud_2\}) = 0.95, m_6(\Theta) = 0.05$
o_7	Dark	Brown	Tall	$m_7(\{ud_1\}) = 1$
o_8	Dark	Brown	Middle	$m_8(\{ud_1\}) = 0.975, m_8(\Theta) = 0.025$

Decision Classes - Example

U	Hair	Eyes	Height	ud
o_1	Dark	Brown	Short	$m_1(\{ud_2\}) = 0.5, m_1(\Theta) = 0.5$
o_2	Blond	Blue	Middle	$m_2(\{ud_2\}) = 1$
o_3	Blond	Brown	Short	$m_3(\{ud_1\}) = 0.7, m_3(\Theta) = 0.3$
o_4	Blond	Brown	Tall	$m_4(\{ud_1\}) = 0.95, m_4(\{ud_2\}) = 0.05$
o_5	Dark	Brown	Short	$m_5(\{ud_2\}) = 1$
o_6	Blond	Blue	Middle	$m_6(\{ud_2\}) = 0.95, m_6(\Theta) = 0.05$
o_7	Dark	Brown	Tall	$m_7(\{ud_1\}) = 1$
o_8	Dark	Brown	Middle	$m_8(\{ud_1\}) = 0.975, m_8(\Theta) = 0.025$

- $X_1 = \{o_1, o_3, o_4, o_7, o_8\}$
- $X_2 = \{o_1, o_2, o_3, o_5, o_6\}$

Decision classes

Given a decision value $v_i \in Val_d$ and a threshold θ, then, the tolerance decision class X_i is defined as:

$$X_{v_i} = \{x_i : \text{dist}(m, m_j) < \theta\}$$

for a m such that $m(\{v_i\}) = 1$

Decision classes

Given a decision value $v_i \in Val_d$ and a threshold θ, then, the tolerance decision class X_i is defined as:

$$X_{v_i} = \{x_i : \text{dist}(m, m_j) < \theta\}$$

for a m such that $m(\{v_i\}) = 1$ the distance function is defined as

$$dist(m_1, m_2) = \sqrt{\frac{1}{2}(\|\overrightarrow{m_1}\|^2 + \|\overrightarrow{m_2}\|^2 - 2\langle\overrightarrow{m_1}, \overrightarrow{m_1}\rangle)}$$

with

$$\langle\overrightarrow{m_1}, \overrightarrow{m_1}\rangle = \sum_{i=1}^{|2^{Val_d}|} \sum_{j=1}^{|2^{Val_d}|} m_1(A_i)m_2(A_j)\frac{|A_i \cap A_j|}{|A_i \cup A_j|}$$

and $A_i, A_j \in 2^{Val_d}$. That is the sum is made over all the possible subsets of values that the decision attribute can assume.

The Approximations - Example

o_1, o_5 have the same features $\rightarrow$ aggregate their bbas

Object	$m(\{ud_1\})$	$m(\{ud_2\})$	$m(\Theta)$
o_1	0	0.5	0.5
o_5	0	1	0
$m_{1,5}$	0	0.75	0.25
o_2	0	1	0
o_6	0	0.95	0.05
$m_{2,6}$	0	0.975	0.025

The approximations (of decision classes)

1. Combine the bbas inside each equivalence class $[x]_C$ defined on the condition attributes to obtain $m_{[x]_C}$
2. the lower approximation for each decision class X_i is computed as

$$L(X_i) = \{x \in X : [x]_C \subseteq X_i \quad \text{and} \quad dist(m, m_{[x]_C}) \leq \theta\}$$

3. The upper approximation definition is not changed

The approximations (of decision classes)

1. Combine the bbas inside each equivalence class $[x]_C$ defined on the condition attributes to obtain $m_{[x]_C}$
2. the lower approximation for each decision class X_i is computed as

$$L(X_i) = \{x \in X : [x]_C \subseteq X_i \quad \text{and} \quad dist(m, m_{[x]_C}) \leq \theta\}$$

3. The upper approximation definition is not changed
4. In the example
$L(X1) = \{o_4, o_7, o_8\}$, $U(X1) = \{o_1, o_3, o_4, o_5, o_7, o_8\}$
$L(X2) = \{o_2, o_6\}$, $U(X2) = \{o_1, o_2, o_3, o_5, o_6\}$

Other connections: possibility theory

- Possibility functions can be seen as a special case of belief/plausibility functions (where the focal sets are nested)

Other connections: possibility theory

- Possibility functions can be seen as a special case of belief/plausibility functions (where the focal sets are nested)
- Some attempt to mix possibility theory with rough sets have also been pursued:
  - Nakata- Sakai group, last work:
    Michinori Nakata, Hiroshi Sakai: Rule Induction Based on Rough Sets from Possibilistic Data Tables. IUKM 2019, LNCS 11471, 86-97
  - Davide Ciucci, Ivan Forcati: Certainty-Based Rough Sets. IJCRS 2017, LNCS 10314, 43-55

Outline

Uncertainty in Machine Learning: a broader view

How many algorithms need data

Patient	Mitral rigurgitation	Acute dyspnea	Bicuspid aortic valve	EKG stress test
P1	No	Heart failure	Yes	Positive
P2	No	COPD	No	Negative
P3	Yes	Pneumonia	No	Positive
P4	Yes	Heart failure	Yes	Negative
P5	Yes	COPD	No	Positive
P6	yes	pneumonia	No	Positive

Uncertainty in Machine Learning: a broader view

How many algorithms need data

Patient	Mitral rigurgitation	Acute dyspnea	Bicuspid aortic valve	EKG stress test
P1	No	Heart failure	Yes	Positive
P2	No	COPD	No	Negative
P3	Yes	Pneumonia	No	Positive
P4	Yes	Heart failure	Yes	Negative
P5	Yes	COPD	No	Positive
P6	yes	pneumonia	No	Positive

Data is not a perfect representation of reality: it's a fundamentally human construct, and therefore subject to biases, limitations and other meaningful and consequential imperfections.
Andrea Jones-Rooy, NYU

Uncertainty in Machine Learning: a broader view

Patient	Mitral rigurgitation	Acute dyspnea	Bicuspid aortic valve	EKG stress test
P1	No	Heart failure (highly confident)	Yes (sure)	Positive (highly confident)
P2	Mild	COPD (highly confident)	No (confident)	Negative (highly confident)
P3	Moderate	Pneumonia (highly confident)	No (confident)	Not performed
P4	Severe	Heart failure (highly confident) or pneumonia (low confident)	Yes (highly confident)	Negative (confident)
P5	Mild or Moderate	COPD or pulmonary embolism or pneumonia	Not applicable	Not available
P6	Undetermined	Heart failure and pneumonia	Not applicable	Positive (confident)

- Several forms of uncertainty can exists in data

Uncertainty in Machine Learning: a broader view

Patient	Mitral rigurgitation	Acute dyspnea	Bicuspid aortic valve	EKG stress test
P1	No	Heart failure (highly confident)	Yes (sure)	Positive (highly confident)
P2	Mild	COPD (highly confident)	No (confident)	Negative (highly confident)
P3	Moderate	Pneumonia (highly confident)	No (confident)	Not performed
P4	Severe	Heart failure (highly confident) or pneumonia (low confident)	Yes (highly confident)	Negative (confident)
P5	Mild or Moderate	COPD or pulmonary embolism or pneumonia	Not applicable	Not available
P6	Undetermined	Heart failure and pneumonia	Not applicable	Positive (confident)

- Several forms of uncertainty can exists in data
- Do not hide them, manage them
- Do not always pretend a clear answer from algorithms.
  Better to say with certainty to a physician that someone is not ill or that there is a chance that he is ill?

Outline

Evidential Clustering

- Objects are assigned to cluster with uncertainty
- Each objects has associated a mass function m_i defined on the collection of clusters

Evidential Clustering

- Objects are assigned to cluster with uncertainty
- Each objects has associated a mass function m_i defined on the collection of clusters
- Rough clustering: all mass functions are logical, that is $m_i(A) = 1$ for some set of clusters A

Evidential Clustering

- Objects are assigned to cluster with uncertainty
- Each objects has associated a mass function m_i defined on the collection of clusters
- Rough clustering: all mass functions are logical, that is $m_i(A) = 1$ for some set of clusters A
- OPEN: what about three-way clustering?

Thierry Denoeux, Orakanya Kanjanatarakul: Beyond Fuzzy, Possibilistic and Rough: An Investigation of Belief Functions in Clustering. SMPS 2016: 157-164

Outline

Environment

- Information/Decision table with missing values or partially specified values
- Similarity relation

$$S_A = \{(x, y) \in X \times X : \forall a \in A,\ F(a, x) \cap F(a, y) \neq \emptyset\}$$

- Belief and plausibility functions

$$Bel(H) = \frac{|L(H)|}{|X|} \qquad Pl(H) = \frac{|U(H)|}{|X|}$$

Environment

- Information/Decision table with missing values or partially specified values
- Similarity relation

$$S_A = \{(x,y) \in X \times X : \forall a \in A,\ F(a,x) \cap F(a,y) \neq \emptyset\}$$

- Belief and plausibility functions

$$Bel(H) = \frac{|L(H)|}{|X|} \qquad Pl(H) = \frac{|U(H)|}{|X|}$$

- idea: new definition of reduct based on belief and plausibility functions

Wu, W., 2008. Attribute reduction based on evidence theory in incomplete decision systems. Inf. Sci. 178, 1355–1371

Belief Reduct in Information Tables

Given two attribute subsets $A \subseteq B$, A is a belief reduct of B if

1. For all similarity classes $[x]_S$, $Bel_A([x]_S) = Bel_B([x]_S)$
2. A minimality condition hold:
   If $C \subseteq B$ is such that $Bel_C([x]_S) = Bel_A([x]_S)$, then $C = B$

RESULT: H is a reduct iff it is a belief reduct

Plausibility Reduct in information tables

Given two attribute subsets $A \subseteq B$, A is a plausibility reduct of B if

1. For all similarity classes $[x]_R$, $Pl_A([x]_S) = Pl_B([x]_S)$
2. A minimality condition hold:
   If $C \subseteq B$ is such that $Pl_C([x]_S) = Pl_A([x]_S)$, then $C = B$

RESULT: If H is a reduct then it is a plausibility reduct

Reduct in decision tables

Two attribute subsets $A \subseteq B$, A is a relative belief reduct of B if

1. For all decision classes $[x]_d$, $Bel_A([x]_d) = Bel_B([x]_d)$
2. A minimality condition hold:
   If $C \subseteq B$ is such that $Bel_C([x]_d) = Bel_A([x]_d)$, then $B = C$.

Reduct in decision tables

Two attribute subsets $A \subseteq B$, A is a relative belief reduct of B if

1. For all decision classes $[x]_d$, $Bel_A([x]_d) = Bel_B([x]_d)$
2. A minimality condition hold:
   If $C \subseteq B$ is such that $Bel_C([x]_d) = Bel_A([x]_d)$, then $B = C$.

Two attribute subsets $A \subseteq B$, A is a plausibility reduct of B if

1. For all decision classes $[x]_d$, $Pl_A([x]_d) = Pl_B([x]_d)$
2. A minimality condition hold:
   If $C \subseteq B$ is such that $Pl_C([x]_d) = Pl_A([x]_d)$, then $B = C$

Results

- If the system is consistent: the standard relative reduct and the plausibility/belief reducts are equivalent

Results

- If the system is consistent: the standard relative reduct and the plausibility/belief reducts are equivalent
- If the system is not consistent: the standard reduct is equivalent to the plausibility reduct and different from the belief reduct

Results

- If the system is consistent: the standard relative reduct and the plausibility/belief reducts are equivalent
- If the system is not consistent: the standard reduct is equivalent to the plausibility reduct and different from the belief reduct
- OPEN how do the belief reduct behaves in applications?

Outline

Idea

- Based on Uncertain Decision Table (UDT)
- Reduct definition with respect to the positive region (unchanged)
- Based on the modified lower approximation

$$L(X_i) = \{x \in X : [x]_C \subseteq X_i \quad \text{and} \quad dist(m, m_{[x]_C}) \leq \theta\}$$

Idea

- Based on Uncertain Decision Table (UDT)
- Reduct definition with respect to the positive region (unchanged)
- Based on the modified lower approximation

$$L(X_i) = \{x \in X : [x]_C \subseteq X_i \quad \text{and} \quad dist(m, m_{[x]_C}) \leq \theta\}$$

- when two or more objects have the same conditions, their bbas are aggregated

Example

U	Hair	Eyes	Height	ud
o_1	Dark	Brown	Short	$m_1(\{ud_2\}) = 0.5, m_1(\Theta) = 0.5$
o_2	Blond	Blue	Middle	$m_2(\{ud_2\}) = 1$
o_3	Blond	Brown	Short	$m_3(\{ud_1\}) = 0.7, m_3(\Theta) = 0.3$
o_4	Blond	Brown	Tall	$m_4(\{ud_1\}) = 0.95, m_4(\{ud_2\}) = 0.05$
o_5	Dark	Brown	Short	$m_5(\{ud_2\}) = 1$
o_6	Blond	Blue	Middle	$m_6(\{ud_2\}) = 0.95, m_6(\Theta) = 0.05$
o_7	Dark	Brown	Tall	$m_7(\{ud_1\}) = 1$
o_8	Dark	Brown	Middle	$m_8(\{ud_1\}) = 0.975, m_8(\Theta) = 0.025$

- Reducts: {Hair, Height} and {Eyes, Height}
- Based on {Eyes, Height} we have the equivalence classes: $\{o_1, o_3, o_5\}$, $\{o_2, o_6\}$, $\{o_4, o_7\}$, $\{o_8\}$

Example

U	ud
o_1, o_3, o_5	$m_{1,3,5}(\{ud_1\}) = 0.24, m_{1,3,5}(\{ud_2\}) = 0.5, m_{1,3,5}(\Theta) = 0.26$
o_2, o_6	$m_{2,6}(\{ud_2\}) = 0.975, m_{2,6}(\Theta) = 0.025$
o_4, o_7	$m_{4,7}(\{ud_1\}) = 0.975, m_{4,7}(\{ud_2\}) = 0.025$
o_8	$m_8(\{ud_1\}) = 0.975, m_8(\Theta) = 0.025$

- Bbas have been combined

Example

U	ud
o_1, o_3, o_5	$m_{1,3,5}(\{ud_1\}) = 0.24, m_{1,3,5}(\{ud_2\}) = 0.5, m_{1,3,5}(\Theta) = 0.26$
o_2, o_6	$m_{2,6}(\{ud_2\}) = 0.975, m_{2,6}(\Theta) = 0.025$
o_4, o_7	$m_{4,7}(\{ud_1\}) = 0.975, m_{4,7}(\{ud_2\}) = 0.025$
o_8	$m_8(\{ud_1\}) = 0.975, m_8(\Theta) = 0.025$

- Bbas have been combined
- Rules: If Eyes = Brown and Height = Short then $m(ud1) = 0.24$, $m(ud2) = 0.5$, $m(\Theta) = 0.26$

Improvements

1. Use of an heuristic to simplify reducts' computation
2. Modified version of previous algorithms to work in parallel and suited to large datasets

1. Salsabil Trabelsi, Zied Elouedi, Pawan Lingras: Heuristic for Attribute Selection Using Belief Discernibility Matrix. RSKT 2012, LNCS 7414, 129-138
2. Salsabil Trabelsi, Zied Elouedi, Pawan Lingras: Belief Discernibility Matrix and Function for Incremental or Large Data. RSFDGrC 2013, LNCS 8170, 67-76

Conclusion

- Knowledge representation
  - Mathematical connections between approximations and belief function have been explored
  - The interpretation of these links is often missing
  - Some attempts on the combined use of rough sets and belief functions (Uncertain Decision Table)

Conclusion

- Knowledge representation
  - Mathematical connections between approximations and belief function have been explored
  - The interpretation of these links is often missing
  - Some attempts on the combined use of rough sets and belief functions (Uncertain Decision Table)
- Machine Learning
  - Connection between evidential and rough clustering
  - Combined use of the two theories
  - There are possibilities for improvements