Yao, Y.Y., Decision-theoretic Rough Set Models (DTRSM)
Computer Science (Ji Suan Ji Ke Xue), Vol. 35, No. 8A, pp. 7-8, 2008.

Decision-Theoretic Rough Set Models (DTRSM)

Yiyu (Y.Y.) Yao
Department of Computer Science, University of Regina
Regina, Saskatchewan, Canada S4S 0A2
Email: yyao@cs.uregina.ca
http://www.cs.uregina.ca/~yyao

The standard rough sets model is a qualitative model that defines three regions for approximating a subset of a universe of objects based on an equivalence relation on the universe. The positive region (i.e., the lower approximation) is the union of equivalence classes that are subsets of the set to be approximated. The boundary region (i.e., the difference of upper approximation and lower approximation) is the union of equivalence classes that have non-empty intersections with the set and at the same time are not subsets of the set. The negative region (i.e., the complement of the upper approximations) is the union of equivalence classes that have an empty intersection with the set. The theory may be formulated and interpreted in terms of three-valued logic, in which each value corresponds to one of the three regions. More importantly, the three-valued logic is non-truth-functional, that is, one may not be able to obtain the value of compound formula based on its component sub-formulas.

A lack of consideration of the degree of overlap between an equivalence class and the set motivates many researchers to study quantitative rough set models. Probabilistic approaches to rough sets are one of the most important and successful schools of quantitative rough sets. The Decision-Theoretic Rough Set Model (DTRSM) was proposed in early 1990's based on the well-established Bayesian decision procedure. In this talk, I will review basic ideas and recent developments of DTRSM. Specifically, the following topics will be covered:

1. **Probabilistic approximations and regions**. We define the probabilistic positive, boundary, and negative regions, or equivalently the probabilistic lower and upper approximations, by using a pair of threshold parameters. An equivalence class is in the probabilistic positive region if its relative overlap with the set (i.e., the conditional probability) is above or equal to a threshold, is in the negative region if its relative overlap is below another threshold, and is in the boundary region if the relative overlap is between the two parameters. Although the formulation is intuitively appealing, a fundamental question that remains is how to determine the pair of parameters. In DTRSM, it is shown that the parameters can be systematically computed from a risk or loss function based on the Bayesian decision procedure. More importantly, the derived parameters would minimize the overall cost of classifying an object into the positive, boundary and negative regions.

2. **Interpretations of probabilistic rules.** Probabilistic rough sets may be viewed as an extension of the standard rough sets. Many authors apply the concepts and notions of the latter to the study of the former in a straightforward manner, without considering the special features of the probabilistic models. In some situations, this may introduce inconsistency and semantics difficulties. One of the semantics difficulties is the interpretation of probabilistic rules induced from the probabilistic regions. Since the precision (i.e., accuracy) of rules induced from three regions may be uncertain, assigning a quantitative measure to rules is no longer sufficient. It is necessary to study the physical meanings of rules and their implications. The use of loss or cost functions in DTRSM provides a meaningful interpretation. Given a class, its positive, boundary and negative regions represent three different types of decisions. For example, consider classifying a set of patients with respect to a particular disease. A patient in the positive region needs an "immediately treatment," a patient in the boundary requires a "further investigation," and a patient in the negative region does not require any treatment. With respect to the three regions, we in fact have three different types of rules, called positive, boundary, and negative rules, respectively. They represent three different decisions with different costs. In many

applications, it may be sufficient to consider only the first two classes.

3. **Reduct construction**. Several proposals have been made regarding reduct construction in probabilistic rough set models. In standard rough sets, a reduct with respect to a classification (i.e., a relative reduct) is defined by requiring the same positive region of the classification, or equivalently the same-sized positive region. A straightforward generalization in the probabilistic models is to require the same probabilistic positive region, or the same-sized probabilistic positive region. However, these two characterizations of probabilistic positive region are no longer equivalent. Other proposals, based on probability distributions, require the same probabilistic distribution, the same ordering of decision classes defined by the probability distributions, or the same decision class defined by the maximum conditional probabilities. In terms of conditional entropy, one may also require the same entropy value. Although these proposals are intuitively appealing, they may not capture the true meaning of probabilistic rough sets. The positive region in the standard rough set model demands 100 percent inclusion of an equivalence class in the set to be approximated. This requirement is relaxed in probabilistic rough sets. The probabilistic positive region only requires the degree of inclusion to be above a certain level. The result is normally a larger positive region. In other words, we trade precision (i.e., the accuracy) for generality (i.e., the size) with the probabilistic positive region. The trade-off is determined by a threshold value representing our tolerance of imprecision. However, existing formulations and interpretations of reducts do not truthfully reflect such a trade-off. In DTRSM, we can easily interpret the trade-off in terms of acceptable loss or cost. Semantically, a probabilistic reduct can be defined as a minimal set of attributes with an acceptable level of loss or cost, which determines the threshold values.

The standard rough set model provides a qualitative data analysis theory and the probabilistic rough set models enable us to perform the corresponding quantitative analysis. The decision-theoretic rough set model offers a solid foundation for probabilistic rough sets based on well-established Bayesian decision theory. With DTRSM, we can resolve several difficulties associated with some studies on probabilistic rough sets, including the determination of threshold values, the interpretation of rules, and the definition and construction of probabilistic reducts.

In a wider context, we need to examine the research efforts on “generalizations” in any theory. We must make useful generalizations rather than useless generalizations and meaningful generalizations rather than meaningless generalizations. Generalizations should not be attempted only for the sake of generalizations; they must be means to useful ends. It seems that DTRSM is a meaningful and useful generalization of the standard rough set model.

More information and applications of DTRSM can be found in the references listed below and a website (http://www.cs.uregina.ca/~DTRS), and papers on DTRSM are also available for download at: http://www.cs.uregina.ca/~yyao/decision_theoretic_rough_set_paper/.

References

[1] Abd El-Monsef, M.M.E. and Kilany, N.M. Decision analysis via granulation based on general binary relation, *International Journal of Mathematics and Mathematical Sciences*, Article ID 12714, 2007.

[2] Ciucci, D. A unifying abstract approach for rough models, *Rough Sets and Knowledge Technology, Third International Conference, RSKT 2008*, LNAI 5009, 371-378, 2008.

[3] Greco, S., Slowinski, R. and Yao, Y.Y. Bayesian decision theory for dominance-based rough set approach, *Proceedings of the Second International Conference on Rough Sets and Knowledge Technology*, RSKT 2007, LNAI 4481, 134-141, 2007.

[4] Herbert, J.P. and Yao, J.T. Rough set model selection for practical decision making, *Proceedings of the Fourth International Conference on Fuzzy Systems and Knowledge Discovery*, FSKD 2007, 203-207, 2007.

[5] Herbert, J.P. and Yao, J.T. Game-theoretic risk analysis in decision-theoretic rough sets,

Proceedings of the Third International Conference on Rough Sets and Knowledge Technology, RSKT 2008, LNAI 5009, 132-139, 2008.

[6] Herbert, J.P. and Yao J.T. Criteria for choosing a rough set model, *Journal of Computers and Mathematics with Applications*, in press, 2008.

[7] Li, Y., Zhang, C. and Swanb, J.R. Rough set based model in information retrieval and filtering, *Proceedings of the Fifth International Conference on Information Systems Analysis and Synthesis*, 398-403, 1999.

[8] Li, Y., Zhang, C. and Swanb, J.R. An information filtering model on the Web and its application in JobAgent, *Knowledge-Based Systems*, 13, 285-296, 2000.

[9] Yao, J.T. and Herbert, J.P. Web-based support systems with rough set analysis, *Proceedings of Rough Sets and Emerging Intelligent Systems Paradigms*, RSEISP 2007, LNAI 4585, 360-370, 2007.

[10] Yao, J.T. and Herbert, J.P. A game-theoretic perspective on rough set analysis, *Journal of Chongqing University of Posts and Telecommunications* (Natural Science Edition), 20, 291-298, 2008.

[11] Yao, Y.Y. and Zhao, Y. Attribute reduction in decision-theoretic rough set models, *Information Sciences*, to appear.

[12] Yao, Y.Y. Probabilistic rough set approximations, *International Journal of Approximate Reasoning*, to appear.

[13] Yao, Y.Y. Decision-theoretic rough set models, *Proceedings of the Second International Conference on Rough Sets and Knowledge Technology*, RSKT 2007, LNAI 4481, 1-12, 2007.

[14] Yao, Y.Y. Probabilistic approaches to rough sets, *Expert Systems*, 20, 287-297, 2003.

[15] Yao, Y.Y. Information granulation and approximation in a decision-theoretical model of rough sets, in: Polkowski, L., Pal, S.K., and Skowron, A. (Eds), *Rough-neuro Computing: Techniques for Computing with Words*, Springer, Berlin, 491-516, 2003.

[16] Yao, Y.Y. and Wong, S.K.M. A decision theoretic framework for approximating concepts, *International Journal of Man-machine Studies*, 37, 793-809, 1992.

[17] Yao, Y.Y., Wong, S.K.M. and Lingras, P. A decision-theoretic rough set model, in: Z.W. Ras, M. Zemankova and M.L. Emrich (Eds.) *Methodologies for Intelligent Systems*, 5, North-Holland, New York, 17-24, 1990.

[18] Zhao, W.Q. and Zhu, Y.L. An email classification scheme based on decision-theoretic rough set theory and analysis of email security, Proceeding of 2005 IEEE Region 10 *TENCON*, Digital Object Identifier: 10.1109/TENCON.2005.301121, 2005.

[19] Zhao, Y., Wong, S.K.M. and Yao, Y.Y. A note on attribute reduction in probabilistic rough set models, manuscript, 2008.

[20] 李秀红，张东升，(α,β)-粗糙集模型中阈值的确定与解释，山东大学学报(工学版)，36，81-85，2006.

[21] 张文修, 仇国芳，基于粗糙集的不确定决策, 北京, 清华大学出版社, 2005.

[22] 张文修, 吴伟志, 梁吉业, 李德玉, 粗糙集理论与方法, 科学出版社, 北京, 2001.

[23] 赵文清，朱永利，高伟，一个基于决策粗糙集理论的信息过滤模型，计算机工程与应用，43，185-194, 2007.

[24] 周辉，王黔英，费颖，袁芳，基于α相同度相似关系的rough集模型，计算机应用，26，666-667，2006.