From Rough Set Theory to Evidence Theory

Roman Słowiński

Laboratory of Intelligent Decision Support Systems
Institute of Computing Science
Poznań University of Technology

Introduction

- We show that rough set theory (RST) proposed by Pawlak 1982 can be treated as a basis for evidence theory proposed by Dempster and Shafer (1976), called Dempster-Shafer Theory (DST)
- Dempster-Shafer Theory annuls the axiom of additivity of exclusive events, which is one of traditional axioms of probabilistic reasoning: $P(A) + P(\neg A) = 1$
- This permits to take into account both a degree of belief in a fact and a degree of ignorance (belief in a negation of the fact) which do not need to sum up to **1** (in extreme case, degrees of belief and ignorance can be 0)
- In DST, increased support for a hypothesis does not affect the support for the complement of this hypothesis
- RST operates on a decision table $S=\langle U, C\cup\{d\}\rangle$ providing information about objects in terms of values of attributes; RST detects inconsistencies due to indiscernibility of objects belonging to different decision classes (hypotheses) $\boldsymbol{Cl} = \{Cl_1,Cl_2,\dots,Cl_n\}$; in consequence, the classes are characterized by lower and upper approximations

Introduction

- DST operates on a frame of discernment being a finite set of names of decision classes (hypotheses) $\boldsymbol{Cl} = \{Cl_1, Cl_2, \ldots, Cl_n\}$; all possible clusters of classes from $\boldsymbol{Cl}$ are considered – they create power set $2^{\boldsymbol{Cl}}$; information about a set of objects from U creating a given cluster is provided directly by some numerical functions called the basic probability assignment (gęstość prawdopodobińtwa, masy), belief function and plausibility function (dolna i górna granica stopnia przekonania)
- Clusters from $2^{\boldsymbol{Cl}}$ are all possible homes of objects from U (hypotheses), given some pieces of evidence (premises): e.g. for $\boldsymbol{Cl}$ = {*ravens*, *cats*, *whales*}, the home (hypothesis) for objects with legs is the cluster {*ravens*, *cats*}, while the home for objects being mammals is the cluster {*cats*, *whales*}
- While in DST the membership of objects in clusters is described by belief and plausibility being functions of bpa, we will show that the values of these functions can be calculated from the decision table using the RST concepts of lower and upper approximations of decision classes

Evidence Theory or Dempster-Shafer Theory

- Frame of discernment : $\boldsymbol{Cl} = \{Cl_1, Cl_2, \ldots, Cl_n\}$
- Basic probablitity assignment (bpa) m on $\boldsymbol{Cl}$: $m: 2^{\boldsymbol{Cl}} \rightarrow \mathbf{R}_+$ satisfying
  - $m(\emptyset) = 0$
  - $\sum_{\Delta \subseteq \boldsymbol{Cl}} m(\Delta) = 1$, where Δ is a cluster of decision classes (from $2^{\boldsymbol{Cl}}$)
- For a given bpa m, two functions are defined:
  - belief function over $\boldsymbol{Cl}$: $Bel: 2^{\boldsymbol{Cl}} \rightarrow \mathbf{R}_+$ iff for any $\theta \subseteq \boldsymbol{Cl}$

$$Bel(\theta) = \sum_{\Delta \subseteq \theta} m(\Delta)$$

  - plausibility function over $\boldsymbol{Cl}$: $Pl: 2^{\boldsymbol{Cl}} \rightarrow \mathbf{R}_+$ iff for any $\theta \subseteq \boldsymbol{Cl}$

$$Pl(\theta) = 1 - Bel(\boldsymbol{Cl} - \theta) = \sum_{\Delta \subseteq \boldsymbol{Cl}} m(\Delta) - \sum_{\Delta \subseteq \boldsymbol{Cl} - \theta} m(\Delta) = \sum_{\Delta \cap \theta \neq \emptyset} m(\Delta)$$

- Some properties:

$$Bel(\emptyset) = Pl(\emptyset) = 0, \quad Bel(\boldsymbol{Cl}) = Pl(\boldsymbol{Cl}) = 1$$

$$\text{if } \Delta \subseteq \theta, \text{ then } Bel(\Delta) \leq Bel(\theta) \text{ and } Pl(\Delta) \leq Pl(\theta)$$

$$Bel(\theta) \leq Pl(\theta)$$

$$Bel(\theta) + Bel(\boldsymbol{Cl} - \theta) \leq 1 \quad \text{and} \quad Pl(\theta) + Pl(\boldsymbol{Cl} - \theta) \geq 1$$

Evidence Theory or Dempster-Shafer Theory

- From a given belief function, a basic probablitity assignment (bpa) can be reconstructed :

$$m(\theta) = \sum_{\Delta \subseteq \theta} (-1)^{card(\theta - \Delta)} Bel(\Delta)\,, \quad \text{for } \theta \subseteq \boldsymbol{\mathit{CI}}$$

- The union of all subsets $\theta \subseteq$ ***CI*** that are focal (i.e. have the property $m(\theta)>0$) is called the core of ***CI***

Evidence Theory or Dempster-Shafer Theory

- Dempster's rule of combination

Suppose that 2 different belief functions Bel_1 and Bel_2 over the same frame of discernment ***Cl*** represent different pieces of evidence (przesłanki) (e.g. „with legs" or „mammals").

The assumption is that both pieces of evidence are independent.

As a result of Dempster's rule of combination a new belief function Bel_3 is computed as their orthogonal sum $Bel_1 \oplus Bel_2$:

$$m_3(\theta) = \frac{\sum_{\theta_1 \cap \theta_2 = \theta} m_1(\theta_1) m_2(\theta_2)}{1 - \sum_{\theta_1 \cap \theta_2 = \varnothing} m_1(\theta_1) m_2(\theta_2)}$$

where θ_1, θ_2 are focal elements of m_1 and m_2, respectively.

If $\sum_{\theta_1 \cap \theta_2 = \varnothing} m_1(\theta_1) m_2(\theta_2) = 1$, then $m_3(\theta)$ cannot be defined, and m_1, m_2 are said to be contradictory bpa.

Relationship between RST and DST

- Let ***Cl*** be the frame of discernment compatible with the decision table $S=\langle U, C\cup\{d\}\rangle$, let also $T(\theta)=\{t : Cl_t\in \boldsymbol{Cl}\}$, where $\theta\subseteq \boldsymbol{Cl}$
- For any $\theta\subseteq \boldsymbol{Cl}$ the belief function can be calculated as:

$$Bel_S(\theta) = \sum_{\Delta\subseteq\theta} m_S(\Delta) = \sum_{\Delta\subseteq\theta} \frac{card(\{x\in U : \delta_S(x)=\Delta\})}{card(U)} =$$

$$= \sum_{t\in T(\theta)} \frac{card(\underline{C}(Cl_t))}{card(U)} + \sum_{\substack{\Delta\subseteq\theta \\ card(\Delta)>1}} \frac{card(\{x\in U : \delta_S(x)=\Delta\})}{card(U)} = \frac{card\left(\underline{C}\left(\bigcup_{t\in T(\theta)} Cl_t\right)\right)}{card(U)}$$

where $\delta_S(x)=\{Cl_t : \exists y\in U, y\in I_C(x) \text{ and } y\in Cl_t, t\in T(\boldsymbol{Cl})\}$

is called generalized decision for object x (cluster of classes, with no possibility of discernment using knowledge about $S=\langle U, C\cup\{d\}\rangle$)

Relationship between RST and DST

- For any $\theta \subseteq \boldsymbol{Cl}$ the plausibility function can be calculated as:

$$Pl_S(\theta) = 1 - Bel_S(\boldsymbol{Cl} - \theta) = 1 - \frac{card\left(\underline{C}\left(\bigcup_{t \in T(\boldsymbol{Cl} - \theta)} Cl_t\right)\right)}{card(U)} =$$

$$= \frac{card(U) - card\left(\underline{C}\left(\bigcup_{t \in T(\boldsymbol{Cl} - \theta)} Cl_t\right)\right)}{card(U)} = \frac{card\left(\overline{C}\left(\bigcup_{t \in T(\theta)} Cl_t\right)\right)}{card(U)}$$

Example

- Let us consider decision table $S = \langle U, C\cup\{d\}\rangle$, where $U = \{1,\ldots,28\}$, $C=\{a, b, c, e, f\}$ and decision d makes partition of U into decision classes $\mathbf{Cl} = \{Cl_1, Cl_2, Cl_3\}$, $T(\mathbf{Cl}) = \{1, 2, 3\}$

U	a	b	c	e	f	d
1	0	0	1	0	0	1
2	1	1	1	0	0	1
3	0	1	0	0	0	2
4	1	0	0	0	1	1
5	0	1	0	0	0	2
6	1	0	0	0	1	3
7	0	0	0	1	1	3
8	1	1	1	1	1	1
9	0	0	0	1	1	3
10	0	0	1	0	0	1
11	1	1	1	0	0	2
12	1	1	1	1	1	2
13	1	1	0	1	1	3
14	1	1	0	0	1	3

U	a	b	c	e	f	d
15	0	0	1	1	1	1
16	1	1	0	1	1	2
17	0	0	0	1	1	3
18	0	0	0	0	0	2
19	0	0	0	0	0	2
20	1	1	1	0	0	3
21	1	1	0	0	1	2
22	0	0	1	0	1	1
23	1	1	1	0	0	2
24	0	0	1	1	1	2
25	1	0	1	0	1	1
26	1	0	1	0	1	3
27	1	0	1	0	1	2
28	1	1	1	1	0	3

$Cl_1=\{1,2,4,8,10,15,22,25\}$

$Cl_2=\{3,5,11,12,16,18,19,21,23,24,27\}$

$Cl_3=\{6,7,9,13,14,17,20,26,28\}$

$I_C(1)=\{1,10\}$, $I_C(2)=\{2,11,20,23\}$,

$I_C(3)=\{3,5\}$, $I_C(4)=\{4,6\}$,

$I_C(7)=\{7,9,17\}$, $I_C(8)=\{8,12\}$,

$I_C(13)=\{13,16\}$, $I_C(14)=\{14,21\}$,

$I_C(15)=\{15,24\}$, $I_C(18)=\{18,19\}$,

$I_C(22)=\{22\}$, $I_C(25=\{25,26,27\}$,

$I_C(28)=\{28\}$

Example

- Lower and upper approximations of decision classes:

$$\underline{C}(Cl_1) = I_C(1) \cup I_C(22) = \{1,10,22\}$$
$$\underline{C}(Cl_2) = I_C(3) \cup I_C(18) = \{3,5,18,19\}$$
$$\underline{C}(Cl_3) = I_C(7) \cup I_C(28) = \{7,9,17,28\}$$

$$\overline{C}(Cl_1) = \underline{C}(Cl_1) \cup I_C(2) \cup I_C(4) \cup I_C(8) \cup I_C(15) \cup I_C(25) =$$
$$= \{1,2,4,6,8,10,11,12,15,20,22,23,24,25,26,27\}$$
$$\overline{C}(Cl_2) = \underline{C}(Cl_2) \cup I_C(8) \cup I_C(13) \cup I_C(14) \cup I_C(15) \cup I_C(25) =$$
$$= \{2,3,5,8,11,12,13,14,15,16,18,19,20,21,23,24,25,26,27\}$$
$$\overline{C}(Cl_3) = \underline{C}(Cl_3) \cup I_C(2) \cup I_C(4) \cup I_C(13) \cup I_C(14) \cup I_C(25) =$$
$$= \{2,4,6,7,9,11,13,14,16,17,20,21,23,25,26,27,28\}$$

$$Bn_C(Cl_1) = \overline{C}(Cl_1) - \underline{C}(Cl_1) = \{2,4,6,8,11,12,15,20,23,24,25,26,27\}$$
$$Bn_C(Cl_2) = \overline{C}(Cl_2) - \underline{C}(Cl_2) = \{2,8,11,12,13,14,15,16,20,21,23,24,25,26,27\}$$
$$Bn_C(Cl_3) = \overline{C}(Cl_3) - \underline{C}(Cl_3) = \{2,4,6,11,13,14,16,20,21,23,25,26,27\}$$

Example

- Generalized decisions for objects from the boundaries of decision classes

$$Bn_C(Cl_1) = \overline{C}(Cl_1) - \underline{C}(Cl_1) = \{2,4,6,8,11,12,15,20,23,24,25,26,27\}$$

$$Bn_C(Cl_2) = \overline{C}(Cl_2) - \underline{C}(Cl_2) = \{2,8,11,12,13,14,15,16,20,21,23,24,25,26,27\}$$

$$Bn_C(Cl_3) = \overline{C}(Cl_3) - \underline{C}(Cl_3) = \{2,4,6,11,13,14,16,20,21,23,25,26,27\}$$

$$\delta(2) = \{1,2,3\},\ \ \delta(4) = \{1,3\},\ \ \delta(6) = \{1,3\},\ \ \delta(8) = \{1,2\},\ \ \delta(11) = \{1,2,3\},\ \ \delta(4) = \{1,2\},$$

$$\delta(13) = \{2,3\},\ \ \delta(14) = \{2,3\},\ \ \delta(15) = \{1,2\},\ \ \delta(16) = \{2,3\},\ \ \delta(20) = \{1,2,3\},\ \ \delta(21) = \{2,3\},$$

$$\delta(23) = \{1,2,3\},\ \ \delta(24) = \{1,2\},\ \ \delta(25) = \{1,2,3\},\ \ \delta(26) = \{1,2,3\},\ \ \delta(27) = \{1,2,3\}$$

Example

- Generalized decisions for objects from the boundaries of decision classes

U	a	b	c	e	f	δ_C
1	0	0	1	0	0	{1}
2	1	1	1	0	0	{1, 2, 3}
3	0	1	0	0	0	{2}
4	1	0	0	0	1	{1, 3}
5	0	1	0	0	0	{2}
6	1	0	0	0	1	{1, 3}
7	0	0	0	1	1	{3}
8	1	1	1	1	1	{1, 2}
9	0	0	0	1	1	{3}
10	0	0	1	0	0	{1}
11	1	1	1	0	0	{1, 2, 3}
12	1	1	1	1	1	{1, 2}
13	1	1	0	1	1	{2, 3}
14	1	1	0	0	1	{2, 3}

U	a	b	c	e	f	δ_C
15	0	0	1	1	1	{1, 2}
16	1	1	0	1	1	{2, 3}
17	0	0	0	1	1	{3}
18	0	0	0	0	0	{2}
19	0	0	0	0	0	{2}
20	1	1	1	0	0	{1, 2, 3}
21	1	1	0	0	1	{2, 3}
22	0	0	1	0	1	{1}
23	1	1	1	0	0	{1, 2, 3}
24	0	0	1	1	1	{1, 2}
25	1	0	1	0	1	{1, 2, 3}
26	1	0	1	0	1	{1, 2, 3}
27	1	0	1	0	1	{1, 2, 3}
28	1	1	1	1	0	{3}

Example

- Values of the basic probability assignment for the frame of discernement
  $\boldsymbol{Cl} = \{Cl_1, Cl_2, Cl_3\}$

θ	$\{Cl_1\}$	$\{Cl_2\}$	$\{Cl_3\}$	$\{Cl_1,Cl_2\}$	$\{Cl_1,Cl_3\}$	$\{Cl_2,Cl_3\}$	$\{Cl_1,Cl_2,Cl_3\}$
$T(\theta)$	$\{1\}$	$\{2\}$	$\{3\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$	$\{1,2,3\}$
$m_S(\theta)$	$\frac{3}{28}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{14}$	$\frac{1}{7}$	$\frac{1}{4}$

sum up to 1

Example

- Values of belief function for the frame of discernement $\boldsymbol{Cl} = \{Cl_1, Cl_2, Cl_3\}$,

  e.g. $Bel_S(\{2,3\} = m_S(\{2\})+m_S(\{2\})+m_S(\{2,3\}) = 1/7+1/7+1/7=12/28=3/7$, i.e. 12 out of 28 objects are in $Cl_2 \cup Cl_3$, among them 4 are in Cl_2, 4 are in Cl_3 and 4 are in Cl_2 or Cl_3 (with no possibility of discernment to which one).

θ	$\{Cl_1\}$	$\{Cl_2\}$	$\{Cl_3\}$	$\{Cl_1,Cl_2\}$	$\{Cl_1,Cl_3\}$	$\{Cl_2,Cl_3\}$	$\{Cl_1,Cl_2,Cl_3\}$
$T(\theta)$	$\{1\}$	$\{2\}$	$\{3\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$	$\{1,2,3\}$
$Bel_S(\theta)$	$\frac{3}{28}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{11}{28}$	$\frac{9}{28}$	$\frac{3}{7}$	1

$Pl_S(\theta)$ = 16/28, 19/28, 17/28, 6/7, 6/7, 25/28, 1

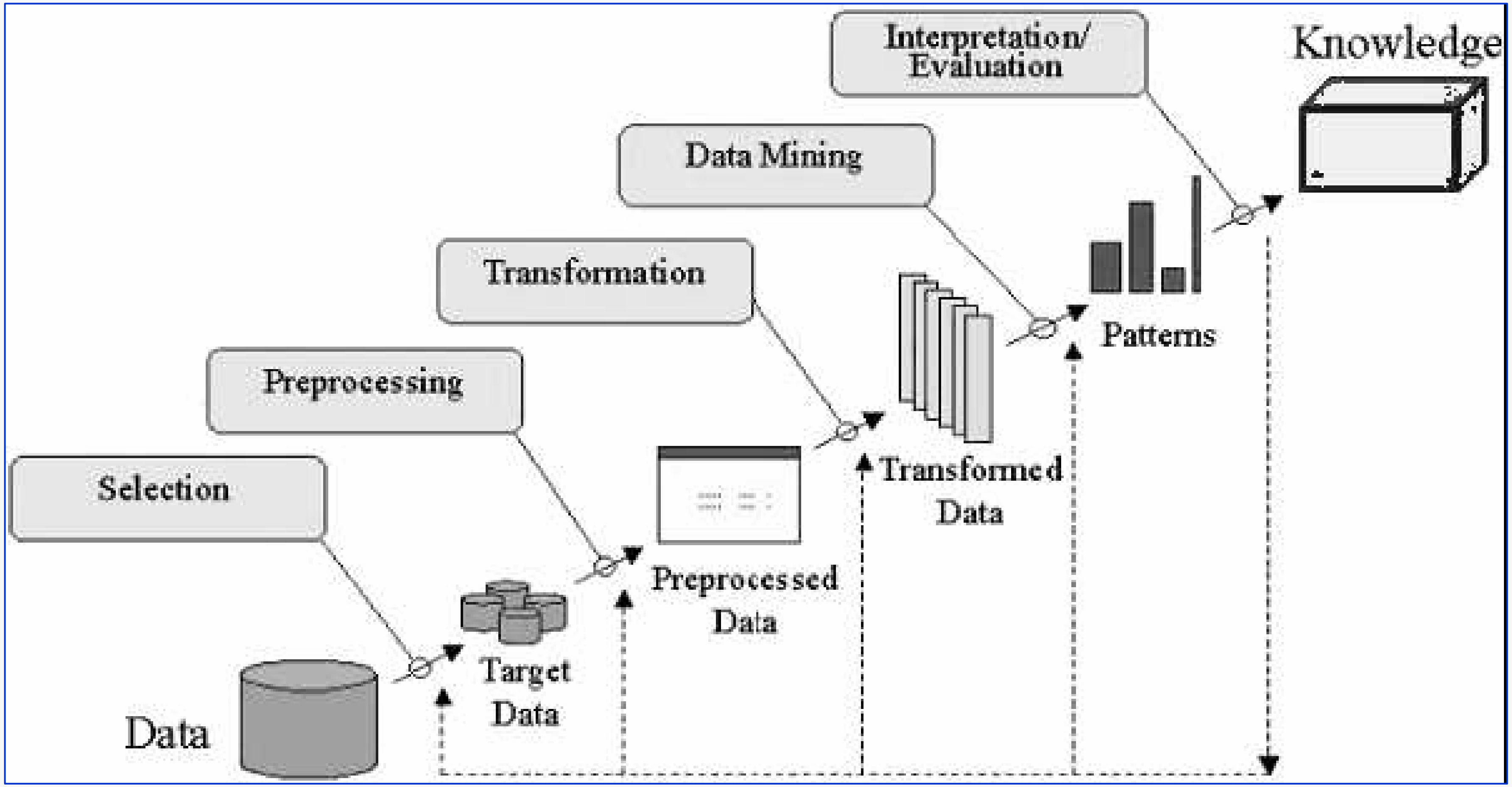
since $Pl_S(\theta) = 1 - Bel(\boldsymbol{Cl} - \theta)$, for $\theta \subseteq \boldsymbol{Cl}$

Final remarks about RST and Knowledge Discovery

Knowledge discovery principle

- Knowledge discovery (KD) is a type of machine learning which enables the extraction of useful information from a set of raw data in a high-level format, which a user can understand.
- The process of KD generally involves developing models (patterns) that describe or classify a set of measurements.
- Some times, the objective of KD is to study the model itself, while other times it is to construct a good classier.
- This process consists of various steps which are executed in a "waterfall" type fashion.
- One important step is the data mining step, whereby patterns and relationships are found in the data.

Knowledge discovery „waterfall”



Knowledge discovery steps

- Selection: This task involves setting the data set into a format that is suitable for the discovery task. This may involve joining together several existing sets of data in order to come to the final data set.
- Preprocessing: This step involves cleaning the data and removing information that is deemed unecessary or filling any missing values in the data.
- Transformation: The data set will be quite large and contain non relevant features or duplicate examples which must be merged. This will reduce the amount of data and also the time taken to execute mining queries.
- Mining: This step involves extracting patterns from the data.
- Evalution: Patterns identified by the system are interpreted into knowledge which can be used to support human-decision making or other tasks.

Rough set theory as a framework for knowledge discovery

- Selection: The data representation for rough sets are flat, two-dimensional data tables.
- Preprocessing: If the data contains missing values, the table may be processed either classically, after completing the data table in some way, or using a new indiscernibility relation able to compare incomplete descriptions.
- Transformation: Data discretization is often performed on numerical attributes. This involves converting the exact observations into intervals or ranges of values. This amounts to defining a coarser view of the world and also results in a reduction on the value set size for the observations.
- Mining: In the rough set framework, *if-then* rules are mined in one of three perspectives: minimal cover (minimal number of rules), exhaustive description (all rules), satisfactory description (interesting rules).
- Evalution: Individual patterns or rules can be measured or manually inspected. Rule sets can be used to classify new objects and their classicatory performance may be assessed.