Rough Sets and Fuzzy Rough Sets: Models and Applications

Chris Cornelis

Department of Applied Mathematics and Computer Science,
Ghent University, Belgium

Introduction

Lotfi Zadeh
(Baku, Feb. 4, 1921)

Zdzisław Pawlak
(Łodz, 1926–
Warsaw, 2006)

Introduction

- Fuzzy Sets (1965)
- Designed for dealing with **gradual** information

- Rough Sets (1982)
- Designed for dealing with **incomplete** information

Introduction

- Fuzzy Rough Sets (1990)
- Didier Dubois & Henri Prade

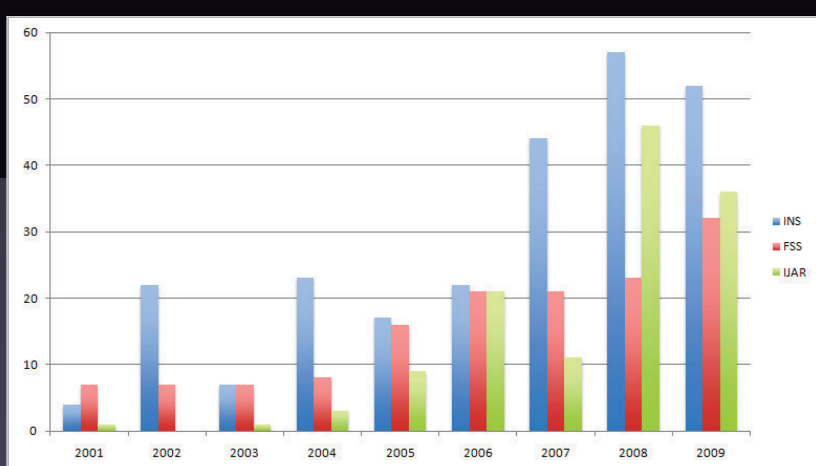
Introduction

International Rough Set Society

- http://www.roughsets.org
- Rough Set Database System (RSDS): 3882 publications (941 in journals, 2187 in proceedings)
- International conferences
  - **RSCTC**: Rough Sets and Current Trends in Computing
    Japan (2006), USA (2008), Poland (2010)
  - **RSKT**: Rough Sets and Knowledge Technology
    China (2008), Australia (2009), China (2010)
  - **RSFDGrC**: Rough Sets, Fuzzy Sets, Data mining and Granular Computing
    Canada (2005,2007), India (2009)
- TRS: Transactions on Rough Sets (LNCS, Springer)

Introduction

Rough set publications in Information Sciences, Fuzzy Sets and Systems and Int. Journal of Approximate Reasoning



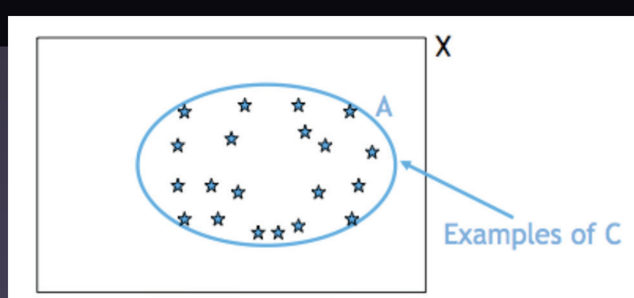
Overview

- Introduction
- Rough Sets (RS)
  - Pawlak's model and generalizations
  - Application: feature selection
- Fuzzy Rough Sets (FRS)
  - Implication/t-norm based model
  - Vaguely quantified rough set model
  - Applications in data analysis
  - Software
- Conclusion

Rough set theory

Goal: to approximate a concept C using

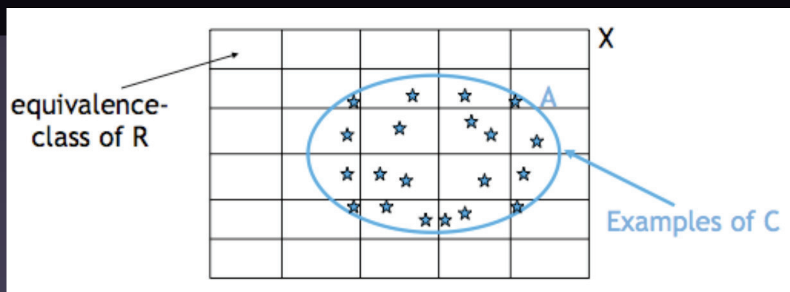
1. a set $A \subseteq X$ of examples of C



Rough set theory

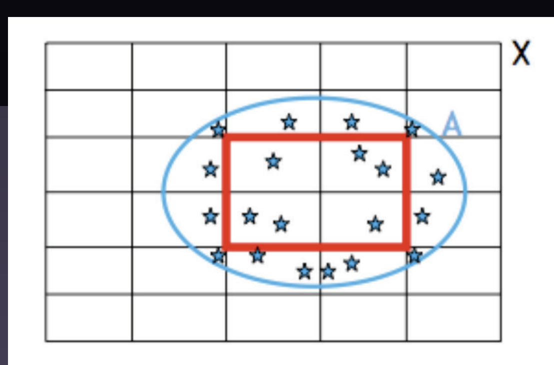
Goal: to approximate a concept C using

1. a set $A \subseteq X$ of examples of C
2. an equivalence relation R in X



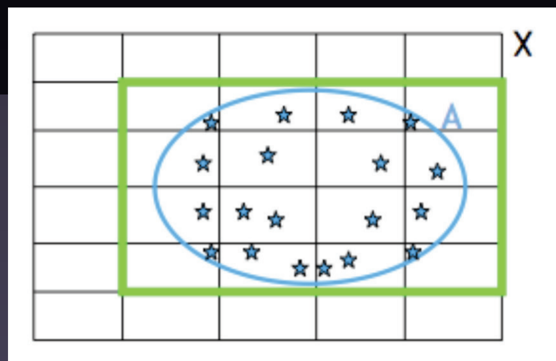
Lower Approximation

$$y \in R{\downarrow}A \Leftrightarrow [y]_R \subseteq A$$



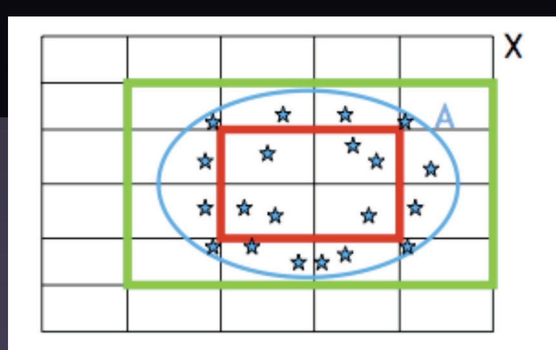
Upper Approximation

$$y \in R{\uparrow}A \Leftrightarrow [y]_R \cap A \neq \emptyset$$



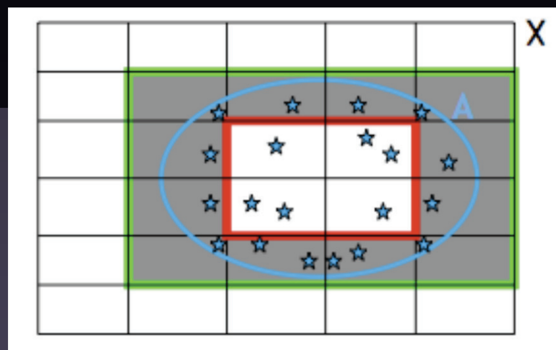
Rough Set $(R{\downarrow}A, R{\uparrow}A)$

$$y \in R{\downarrow}A \quad \Leftrightarrow \quad [y]_R \subseteq A$$
$$y \in R{\uparrow}A \quad \Leftrightarrow \quad [y]_R \cap A \neq \emptyset$$



Boundary region

$$y \in R{\downarrow}A \quad \Leftrightarrow \quad [y]_R \subseteq A$$
$$y \in R{\uparrow}A \quad \Leftrightarrow \quad [y]_R \cap A \neq \emptyset$$



Rough sets: application domains

- Machine learning
  - Supervised learning, e.g. feature selection and rule induction
  - Unsupervised learning, e.g. rough clustering
- Data warehousing
- Information retrieval
- Multiple Criteria Decision Making
- Semantic Web
- …

Example: data analysis

Applicant	Diploma	Experience	Spanish	Decision
x_1	MSc	Medium	Yes	Accept
x_2	MSc	High	No	Accept
x_3	MSc	High	Yes	Accept
x_4	MBA	High	No	Reject
x_5	MCE	Low	Yes	Reject
x_6	MSc	Medium	Yes	Reject
x_7	MCE	Low	No	Reject

Example: data analysis

Applicant	Diploma	Experience	Spanish	Decision
x_1	MSc	Medium	Yes	Accept
x_2	MSc	High	No	Accept
x_3	MSc	High	Yes	Accept
x_4	MBA	High	No	Reject
x_5	MCE	Low	Yes	Reject
x_6	MSc	Medium	Yes	Reject
x_7	MCE	Low	No	Reject

Example: data analysis

Applicant	Diploma	Experience	Spanish	Decision
x_1	MSc	Medium	Yes	Accept
x_2	MSc	High	No	Accept
x_3	MSc	High	Yes	Accept
x_4	MBA	High	No	Reject
x_5	MCE	Low	Yes	Reject
x_6	MSc	Medium	Yes	Reject
x_7	MCE	Low	No	Reject

$$(x_i, x_j) \in R \Leftrightarrow \begin{cases} Diploma(x_i) & = & Diploma(x_j) \\ Experience(x_i) & = & Experience(x_j) \\ Spanish(x_i) & = & Spanish(x_j) \end{cases}$$

Example: data analysis

Applicant	Diploma	Experience	Spanish	Decision
x_1	MSc	Medium	Yes	Accept
x_2	MSc	High	No	Accept
x_3	MSc	High	Yes	Accept
x_4	MBA	High	No	Reject
x_5	MCE	Low	Yes	Reject
x_6	MSc	Medium	Yes	Reject
x_7	MCE	Low	No	Reject

$$(x_1, x_6) \in R, x_1 \in A, x_6 \notin A$$

Example: data analysis

Applicant	Diploma	Experience	Spanish	Decision
x_1	MSc	Medium	Yes	Accept
x_2	MSc	High	No	Accept
x_3	MSc	High	Yes	Accept
x_4	MBA	High	No	Reject
x_5	MCE	Low	Yes	Reject
x_6	MSc	Medium	Yes	Reject
x_7	MCE	Low	No	Reject

$$R{\downarrow}A \quad = \quad \{x_2, x_3\}$$

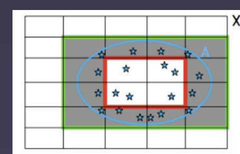
Example: data analysis

Applicant	Diploma	Experience	Spanish	Decision
x_1	MSc	Medium	Yes	Accept
x_2	MSc	High	No	Accept
x_3	MSc	High	Yes	Accept
x_4	MBA	High	No	Reject
x_5	MCE	Low	Yes	Reject
x_6	MSc	Medium	Yes	Reject
x_7	MCE	Low	No	Reject

$$R{\uparrow}A \quad = \quad \{x_1, x_2, x_3, x_6\}$$

Rough set feature selection

- Data reduction method
- Dependent **only** on the data itself
- Reduct: minimal feature subset such that objects' discernibility is preserved
- Decision reduct: minimal feature subset such that objects in different classes can still be discerned



Example: finding a decision reduct

Applicant	Diploma	Experience	Spanish	Decision
x_1	MSc	Medium	Yes	Accept
x_2	MSc	High	No	Accept
x_3	MSc	High	Yes	Accept
x_4	MBA	High	No	Reject
x_5	MCE	Low	Yes	Reject
x_6	MSc	Medium	Yes	Reject
x_7	MCE	Low	No	Reject

Example: finding a decision reduct

Applicant	Experience	Spanish	Decision
x_1	Medium	Yes	Accept
x_2	High	No	Accept
x_3	High	Yes	Accept
x_4	High	No	Reject
x_5	Low	Yes	Reject
x_6	Medium	Yes	Reject
x_7	Low	No	Reject

$\{Experience, Spanish\}$ is no decision reduct

Example: finding a decision reduct

Applicant	Diploma	Experience	Decision
x_1	MSc	Medium	Accept
x_2	MSc	High	Accept
x_3	MSc	High	Accept
x_4	MBA	High	Reject
x_5	MCE	Low	Reject
x_6	MSc	Medium	Reject
x_7	MCE	Low	Reject

$\{Diploma, Experience\}$ is a decision reduct

Finding decision reducts

Theorem (Skowron and Rauszer, 1992)

Given a set of objects $X = \{x_1, \ldots, x_n\}$, a set of conditional attributes $\mathcal{A} = \{a_1, \ldots, a_m\}$ and a decision attribute d. The decision reducts of $(X, \mathcal{A} \cup \{d\})$ are the prime implicants of the boolean function

$$f(a_1^*, ..., a_m^*) = \bigwedge \{\bigvee O_{ij}^* | 1 \leq j < i \leq n \text{ and } O_{ij} \neq \emptyset\}$$

$$O_{ij} = \begin{cases} \emptyset & \text{if } d(x_i) = d(x_j) \\ \{a \in \mathcal{A} | a(x_i) \neq a(x_j)\} & \text{otherwise} \end{cases}$$

Finding decision reducts

Theorem (Skowron and Rauszer, 1992)

Given a set of objects $X = \{x_1, \ldots, x_n\}$, a set of conditional attributes $\mathcal{A} = \{a_1, \ldots, a_m\}$ and a decision attribute d. The decision reducts of $(X, \mathcal{A} \cup \{d\})$ are the prime implicants of the boolean function

$$f(a_1^*, ..., a_m^*) = \bigwedge \{\bigvee O_{ij}^* | 1 \leq j < i \leq n \text{ and } O_{ij} \neq \emptyset\}$$

$$O_{ij} = \begin{cases} \emptyset & \text{if } d(x_i) = d(x_j) \\ \{a \in \mathcal{A} | a(x_i) \neq a(x_j)\} & \text{otherwise} \end{cases}$$

- Problem of finding all (decision) reducts is NP-complete
- Solution: heuristic approaches for finding (approximate) decision reducts

Positive region

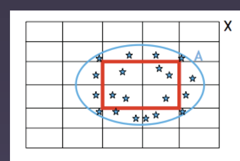
Given a set of objects $X = \{x_1, \ldots, x_n\}$, a set of conditional attributes $\mathcal{A} = \{a_1, \ldots, a_m\}$ and a set of decision classes $\mathcal{C}$.

- For $B \subseteq \mathcal{A}$,

$$R_B = \{(x, y) \in X^2 \mid (\forall a \in B)(a(x) = a(y))\}$$

- Positive region:

$$POS_B = \bigcup_{C \in \mathcal{C}} R_B{\downarrow}C$$



Degree of dependency

Given a set of objects $X = \{x_1, \ldots, x_n\}$, a set of conditional attributes $\mathcal{A} = \{a_1, \ldots, a_m\}$ and a set of decision classes $\mathcal{C}$.

- For $B \subseteq \mathcal{A}$,

$$R_B = \{(x, y) \in X^2 \mid (\forall a \in B)(a(x) = a(y))\}$$

- Positive region:

$$POS_B = \bigcup_{C \in \mathcal{C}} R_B{\downarrow}C$$

- Degree of dependency:

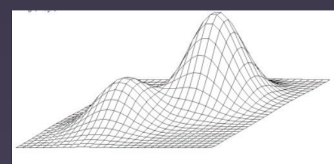
$$\gamma_B = \frac{|POS_B|}{|X|}$$

Theorem

B is a decision reduct if $\gamma_B = \gamma_{\mathcal{A}}$ and $\gamma_{B'} < \gamma_B$ for all $B' \subset B$.

Heuristic search

- Goal: to find a subset $B \subseteq \mathcal{A}$ such that
  - γ_B is maximal
  - $|B|$ is minimal
- Greedy approaches (hillclimbing)
- More complex heuristics: genetic algorithms, ant colony optimization, …



Generalizations of Pawlak rough sets

The definition of lower and upper approximation may be weakened

- Variable Precision Rough Sets (Ziarko, 1993): given $1 \geq u > l \geq 0$,

$$y \in R{\downarrow}A \quad \Leftrightarrow \quad \frac{|[y]_R \cap A|}{|[y]_R|} \geq u$$
$$y \in R{\uparrow}A \quad \Leftrightarrow \quad \frac{|[y]_R \cap A|}{|[y]_R|} > l$$

- If $u = 1$ and $l = 0$, Pawlak's approximations are recovered
- Intuition: introduce noise tolerance into approximations

Generalizations of Pawlak rough sets

The requirement that R is an equivalence relation may be weakened

- Reflexive + transitive: dominance based rough sets (Greco, Matarazzo and Słowiński, 2001) $\rightarrow$ MCDM
- Reflexive + symmetric: tolerance rough sets
  E.g. proximity-based

$$(x, y) \in R \Leftrightarrow d(x, y) \leq \alpha$$

Overview

- Introduction
- Rough Sets (RS)
  - Pawlak's model and generalizations
  - Application: feature selection
- Fuzzy Rough Sets (FRS)
  - Implication/t-norm based model
  - Vaguely quantified rough set model
  - Applications in data analysis
  - Software
- Conclusion

Fuzzy rough sets: motivation

Indiscernibility may be gradual rather than binary

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	d
x_1	1	101	50	15	36	24.2	0.526	26	0
x_2	8	176	90	34	300	33.7	0.467	58	1
x_3	7	150	66	42	342	34.7	0.718	42	0
x_4	7	187	68	39	304	37.7	0.254	41	1
x_5	0	100	88	60	110	46.8	0.962	31	0
x_6	0	105	64	41	142	41.5	0.173	22	0
x_7	1	95	66	13	38	19.6	0.334	25	0

(Diabetes dataset–partim, UCI Machine Learning Repository)

Fuzzy rough sets: motivation

Indiscernibility may be gradual rather than binary

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	d
x_1	1	101	50	15	36	24.2	0.526	26	0
x_2	8	176	90	34	300	33.7	0.467	58	1
x_3	7	150	66	42	342	34.7	0.718	42	0
x_4	7	187	68	39	304	37.7	0.254	41	1
x_5	0	100	88	60	110	46.8	0.962	31	0
x_6	0	105	64	41	142	41.5	0.173	22	0
x_7	1	95	66	13	38	19.6	0.334	25	0

(Diabetes dataset–partim, UCI Machine Learning Repository)

Allow that R is a fuzzy relation

Fuzzy rough sets: motivation

Concepts may be fuzzy rather than crisp

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	d
x_1	0.0351	95	2.68	0	0.4161	7.853	33.2	5.118	48.5
x_2	0.0837	45	3.44	0	0.437	7.185	38.9	4.567	34.9
x_3	0.1061	30	4.93	0	0.428	6.095	65.1	6.336	20.1
x_4	0.0883	12.5	7.87	0	0.524	6.012	66.6	5.561	22.9
x_5	1.4139	0	19.58	1	0.871	6.129	96.0	1.749	17.0
x_6	2.1492	0	19.58	0	0.871	5.709	98.5	1.623	19.4
x_7	3.3211	0	19.58	1	0.871	5.403	100	1.322	13.4

(Housing dataset–partim, UCI Machine Learning Repository)

Fuzzy rough sets: motivation

Concepts may be fuzzy rather than crisp

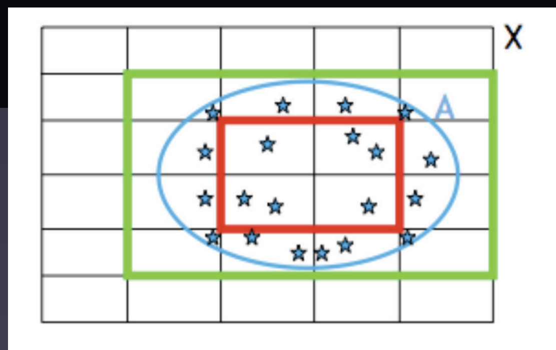
	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	d
x_1	0.0351	95	2.68	0	0.4161	7.853	33.2	5.118	48.5
x_2	0.0837	45	3.44	0	0.437	7.185	38.9	4.567	34.9
x_3	0.1061	30	4.93	0	0.428	6.095	65.1	6.336	20.1
x_4	0.0883	12.5	7.87	0	0.524	6.012	66.6	5.561	22.9
x_5	1.4139	0	19.58	1	0.871	6.129	96.0	1.749	17.0
x_6	2.1492	0	19.58	0	0.871	5.709	98.5	1.623	19.4
x_7	3.3211	0	19.58	1	0.871	5.403	100	1.322	13.4

(Housing dataset–partim, UCI Machine Learning Repository)

Allow that A is a fuzzy set

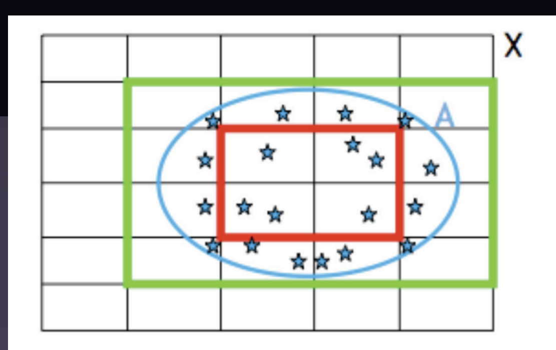
Rough set $(R{\downarrow}A, R{\uparrow}A)$

$$y \in R{\downarrow}A \quad \Leftrightarrow \quad [y]_R \subseteq A$$
$$y \in R{\uparrow}A \quad \Leftrightarrow \quad [y]_R \cap A \neq \emptyset$$



Rough set $(R{\downarrow}A, R{\uparrow}A)$

$$y \in R{\downarrow}A \quad \Leftrightarrow \quad (\forall x \in X)((x,y) \in R \Rightarrow x \in A)$$
$$y \in R{\uparrow}A \quad \Leftrightarrow \quad (\exists x \in X)((x,y) \in R \wedge x \in A)$$



Fuzzy rough set $(R{\downarrow}A, R{\uparrow}A)$

$$(R{\downarrow}A)(y) = \inf_{x \in X} \mathcal{I}(R(x,y), A(x))$$
$$(R{\uparrow}A)(y) = \sup_{x \in X} \mathcal{T}(R(x,y), A(x))$$

- $\mathcal{I}(x,y) = \max(1-x,y)$, $\mathcal{T}(x,y) = \min(x,y)$ (Dubois and Prade, 1990)
- S-, R- and QL-implications (Radzikowska and Kerre, 2002)
- If A and R are crisp, we retrieve Pawlak's approximations

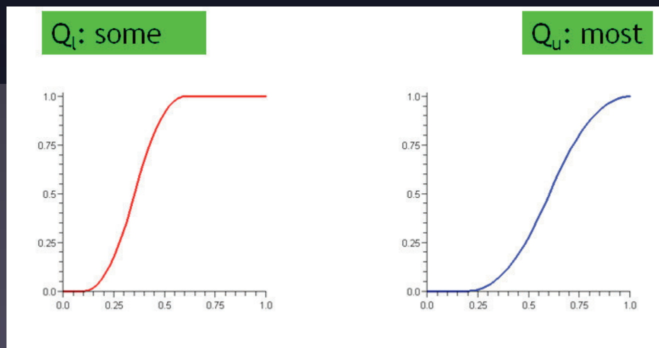
Vaguely Quantified Rough Sets

Principle: soften the quantifiers inside the definitions of lower and upper approximation

- y belongs to the lower approximation of A iff
  - Pawlak: **all** elements of $[y]_R$ belong to A
  - VPRS: **at least a fraction** u of $[y]_R$ belongs to A
  - VQRS: **most** elements of $[y]_R$ belong to A
- y belongs to the upper approximation of A iff
  - Pawlak: **at least one** element of $[y]_R$ belongs to A
  - VPRS: **more than a fraction** l of $[y]_R$ belongs to A
  - VQRS: **some** elements of $[y]_R$ belong to A

Vaguely Quantified Rough Sets

y belongs to the lower approximation of A iff **most** elements of $[y]_R$ belong to A
y belongs to the upper approximation of A iff **some** elements of $[y]_R$ belong to A



Vaguely Quantified Rough Sets

$$R{\downarrow}A(y) = Q_u\left(\frac{|[y]_R \cap A|}{|[y]_R|}\right)$$
$$R{\uparrow}A(y) = Q_l\left(\frac{|[y]_R \cap A|}{|[y]_R|}\right)$$

(Cornelis, De Cock and Radzikowska, 2007)

- If R and A are crisp, Pawlak's approximations are NOT retrieved
- VQRS uses cardinality-based inclusion/overlap measures, while classical FRS uses logic-based measures

Fuzzy-rough feature selection

- Given
  - a set of objects $X = \{x_1, \ldots, x_n\}$,
  - a set of conditional attributes $\mathcal{A} = \{a_1, \ldots, a_m\}$
  - a fuzzy tolerance relation R_B for any $B \subseteq \mathcal{A}$
  - a set of decision classes $\mathcal{C}$
- Positive region:

$$POS_B(x) = \left(\bigcup_{C \in \mathcal{C}} R_B{\downarrow}C\right)(x)$$

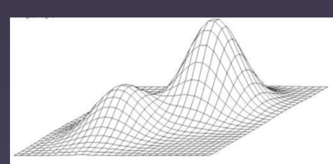
- Degree of dependency:

$$\gamma_B = \frac{|POS_B|}{|X|} = \frac{|\sum_{x \in X} POS_B(x)|}{|X|}$$

Fuzzy-rough feature selection

Definition (Jensen and Shen, 2007)
B is a decision reduct if $\gamma_B = \gamma_{\mathcal{A}}$ and $\gamma_{B'} < \gamma_B$ for all $B' \subset B$.

- Heuristic approaches to find a subset $B \subseteq \mathcal{A}$ such that
  - γ_B is maximal
  - $|B|$ is minimal
- Other extensions of decision reducts have been considered in e.g. (Cornelis, Jensen, Hurtado and Ślęzak, 2010)



Fuzzy-rough K-nearest neighbours

- Goal: classification of test object y given training data T
- K nearest neighbours in T determine y's membership to lower and upper approximation of each class
- Class with highest membership is chosen (Jensen and Cornelis, 2008)

(1) GetNearestNeighbours(y,K)
(2) $\mu_1(y) \leftarrow 0$, $\mu_2(y) \leftarrow 0$, *Class* $\leftarrow \emptyset$
(3) $\forall C \in \mathcal{C}$
(4) **if** $((R{\downarrow}C)(y) \geq \mu_1(y) \wedge (R{\uparrow}C)(y) \geq \mu_2(y))$
(5)   *Class* $\leftarrow C$
(6) $\mu_1(y) \leftarrow (R{\downarrow}C)(y)$, $\mu_2(y) \leftarrow (R{\uparrow}C)(y)$
(7) **output** *Class*

QuickRules

- Goal: generate fuzzy classification rules using minimum number of attributes
- Integrates feature selection and rule induction
  - Decision reduct is obtained by a hillclimbing search
  - On the fly, decision rules are generated for fully covered training objects
- (Jensen, Cornelis and Shen, 2009)

Fuzzy-rough data analysis in practice

WEKA
The University of Waikato

Software

Several fuzzy-rough feature selection and classification methods have been ported to WEKA and are available at Richard Jensen's homepage

http://users.aber.ac.uk/rkj/home/

Conclusion

- Fuzzy sets model gradual information
- Rough sets model incomplete information
- They are highly complementary soft computing paradigms
- They have many applications, in particular in data analysis
- (Fuzzy) rough sets raise many research challenges, both practical and theoretical

Bibliography

- D. Chen, E. Tsang, S. Zhao, An approach of attributes reduction based on fuzzy rough sets, Proc. IEEE Int. Conf. on Systems, Man, and Cybernetics, 2007, pp. 486–491.
- D. Chen, E. Tsang, S. Zhao, Attribute reduction based on fuzzy rough sets, Proc. Int. Conf. on Rough Sets and Intelligent Systems Paradigms, 2007, pp. 73–89.
- C. Cornelis, M. De Cock, A. Radzikowska, Vaguely quantified rough sets, Proceedings of 11th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing (RSFDGrC2007), Lecture Notes in Artificial Intelligence 4482, 2007, pp. 87–94.
- C. Cornelis, M. De Cock, A.M. Radzikowska, Fuzzy rough sets: from theory into practice, Handbook of Granular Computing (W. Pedrycz, A. Skowron, V. Kreinovich, eds.), John Wiley and Sons, 2008, pp. 533–552.
- R. Jensen, C. Cornelis, A new approach to fuzzy-rough nearest neighbour classification, Proceedings of the 6th International Conference on Rough Sets and Current Trends in Computing (RSCTC 2008), 2008, pp. 310-319
- C. Cornelis, R. Jensen, G. Hurtado Martín D. Ślęzak, Attribute selection with fuzzy decision reducts, Information Sciences 180(2) (2010) 209–224.
- M. De Cock, C. Cornelis, E.E. Kerre, Fuzzy rough sets: the forgotten step, IEEE Transactions on Fuzzy Systems 15(1) (2007) 137–153.
- R. Jensen, Q. Shen, Fuzzy-rough sets assisted attribute selection, IEEE Transactions on Fuzzy Systems 15(1) (2007) 73–89.
- R. Jensen, Q. Shen, New approaches to fuzzy-rough feature selection, IEEE Transactions on Fuzzy Systems 17(4) (2009) 824–838.
- R. Jensen, C. Cornelis, Q. Shen, Hybrid fuzzy-rough rule induction and feature selection, Proceedings of the 18th IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2009), 2009, pp. 1151-1156.

Bibliography

- Z. Pawlak, Rough Sets, International Journal of Computer and Information Sciences 11(5) (1982) 341–356.
- Z. Pawlak, Rough Sets — Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, Dordrecht, Netherlands, 1991.
- A.M. Radzikowska, E.E. Kerre, A comparative study of fuzzy rough sets, Fuzzy Sets and Systems 126 (2002) 137–156.
- A. Skowron, C. Rauszer, The Discernibility Matrices and Functions in Information Systems, Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory (R. Słowiński, ed.), Kluwer Academic Publishers, Dordrecht, Netherlands, 1992, pp. 331–362.
- J. Stepaniuk, Tolerance Information Granules, Monitoring, Security, and Rescue Techniques in Multiagent Systems. Advances in Soft Computing, Springer, 2005, pp. 305–316.
- E.C.C. Tsang, D.G. Chen, D.S. Yeung, X.Z. Wang, J.W.T Lee, attributes reduction using fuzzy rough sets, IEEE Transactions on Fuzzy Systems 16(5) (2008) 1130–1141.
- I.H. Witten, E. Frank, Data Mining: Practical machine learning tools and techniques, 2nd Edition, Morgan Kaufmann, San Francisco, 2005.
- M. Yang, S. Chen, X. Yang, A novel approach of rough set-based attribute reduction using fuzzy discernibility matrix, Proc. 4th Int. Conf. on Fuzzy Systems and Knowledge Discovery, 2007, pp. 96–101.
- S. Zhao, E.C.C. Tsang, On fuzzy approximation operators in attribute reduction with fuzzy rough sets, Information Sciences 178(16), (2007) 3163–3176.

Para terminar

- Gracias por su atención!
- Preguntas?
  (en inglés, por favor;-))