2011 International Conference on Information Communication and Management
IPCSIT vol.16 (2011) © (2011) IACSIT Press, Singapore

Learning Influence Diagram with Rough Sets for Bacterial Infection Treatments

Chia-Hui Huang [1 +] and Han-Ying Kao [2]

[1] Department of Business Administration, National Taipei College of Business,
No. 321, Sec. 1, Ji-Nan Road, Zhongzheng District, Taipei 100, Taiwan, R.O.C.

[2] Institute of Networking and Multimedia Technology, National Dong Hwa University,
No. 1, Sec. 2, Da-Hsueh Road, Shoufeng, Hualien 97401, Taiwan, R.O.C.

teresak@mail.ndhu.edu.tw

Abstract. There are several advantages to evaluating a problem with influence diagram operations. The analyst can use a representation that is natural to the decision maker since the algorithm executes all of the inference and analysis automatically. The influence diagram solution procedure can also result in significant gains in efficiency. Conditional independence is clearly exhibited in the diagram, so the size of intermediate calculations can be reduced which is resulted in considerable reductions in processing time and memory requirements. However, when imprecise knowledge from large-scaled data set is involved in the systems, how to reason from approximate information becomes a main issue in evaluating influence diagrams effectively. This study proposes an alternative numerical framework for influence diagrams, rough sets and also develops a rough set-based influence diagrams which combine rough set decision rules with the graphical structure of the influence diagrams in medical settings. The proposed knowledge model provides a comprehensive way for knowledge representation and decision support.

Keywords: Influence diagrams, rough sets, Bayesian networks, bacterial infection.

1. Introduction

There are several important missions of a medical reasoning system: diagnosis, prediction, treatment planning, etc. [3-8]. Diagnosis is the process of reconstructing the past facts from the observed evidence. Prediction is the process of projecting the evidences from hypotheses. Treatment planning is reasoning about the costs and effects of treatments on patients. Usually, medical practice requires various kinds of reasoning simultaneously. Hence, the capability for multiple reasoning tasks is critical to the performance of medical decision support systems. Besides, medical expert systems become more complex when considering the mechanism of human bodies and their mutual interactions with the environmental factors.

In medical informatics, graphical decision models such as Bayesian networks and influence diagrams have been widely used as knowledge representation and decision models [2-4, 7, 9-13, 19]. Influence diagrams are a graphical technique for a decision problem under uncertainty [1, 11, 21], which have been widely used as knowledge representation and decision models [2, 11, 12, 20, 21]. Various methods have been developed for learning or evaluating influence diagrams [2, 11, 12, 20, 21]. In previous researches, the numerical models of the influence diagrams used to be limited in probability distributions [1, 20].

However, when imprecise knowledge from large-scaled data set is involved in the systems, how to reason from approximate information becomes a main issue in evaluating influence diagrams effectively. This study proposes an alternative numerical model for the knowledge in influence diagrams with rough sets.

[+] Corresponding author. Tel.: +886-2-2322-6402; fax: +886-2-2322-6404.
E-mail address: leohuang@webmail.ntcb.edu.tw.

Rough set theory was first introduced by Pawlak [14] as a tool dealing with risk and impreciseness in decision-making. The probabilistic approaches have been previously applied to rough set theory [15-18, 22]. This study intends to propose an alternative numerical model for influence diagrams, which extend influence diagrams into rough-set influence diagrams.

2. Learning Influence Diagram with Rough Sets

Most literatures on influence diagrams [1,11,19,20] used to describe the dependency and its associated numerical model with probability theory. However, when impreciseness and large data volume involved in the domain, the decision makers may need more flexible uncertainty measures for analysis. Rough set theory can be an alternative measure in such a problem.

Given the influence diagrams structure and the original data set, rough set theory provides a basis for extracting the knowledge and expressing the dependency among nodes in the influence diagrams. To represent the ontology, we define that rough set-based influence diagram is a directed acyclic graph RSID = (*U*, *A*, *f*), where:

(1) *U* is the universe, a nonempty finite set of objects.

(2) *A* is a set of nodes where $A \equiv C \bigcup D \equiv V_D \bigcup V_R \bigcup V_U$.

(3) *f* is the flow function representing the strength, certainty factor, and coverage factor of the decision rule.

Given the influence diagrams structure and the original data set, rough set theory provides a basis for extracting the knowledge and expressing the dependency among nodes in the influence diagrams. To represent the ontology, we define that rough set-based influence diagram is a directed acyclic graph **RSID** = (*U*, *A*, *f*), where:

With every branch of (x_i, x_{i+1}) there is a directed arc from node x_i to x_{i+1} as shown in Fig. 1. The certainty, coverage, and strength of a branch (x_i, x_{i+1}) are defined as $cer(x_i, x_{i+1})$, $cov(x_i, x_{i+1})$, and $\sigma(x_i, x_{i+1})$, respectively.

Fig. 1: A directed arc from node x_i to x_{i+1}

A directed path from x_1 to x_n for x_1 , $x_n \in A$ denoted $[x_1, x_n]$, is a sequence of nodes $x_1, x_2, \ldots, x_i, \ldots, x_n$, $1 \le i \le n$, as shown in Fig. 2.

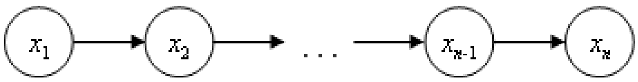

Fig. 2: A directed path $[x_1, x_n]$ from node x_1 to x_n

The certainty of a directed path $[x_1, x_n]$ is defined as

$$cer[x_1, x_n] = \prod_{i=1}^{n-1} cer(x_1, x_n) \tag{1}$$

The coverage of a directed path $[x_1, x_n]$ is defined as

$$cov[x_1, x_n] = \prod_{i=1}^{n-1} cov(x_1, x_n) \tag{2}$$

The strength of a directed path $[x_1, x_n]$ is defined as

$$\sigma[x_1, x_n] = \begin{cases} \sigma(x_1)cer[x_1, x_n] = \sigma(x_1)\prod_{i=1}^{n-1} cer(x_1, x_n) \\ \sigma(x_n)cov[x_1, x_n] = \sigma(x_n)\prod_{i=1}^{n-1} cov(x_1, x_n) \end{cases} \tag{3}$$

The set of all directed paths from x_1 to x_n denoted $\langle x_1, x_n \rangle$ as shown in Fig. 3, where $[x_1, x_n] \in \langle x_1, x_n \rangle$.

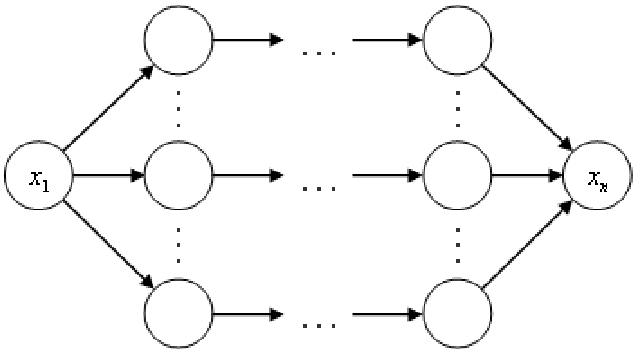

Fig. 3: The set, $\langle x_1, x_n \rangle$, of all directed paths from node x_1 to x_n

The certainty of $\langle x_1, x_n \rangle$ is defined as

$$cer\langle x_1, x_n \rangle = \sum_{[x_1, x_n] \in \langle x_1, x_n \rangle} cer[x_1, x_n] \tag{4}$$

The coverage of $\langle x_1, x_n \rangle$ is defined as

$$cov\langle x_1, x_n \rangle = \sum_{[x_1, x_n] \in \langle x_1, x_n \rangle} cov[x_1, x_n] \tag{5}$$

The strength of $\langle x_1, x_n \rangle$ is defined as

$$\sigma\langle x_1, x_n \rangle = \sum_{[x_1, x_n] \in \langle x_1, x_n \rangle} \sigma[x_1, x_n] = \begin{cases} \sum_{[x_1, x_n] \in \langle x_1, x_n \rangle} \sigma(x_1)\prod_{i=1}^{n-1} cer(x_1, x_n) \\ \sum_{[x_1, x_n] \in \langle x_1, x_n \rangle} \sigma(x_n)\prod_{i=1}^{n-1} cov(x_1, x_n) \end{cases} \tag{6}$$

3. Bacterial Infection Treatments

An example of influence diagrams for metastatic cancer and treatment is modified from Pearl [19] as shown in Fig. 4, where $V = V_D \cup V_R \cup V_U$, $V_D = \{T\}$, $V_R = \{B, C, I, M\}$, and $V_U = \{Q\}$.

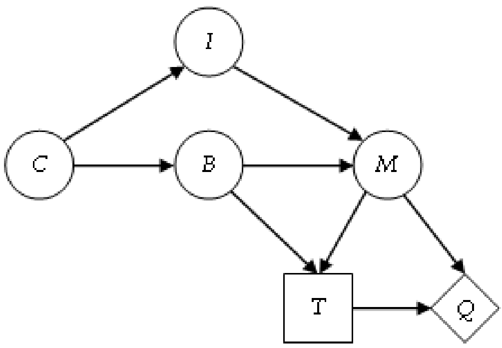

Fig. 4: An example of influence diagrams for metastatic cancer and treatment

In Fig. 4, there is one decision node *T* (Treatment?), which has two alternatives to take: yes or no. Four chance variables are relevant to the biological test and treatment problems: *B* (Brain Tumor), *C* (Coma), *I*

(Increased Total Serum Calcium), and M (Metastatic Cancer). The utility function Q (Quality-adjusted Life Expectancy) is to be maximized. The meaning and states of the nodes are summarized in Table 1.

In this example, the states of I and B are conditioned on the state of C. The states of I and B will influence the manifestation of M. The outcome of M and the decision on T will determine the value of Q, where B and M provide the information prior to decision making. Usually, the causal relationships and decision rules in influence diagrams are expressed with probability distributions and a value table.

Table 1: The Descriptions and States of the Nodes in Fig. 4

Node	Description	State
B	Brain Tumor	$B \in \{0,1\}$, 0: poor, 1: good
C	Coma	$C \in \{0,1\}$, 0: poor, 1: good
I	Increased Total Serum Calcium	$I \in \{0,1\}$, 0: low, 1: high
M	Metastatic Cancer	$M \in \{0,1\}$, 0: poor, 1: good
T	Treatment	$T \in \{\text{yes}, \text{no}\}$
Q	Quality-adjusted Life Expectancy	

The objective of this problem is to maximize the expected utilities as follows

$$\max EU(Q = q(m,t)) = \sum_{b,c,i,m} q(m,t)\sigma(m \mid b,i)\sigma(b \mid c)\sigma(i \mid c)\sigma(c) \tag{7}$$

where $EU(*)$ stands for the expected value of "*". Note that the uppercase and lowercase letter represents the variable and the value of a variable, respectively. The expected utilities of Q based on T = no and T = yes are computed as follows, respectively.

$$EU_1 = EU(Q = q(m,t)) = \sum_{b,c,i,m} q(p,\text{no})\sigma(m \mid b,i)\sigma(b \mid c)\sigma(i \mid c)\sigma(c) \tag{8}$$

$$EU_2 = EU(Q = q(m,t)) = \sum_{b,c,i,m} q(p, yes)\sigma(m \mid b,i)\sigma(b \mid c)\sigma(i \mid c)\sigma(c) \tag{9}$$

Hence, the optimal decision to maximize the utilities is

$$\max EU(Q = q(m,t)) = \max\{EU_1, EU_2\} \tag{10}$$

4. Conclusion

This study proposes an alternative numerical framework for influence diagrams, rough sets and also develops a rough set-based influence diagrams which combine rough set decision rules with the graphical structure of the influence diagrams in medical settings. Considering the imprecise knowledge from large-scaled data set, this study formulates the causal relationships and the decision rules among the nodes (attributes) with rough sets from information systems. The proposed knowledge model provides a comprehensive way for knowledge representation and decision support.

For future studies, there are some potential themes:

(1) Integrated analysis with rough sets in various graphical decision models, including Bayesian networks, decision trees, and influence diagrams.

(2) Hybrid decision analysis with fuzzy sets and rough sets in graphical decision models.

(3) Potential applications of intelligent decision support with rough sets, such as supply chain management, business strategic analysis, and biomedicine.

5. Acknowledgements

The authors appreciate the anonymous referees for their careful reading and the fruitful comments for the manuscript. Also special thanks to National Science Council in Taiwan, R.O.C. for supporting this study under grant No. NSC 100-2410-H-141-006-MY2 (C.-H. Huang) and No. NSC 99-2410-H-259-045-MY2 (H.-Y. Kao).

6. References

[1] R.A. Howard, J.E. Matheson. Influence Diagrams. In: R.A. Howard, J.E. Matheson (eds.). *The Principles and Applications of Decision Analysis*. Strategic Decisions Group, Menlo Park. 1981, pp. 719-762.

[2] F.V. Jensen. *Bayesian Networks and Decision Graphs*. Springer-Verlag, Inc., New York, 2001.

[3] H.Y. Kao. Diagnostic reasoning and medical decision-making with fuzzy influence diagrams. *Computer Methods and Programs in Biomedicine*. 2008, **90** (1): 9-16.

[4] H.Y. Kao, H.L Li. A diagnostic reasoning and optimal treatment model for bacterial infections with fuzzy information. *Computer Methods and Programs in Biomedicine*. 2005, **77** (1): 23-37.

[5] I. Kononenko. Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine*. 2001, **23** (1): 89-109.

[6] L. Leibovici, M. Fishman, H.C. Schonheyder, C. Riekehr, B. Kristensen, I. Shraga, S. Andreassen. A causal probabilistic network for optimal treatment of bacterial infections. *IEEE Transactions on Knowledge and Data Engineering*. 2000, 12 (2): 517-528.

[7] H.L Li, H.Y. Kao. Constrained abductive reasoning with fuzzy parameters in Bayesian networks. *Computers & Operations Research*. 2005, **32** (1): 87-105.

[8] W.J. Long. Medical informatics: reasoning methods. *Artificial Intelligence in Medicine*. 2001, **23** (1): 71-87.

[9] M. López-Díaz, L.J. Rodríguez-Muñiz. Influence diagrams with super value nodes involving imprecise information. *European Journal of Operational Research*. 2007, **179** (1): 203-219.

[10] R.E. Neapolian. *Learning Bayesian Networks*. Pearson Education Inc., 2004.

[11] R.F. Nease, D.K. Owens. Use of influence diagrams to structure medical decisions. *Medical Decision Making*. 1997, **17** (3): 263-275.

[12] R.M. Oliver, J.Q. Smith. *Influence Diagrams, Belief Nets and Decision Analysis*. John Wiley & Sons, 1990.

[13] D.K. Owens, R.D. Shachter, R.F. Nease. Representation and analysis of medical decision problems with influence diagrams. *Medical Decision Making*. 1997, **17** (3): 251-262.

[14] Z. Pawlak. Rough sets. *Informational Journal of Information and Computer Sciences*. 1982, **11** (5): 341-356.

[15] Z. Pawlak. *Rough Sets-Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, Dordrecht, 1991.

[16] Z. Pawlak. Rough sets, decision algorithms and Bayes' theorem. *European Journal of Operational Research*. 2002, **136** (1): 181-189.

[17] Z. Pawlak. Probability, truth and flow graph. *Electronic Notes in Theoretical Computer Science*. 2003, **82** (4): 1-9.

[18] Z. Pawlak, S.K.M. Wong, W. Ziarko. Rough sets: probabilistic versus deterministic approach. *International Journal of Man-Machine Studies*. 1988, **29** (1): 81-95.

[19] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, Inc., 1997.

[20] R.D. Shacher. Evaluating influence diagrams. *Operations Research*. 1986, **34** (6): 871-882.

[21] J.A. Tatman, R.D. Shacher. Dynamic programming and influence diagrams. *IEEE Transactions on Systems, Man and Cybernetics*. 1990, **20** (2): 365-379.

[22] Y.Y. Yao. Probabilistic approaches to rough sets. *Expert Systems*. 2003, **20** (5): 287-297.