Knowledge-Based Systems 91 (2016) 219–240

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Double-quantitative fusion of accuracy and importance: Systematic measure mining, benign integration construction, hierarchical attribute reduction

Xianyong Zhang[a,b,c,*], Duoqian Miao[b,c]

[a] College of Mathematics and Software Science, Sichuan Normal University, Chengdu 610068, PR China
[b] Department of Computer Science and Technology, Tongji University, Shanghai 201804, PR China
[c] Key Laboratory of Embedded System and Service Computing, Ministry of Education, Shanghai 201804, PR China

ARTICLE INFO

Article history:
Received 4 February 2015
Revised 30 August 2015
Accepted 1 September 2015
Available online 7 September 2015

Keywords:
Rough set theory
Granular computing
Attribute reduction
Uncertainty measure
Double quantification

ABSTRACT

Uncertainty measure mining and applications are fundamental, and it is possible for double-quantitative fusion to acquire benign measures via heterogeneity and complementarity. This paper investigates the double-quantitative fusion of relative accuracy and absolute importance to provide systematic measure mining, benign integration construction, and hierarchical attribute reduction. (1) First, three-way probabilities and measures are analyzed. Thus, the accuracy and importance are systematically extracted, and both are further fused into importance-accuracy (IP-Accuracy), a synthetic causality measure. (2) By sum integration, IP-Accuracy gains a bottom-top granulation construction and granular hierarchical structure. IP-Accuracy holds benign granulation monotonicity at both the knowledge concept and classification levels. (3) IP-Accuracy attribute reduction is explored based on decision tables. A hierarchical reduct system is thereby established, including qualitative/quantitative reducts, tolerant/approximate reducts, reduct hierarchies, and heuristic algorithms. Herein, the innovative tolerant and approximate reducts quantitatively approach/expand/weaken the ideal qualitative reduct. (4) Finally, a decision table example is provided for illustration. This paper performs double-quantitative fusion of causality measures to systematically mine IP-Accuracy, and this measure benignly constructs a granular computing platform and hierarchical reduct system. By resorting to a monotonous uncertainty measure, this study provides an integration-evolution strategy of granular construction for attribute reduction.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Rough set theory (RS-Theory) [35,36,59,63,70,71] represents a fundamental *granular computing (GrC)* pattern for handling uncertainty issues. The initial Pawlak-Model [35] acts only as a qualitative model, so it lacks the quantitative mechanism regarding fault-tolerance and robustness. Thus, quantitative models exhibit improvements and have applications, and they could in part be unified by the subsethood measure [61]. In particular, *the probabilistic rough set (PRS)* [1,27,28,30,51,57,58,60,78] introduces the probability uncertainty measure into RS-Theory, which forms the basis of mainstream quantitative models. PRS offers measurability, generality, and flexibility and exhibits a series of concrete models, including the *decision-theoretic rough set (DTRS)* [64], game-theoretic rough set [1,2], variable precision rough set [81], Bayesian rough set [47], and parameterized rough set [6]. With the exception of PRS, the graded rough set [25,62] depends on the grade measure to become another basic type of quantitative model.

Herein, DTRS is introduced as a model example. DTRS utilizes conditional probability and the Bayesian risk decision to establish three-way decisions and threshold-quantitative semantics [64]. As a result, DTRS improves upon some basic models and provides a quantitative exploration platform. In terms of relevant studies, three-way decisions were analyzed in [17–19,58,60,79]; model development and threshold calculation were discussed in [15,16,27,45,48]; attribute reduction was studied in [10,12,31,65,74,75]; and model applications (regarding clustering, regression, and semi-supervised learning) were addressed in [13,23,24,26,66]. In fact, three-way decisions have been expanded into three-way decision theory, and this fundamental theory has been the subject of extensive study and used in a number of

* Corresponding author at: College of Mathematics and Software Science, Sichuan Normal University, Chengdu 610068, PR China.

E-mail addresses: xianyongzh@sina.com.cn (X. Zhang), miaoduoqian@163.com.cn (D. Miao).

http://dx.doi.org/10.1016/j.knosys.2015.09.001
0950-7051/© 2015 Elsevier B.V. All rights reserved.

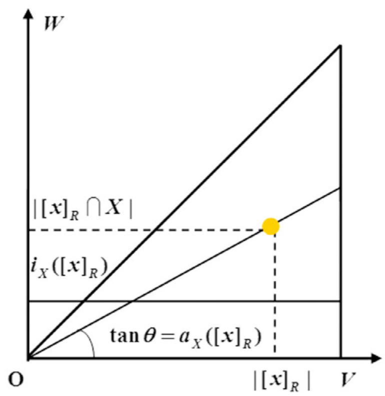

Fig. 1. Accuracy and importance in the cardinality plane.

Table 1
Main abbreviations of this paper.

Abbreviation	Original term
RS-Theory	Rough set theory
GrC	Granular computing
DTRS	Decision-theoretic rough set
Approx-Space	Approximate space
C-POS/C-BND	Classification-positive/boundary region
POS/NEG/BND	Set-positive/negative/boundary region
D-Table	Decision table
Kn-Coarsening	Knowledge coarsening
Gr-Merging	Granule merging
Gr-Preservation	Granule preservation
IP-Accuracy	Importance-accuracy
IIP-Accuracy	Internal importance-accuracy
Gr/Kn/Cl/Con	Granule/knowledge/classification/concept
Gr-Con/Kn-Con/ Kn-Cl	Granular concept/knowledge's concept/knowledge's classification
MT	Monotonicity target

Table 2
Three-way probabilities and three-way measures in Bottom-System (U, R, X).

Metrical essence	Causality relevance	Metrical features	Three-way probabilities	Probability formula	Three-way measures	Measure formula
$\frac{\|[x]_R \cap X\|}{\|[x]_R\|}$	Has	Relativity, concentration, locality	Likelihood probability	$p(X\|[x]_R)$	Accuracy	$a_X([x]_R)$
$\frac{\|[x]_R \cap X\|}{\|X\|}$	Has	Absoluteness, directness, globality (regarding X)	Posterior probability	$p([x]_R\|X)$	Importance	$i_X([x]_R)$
$\frac{\|[x]_R\|}{\|U\|}$	Never has	Absoluteness, vividness, globality (regarding U)	Prior probability	$p([x]_R)$	Generality	$g([x]_R)$

Table 3
Granule-statistical information in Bottom-System $(U, \mathcal{C}, X)$.

Granule $[x]_{\mathcal{C}}^i$	Interaction cardinality $\|[x]_{\mathcal{C}}^i \cap X\|$	Granule cardinality $\|[x]_{\mathcal{C}}^i\|$	Generality $g([x]_{\mathcal{C}}^i)$	Importance $i_X([x]_{\mathcal{C}}^i)$	Accuracy $a_X([x]_{\mathcal{C}}^i)$	IP-Accuracy $ia_X([x]_{\mathcal{C}}^i)$
$[x]_{\mathcal{C}}^1$	8	16	0.16	0.200	0.5	0.1
$[x]_{\mathcal{C}}^2$	6	10	0.10	0.150	0.6	0.09
$[x]_{\mathcal{C}}^3$	6	15	0.15	0.150	0.4	0.06
$[x]_{\mathcal{C}}^4$	5	10	0.10	0.125	0.5	0.0625
$[x]_{\mathcal{C}}^5$	4	16	0.16	0.100	0.25	0.025
$[x]_{\mathcal{C}}^6$	3	3	0.03	0.075	1.0	0.075
$[x]_{\mathcal{C}}^7$	3	4	0.04	0.075	0.75	0.05625
$[x]_{\mathcal{C}}^8$	2	3	0.03	0.050	2/3	1/30
$[x]_{\mathcal{C}}^9$	2	6	0.06	0.050	1/3	1/60
$[x]_{\mathcal{C}}^{10}$	1	1	0.01	0.025	1.0	0.025
$[x]_{\mathcal{C}}^{11}$	0	8	0.08	0.000	0.0	0.0
$[x]_{\mathcal{C}}^{12}$	0	8	0.08	0.000	0.0	0.0

Table 4
Knowledge's relevant uncertainty measures regarding IP-Accuracy.

Knowledge $\mathcal{C}_*$	Kn-Con IP-Accuracy $ia_X(\mathcal{C}_*)$, $ia_{\neg X}(\mathcal{C}_*)$	Kn-Con IIP-Accuracy $Iia_X(\mathcal{C}_*)$, $Iia_{\neg X}(\mathcal{C}_*)$	Kn-Cl IP-Accuracy $ia_{\mathcal{D}}(\mathcal{C}_*)$	Kn-Cl IIP-Accuracy $Iia_{\mathcal{D}}(\mathcal{C}_*)$	IP-Accuracy discrepancy, equality rates $dr_{ia_{\mathcal{D}}}$, $er_{ia_{\mathcal{D}}}$	IIP-Accuracy discrepancy, equality rates $dr_{Iia_{\mathcal{D}}}$, $er_{Iia_{\mathcal{D}}}$
$\{a, b, c, d, e, f\}$	0.5437, 0.6958	0.1000, 0.2667	1.2395	0.3667	0.00%, 100.00%	0.00%, 100.00%
$\{a, b, c, d, e\}$	0.5437, 0.6958	0.1000, 0.2667	1.2395	0.3667	0.00%, 100.00%	0.00%, 100.00%
$\{a, b, c, d\}$	0.5376, 0.6917	0.1000, 0.2667	1.2293	0.3667	0.82%, 99.18%	0.00%, 100.00%
$\{a, b, c\}$	0.5366, 0.6910	0.1000, 0.2667	1.2276	0.3667	0.96%, 99.04%	0.00%, 100.00%
$\{a, b\}$	0.5193, 0.6796	0.1000, 0.0000	1.1989	0.1000	3.28%, 96.72%	72.73%, 27.27%
$\{a, c\}$	0.5050, 0.6700	0.1000, 0.2667	1.1750	0.3667	5.20%, 94.80%	0.00%, 100.00%
$\{b, c\}$	0.5314, 0.6876	0.0000, 0.2667	1.2190	0.1000	1.65%, 98.35%	27.27%, 72.73%
$\{a\}$	0.4375, 0.6250	0.1000, 0.0000	1.0625	0.1000	14.28%, 85.72%	72.73%, 27.27%
$\{b\}$	0.5141, 0.6761	0.0000, 0.0000	1.1902	0.1000	3.98%, 96.02%	100.00%, 0.00%
$\{c\}$	0.4762, 0.6508	0.0000, 0.2667	1.1270	0.1000	9.08%, 90.92%	27.27%, 72.73%
$\varnothing$	0.4000, 0.6000	0.0000, 0.0000	1.0000	0.0000	19.32%, 80.68%	100.00%, 0.00%

Download English Version:

https://daneshyari.com/en/article/402201

Download Persian Version:

https://daneshyari.com/article/402201

Daneshyari.com