

Comparisons between Rough Set Based and Computational Applications in Data Mining

En-Bing Lin and Yu-Ru Syau

Abstract—Rough set theory and wavelet theory are totally different areas of research in mathematics. We briefly describe each theory and apply them respectively to the same problem as an example of application in data mining. Furthermore, we compare the results we obtained from these two different approaches of the same application. Future study along this line of research is also mentioned.

Index Terms—Information system, rough set theory, wavelet, denoising.

I. INTRODUCTION

Rough set theory is a set theory for the study of information systems which are characterized by insufficient and incomplete information. An information system can be regarded as a set-valued system. Some of its attribute values may be subsets of an attribute domain. One of our objectives of this study is to find rules, relationships and classifications of such a system and to develop applications to data mining. Information systems can be represented in various ways. One approach is to use attribute systems in which each system can be interpreted as an ordered pair (U, R) ; where U is a non-empty set of all finite objects under consideration and R is an equivalence relation on U . This approach is called the rough set approach. Rough Set theory originated from Pawlak's seminal work [1]. It has been conceived as a tool to conceptualize, analyze and classify various types of data. It has been developed as a tool to classify objects which are only roughly described. The available information provides a partial discrimination among them although they are considered as different objects. In other words, objects considered distinct could happen to have the same or similar description, as far as a set of attributes is considered. The theory extends the classical crisp set to a rough (or approximate) set by defining lower and upper approximations for any subset of a non-empty universe. It is based on the concept that every object of the universe is associated with some information (data or knowledge). Objects characterized by the same information are considered indiscernible. Thus an elementary set can be any set of all indiscernible entities, and it forms the basic granule of knowledge [2]-[4]. Information granulation is a collection of granules, with a granule being a clump of objects (points) which are drawn toward an object.

Each object is associated with a family of clumps. An unstructured collection of clumps have some mathematical meaning in a crisp world. This has to do with the notion of neighborhood systems. If there is at most one clump per object, then the neighborhood system is defined by a binary relation, and is called a binary neighborhood system. If we assume the binary relation is an equivalence relation, the neighborhood system is a rough set system. If we assume a neighborhood system satisfies certain axioms then the neighborhood system defines a topological space; such neighborhood system is called a topological neighborhood system. In general, we are not able to observe individual objects from U . Hence, we introduce the following structures. The set of all objects which can be classified with certainty as members of X with respect to R is called the R -lower approximation of a set X . The set of all objects which can be classified as possible members of X with respect to R is called the R -upper approximation of a set X with respect to R . The set of boundary region can be classified as members of the R -upper approximation of X but not members of the R -lower approximation of X . A set X is called rough with respect to R if and only if the boundary region of X is nonempty.

The approximations mentioned above can be expressed in terms of granules of knowledge as follows. The lower approximation of a set is a union of all granules which are entirely included in the set; the upper approximation is a union of all granules which have nonempty intersection with a set; the boundary region of a set is the difference between the upper and lower approximation of the set. Fig. 1 illustrates these concepts.

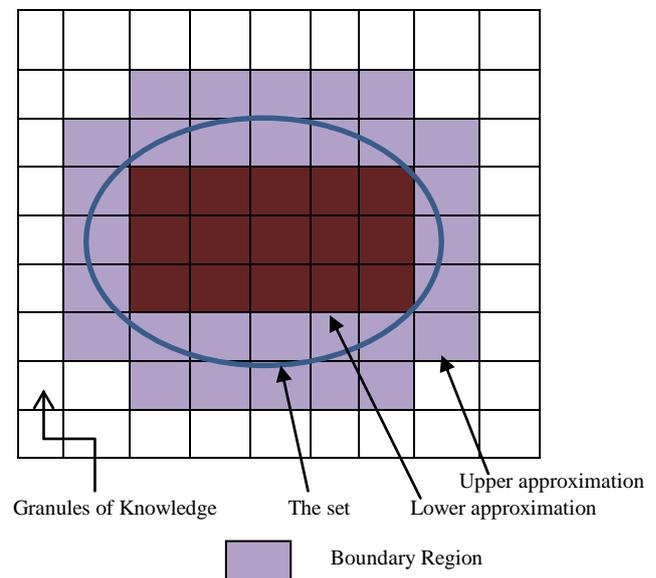


Fig. 1. Illustration of set, region, lower and upper approximations.

Manuscript received February 19, 2014; revised May 13, 2014.

E. B. Lin is with Central Michigan University, Mt. Pleasant, MI, USA (e-mail: enbing.lin@cmich.edu).

Y. R. Syau is with National Formosa University, Yunlin, Taiwan (e-mail: yrsyau@nfu.edu.tw).

Research efforts to advance the classical rough set model have been made by generalizing the Pawlak's approximation space, particularly by exploring the use of a more general binary relation [4]-[7], or replacing the induced partition with a covering [2], [3], [8], [9]. Such a generalization is usually called an approximation space. We were able to provide a clear and more general framework of lower and upper covering approximations [9]. We not only provide duality property but also obtain optimal lower and upper approximations of a covering of an approximation space. In addition, we developed several properties of the so-called total pure reflexive binary neighborhood systems. The techniques we use are essentially tools in topology. Incomplete information systems, reduction of redundant attributes in imprecise information systems, decision-making, data analysis, knowledge presentation, expert systems, data preprocessing, modeling complex system and many other applications can use our framework in rough set theory. The idea is to introduce a set of data, then use our covering approximation techniques to find accuracy, classification, and reduction attributes of a given information system to pursue our objectives for incomplete information systems. Furthermore, we use our methods to couple data and knowledge in data mining and database in various contexts.

Our computational method will be based on wavelets which have been applied to a large variety of data. Wavelet analysis is considered as applied harmonic analysis and has proved to be a powerful technique to analyze, classify or process data efficiently [10]. Wavelets can be used to extract information from many different kinds of data. In this paper, we use wavelets to denoise data based on wavelet coefficients. We intent to handle similar problems as described above by using wavelet method and compare the results obtained via different approaches.

In what follows, we briefly outline some basic elements and fundamental properties of rough set theory and mention an example of some practical application in the second section. In the third section, we give a short description of wavelet theory and apply a wavelet method to the same example in Section II. We conclude with some remarks in the last section.

II. ROUGH SET THEORY

Let U denote a nonempty set (may be finite or infinite) called the universe (of discourse). The power set of U , denoted by 2^U , is the collection of all subsets of U . Let $R \subseteq U \times U$ be an equivalence relation on U . The equivalence relation R partitions the set U into disjoint subsets. Let U/R denote the quotient set consisting of equivalence classes of R .

A. Pawlak's Lower and Upper Approximations

Let I be an index set, and $\beta = \{E_i \subseteq U \mid i \in I\}$ be a partition of U . Pawlak's lower and upper approximations, $\underline{\beta}(X)$ and $\overline{\beta}(X)$, respectively, are defined as follows:

$$\begin{aligned} \underline{\beta}(X) &= \cup\{E_i \in \beta \mid E_i \subseteq X\}, \\ \overline{\beta}(X) &= \cup\{E_i \in \beta \mid E_i \cap X \neq \phi\}. \end{aligned} \quad (1)$$

Equivalently, $\underline{\beta}(X)$ and $\overline{\beta}(X)$ can also be presented as follows [11]:

$$\underline{\beta}(X) = \{x \in U \mid \exists E_i \in \beta : x \in E_i \text{ and } E_i \subseteq X\} \quad (2)$$

$$\overline{\beta}(X) = \{x \in U \mid \forall E_i \in \beta : x \in E_i \Rightarrow E_i \cap X \neq \phi\} \quad (3)$$

B. Covering Lower and Upper Approximations

Let $C = \{C_i \subseteq U \mid C_i \neq \phi, i \in I\}$, be a covering of U . The pair (U, C) is called a covering approximation space. A covering C is called a finite covering if it is a finite family.

Definition 1. [11] Let $C = \{C_i\}$ be a covering of U . For any $X \subseteq U$, the C -lower and C -upper approximations, $\underline{C}(X)$ and $\overline{C}(X)$, respectively, are defined as follows:

$$\underline{C}(X) = \{x \in U \mid \exists C_i \in C : x \in C_i \text{ and } C_i \subseteq X\}$$

$$\overline{C}(X) = \{x \in U \mid \forall C_i \in C : x \in C_i \Rightarrow C_i \cap X \neq \phi\} \quad (4)$$

C. Topological Neighborhood System

To define a topology on U , there are many other ways to set up axioms, including the so-called Kuratowski closure axioms (KC-axioms) and the topological neighborhood axioms (TN-axioms) that can be used to define the structure of topology. Based on the topology the universe has, one can define the topological neighborhood system.

A neighborhood system on U is a mapping $NS : U \rightarrow 2^{2^U}$ which is defined by assigning to each x of U a nonempty collection $NS(x)$ of subsets of U . Such a nonempty collection $NS(x)$ and each of its members is called a neighborhood system at x and a neighborhood of x , respectively.

A neighborhood system $NS : U \rightarrow 2^{2^U}$ will be called a topological neighborhood system, or a topology, on U if for each $x \in U$, $NS(x)$ satisfies the following topological neighborhood axioms:

(TN 1) If $N \in NS(x)$, then $x \in N(x)$.

(TN2) If N, M are members of $NS(x)$, then $N \cap M \in NS(x)$.

(TN 3) Superset condition: If $M \supseteq N$ for a nonempty $N \in NS(x)$, then $M \in NS(x)$.

(TN 4) If $N \in NS(x)$, then there is a member M of $NS(x)$ such that $M \subseteq N$ and $M \in NS(y)$ for each $y \in M$ (that is, M is a neighborhood of each of its points).

D. Total Pure Reflexive Neighborhood Systems

As it is known that a covering C of U can be regarded as a special form of neighborhood system which we call a covering neighborhood system. We study the structure of the covering approximation space (U, C) via a special kind of

neighborhood systems, called total pure reflexive neighborhood systems, which include covering neighborhood systems as a special case.

A neighborhood system $NS : U \rightarrow 2^{2^U}$ is called a total pure reflexive (TPR) neighborhood system on U iff

$$x \in N, \forall N \in NS(x).$$

Definition 2. Let $NS : U \rightarrow 2^{2^U}$ be a TPR neighborhood system. For each subset X of U , the lower and upper approximations of X , $\underline{NS}(X)$ and $\overline{NS}(X)$, respectively, are defined as follows:

$$\begin{aligned} \underline{NS}(X) &= \{x \in U \mid N \subseteq X, \text{ for some } N \in NS(x)\} \\ \overline{NS}(X) &= \{x \in U \mid N \cap X \neq \phi, \forall N \in NS(x)\} \end{aligned} \quad (5)$$

The lower and upper approximations are mutually dual in the sense that

$$\overline{NS}(X) = U - \underline{NS}(U - X), \quad \forall X \subseteq U \quad (6)$$

Theorem 1. Let $NS : U \rightarrow 2^{2^U}$ be a neighborhood system in which each $x \in U$, $NS(x)$ satisfies axioms (TN 1) and (TN 2), and let

$$T = \{X \subseteq U \mid \underline{NS}(X) = X\}$$

Then:

1. The family T forms a topology on U .
2. The upper approximation $\overline{NS} : 2^U \rightarrow 2^U$ is a preclosure on U .
3. The lower approximation $\underline{NS} : 2^U \rightarrow 2^U$ satisfies
 - a) $\underline{NS}(U) = U$
 - b) $\underline{NS}(X) \subseteq X, \forall X \subseteq U$.
 - c) $\underline{NS}(X \cap Y) = \underline{NS}(X) \cap \underline{NS}(Y), \forall X, Y \subseteq U$.

E. Total Pure Reflexive Binary Neighborhood Systems

We call a TPR neighborhood system $NS : U \rightarrow 2^{2^U}$ as a TPR binary neighborhood system if each $x \in U$, $NS(x)$ is a singleton. It is clear that $NS(x)$ satisfies the first two TN-axioms. We will use the element instead of singleton. So $NS : U \rightarrow 2^{2^U}$ will be replaced by $BN : U \rightarrow 2^U$, and this gives

$$\begin{aligned} \underline{BN}(X) &= \{x \in U \mid BN(x) \subseteq X\}, \forall X \subseteq U \\ \overline{BN}(X) &= \{x \in U \mid BN(x) \cap X \neq \phi\}, \forall X \subseteq U. \end{aligned} \quad (7)$$

Then, by Theorem 1, the collection $T = \{X \subseteq U \mid \underline{BN}(X) = X\}$ is a topology on U . It follows that T is an Alexandroff topology if it is shown that the intersection of the members of any nonempty subfamily of

T is a member of T .

Theorem 2. Let $BN : U \rightarrow 2^U$ be a TPR binary neighborhood system, and let

$$T = \{X \subseteq U \mid \underline{BN}(X) = X\}$$

Then:

1. The upper approximation $\overline{BN} : 2^U \rightarrow 2^U$ is a preclosure on U .
2. The lower approximation $\underline{BN} : 2^U \rightarrow 2^U$ satisfies
 - a) $\underline{BN}(U) = U$.
 - b) $\underline{BN}(X) \subseteq X, \forall X \subseteq U$.
 - c) $\underline{BN}(X \cap Y) = \underline{BN}(X) \cap \underline{BN}(Y), \forall X, Y \subseteq U$.
3. The family T forms an Alexandroff topology on U .

In an information system, each attribute corresponds to an equivalence relation. The following example is an application of rough set theory.

Example 1. We consider an example from [11]. Briefly, it is shown in Table I which includes Solar Energy, Volcanic Activity and Residual CO_2 as condition attributes, and Temperature as a decision attribute.

TABLE I: CONDITION AND DECISION ATTRIBUTES

Fact	Days	Solar energy	Volcanic activity	Residual CO_2	Temperature
1	20	Medium	High	Low	High
2	30	High	High	High	High
3	34	Medium	Low	High	Low
4	70	High	High	Medium	High
5	90	Medium	Low	High	High
6	120	Low	Low	Low	Low

Using rough set theory, some of its conclusions are [11]:
 Low solar energy certainly causes low temperature;
 Medium solar energy and high volcanic activity or high solar energy certainly causes high temperature.

III. WAVELET BASED METHOD

In wavelet theory, one usually deals with wavelets and scaling functions which are key elements [10]. The wavelet function is defined by building a sequence upon scaling functions generated by $\varphi(x)$. Choosing some suitable sequence, $\{a_p, p \in Z\}$, we obtain the following dilation equation,

$$\varphi(x) = \sum_p a_p \varphi(2^j x - p) = \sum_p a_p \varphi_{j,p}(x). \quad (8)$$

With this scaling function, we have a nested of subspaces $\{V_j, j \in Z\}$ of $L^2(R)$ which is defined as:

$$V_j = \overline{Span\{\varphi_{j,p}(x)\}_p}, \quad j \in Z \quad (9)$$

This means that for any function $f(x) \in V_j$ it can be expressed as:

$$f(x) = \sum_p \alpha_p \varphi_{j,p}(x). \quad (10)$$

If the basis functions of a subspace are orthogonal at the same level, then a given function $f(x) \in V_j$ can be expressed as follows:

$$f(x) = \sum_p \langle f, \varphi_{j,p} \rangle \varphi_{j,p}(x) \quad (11)$$

where

$$\langle f, \varphi_{j,p} \rangle = \int_{-\infty}^{\infty} f(x) \varphi_{j,p}(x) dx. \quad (12)$$

If the nested sequence of the subspaces $\{V_j, j \in Z\}$ has the following properties then it is called a multiresolution analysis (MRA):

- 1) $V_j \subset V_{j+1}$,
- 2) $\bigcap_{j \in Z} V_j = \{0\}$,
- 3) $\overline{\bigcup_{j \in Z} V_j} = L^2(\mathbb{R})$,
- 4) $f(x) \in V_n \Leftrightarrow f(2x) \in V_{n+1}$,
- 5) There exists a function $\varphi \in V_0$ such that $\{\varphi(x-k), k \in Z\}$ is an orthogonal basis for V_0 .

The wavelet function is constructed in the orthogonal complement of each subspace V_j in V_{j+1} which is denoted by W_j . Therefore $V_{j+1} = V_j \oplus W_j$. Since

$$V_j \rightarrow \begin{cases} 0, & \text{as } j \rightarrow -\infty \\ L^2(\mathbb{R}), & \text{as } j \rightarrow \infty \end{cases},$$

we have $V_{j+1} = V_j \oplus W_j$ and $L^2(\mathbb{R}) = \bigoplus_{j=-\infty}^{\infty} W_j$. The

set $\{\psi_{j,p}(x) = \psi(2^j x - p)\}$ forms a basis for W_j , and can be obtained from the following equation:

$$\psi(x) = \sum_p b_p \varphi_{j,p}(x), \text{ for some } b_p. \quad (13)$$

The orthogonality of W_j on V_j means that any member of V_j is orthogonal to the members of W_j , that is,

$$\langle \varphi_{j,p}, \psi_{j,k} \rangle = \int \varphi_{j,p}(x) \psi_{j,k}(x) dx = \delta_{p,k}. \quad (14)$$

In fact, scaling function and wavelet have the following properties:

$$\begin{aligned} \int \varphi(x) dx &= 1, \\ \int x \varphi(x) dx &= \frac{1}{2} \sum_p p a_p, \\ \int x^r \psi(x) dx &= 0, \quad r = 0, \dots, N-1, \end{aligned} \quad (15)$$

where $[0, N]$ is the compact support of $\varphi(x)$ and $\psi(x)$.

Wavelet transform is defined by using wavelets. Wavelet coefficients can be obtained by applying wavelet transform to a given image. Depending on the purpose of the application, there are many different ways to use these coefficients. Denosing is a process to remove noise from a given image. One can remove certain small coefficients from the image to serve the purpose.

More precisely, wavelet transform and wavelet coefficients are defined as follows.

$$[W_\Psi f](a, b) = \frac{1}{\sqrt{|a|}} \int \overline{\Psi\left(\frac{x-b}{a}\right)} f(x) dx, \quad (16)$$

$$c_{j,k} = [W_\Psi f](2^{-j}, k2^{-j}). \quad (17)$$

As an application of denosing image, we consider an image in Fig. 2 below which is obtained by converting the data in Table I. We then apply wavelet denosing method to obtain the image in Fig. 3.

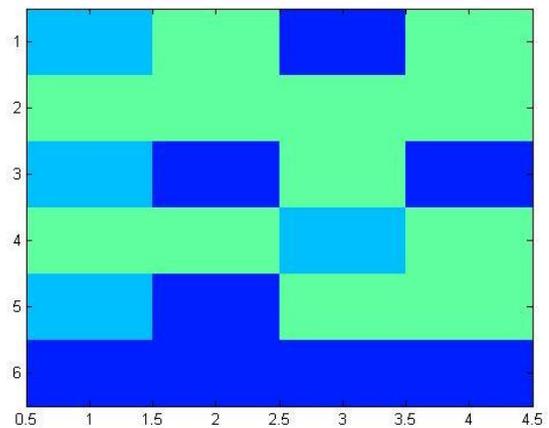


Fig. 2. The data in Table I is converted to this image.

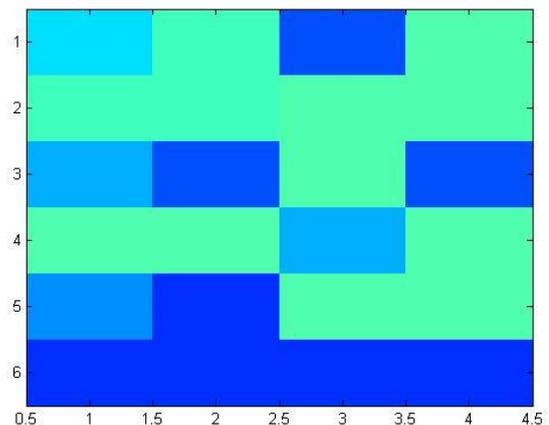


Fig. 3. Denoised image of Fig. 1.

If we denote the image in Fig. 2 by A and Fig. 3 by B, then the difference between noisy and denoised image is $C=A-B$ which is as follows.

$$C = \begin{bmatrix} -2 & 2 & -3 & 1 \\ 2 & 2 & 1 & 1 \\ 1 & -3 & 1 & -3 \\ 1 & 1 & 1 & 1 \\ 3 & -1 & 1 & 1 \\ -1 & -1 & -1 & -1 \end{bmatrix} \quad (18)$$

The smaller variations of the difference of rows between the original and denoised images indicate that the phenomena are quite normal. Consequently, this results in similar conclusions in Example 1.

IV. CONCLUSION

Although rough set theory and wavelet theory are originated from different areas of research in mathematics, the former is in topology and the latter is in analysis, we can use them to work on the same application and obtain similar results. As it is shown in this paper, we compare the results in two different approaches via rough set theory and wavelet method respectively, we obtain similar conclusions. It would be interesting to further study other applications such as missing data problems and many other problems in data mining.

REFERENCES

[1] Z. Pawlak, "Rough sets," *International Journal of Computer and Information Science*, vol. 11, 341-356, 1982.
 [2] T. Y. Lin, "Granular models for partitions, coverings, and Pre-Topologies," in *Proc. the IEEE International Conference on Systems, Man and Cybernetics*, 2007, pp. 3767 - 3771.
 [3] T. Y. Lin, "Granular computing on partitions, coverings and neighborhood systems," *Journal of Nanchang Institute of Technology*, vol. 5, no. 2, 2006, 1-7.

[4] T. Y. Lin, "Granular computing on binary relations I: data mining and neighborhood systems," in *Rough Sets and Knowledge Discovery*, A. Skowron and L. Polkowski, Eds., Physica -Verlag, 1998, pp. 107-121.
 [5] Z. Bonikowski, E. Bryniarski, and U. Wybraniec-Skardowska, "Extensions and intentions in the rough set theory," *Information Sciences*, vol. 107, pp. 149-167, 1998.
 [6] R. Slowinski and D. Vanderpooten, "A generalized definition of rough approximations based on similarity," *IEEE Transactions on Data and Knowledge Engineering*, vol. 12, pp. 331-336, 2000.
 [7] Y. R. Syau and L. Jia, "Generalized rough sets based on reflexive relations," *Communications in Information and Systems*.
 [8] R. B. Barot and T.Y. Lin, "Granular computing on covering from the aspects of knowledge theory," in *Proc. Fuzzy Information Processing Society, NAFIS 2008*.
 [9] Y. R. Syau and E. B. Lin, "Neighborhood systems and covering approximation spaces," *Knowledge-Based Systems*, 2014.
 [10] E. B. Lin and Y. Al-Jarrah, "A wavelet based method for the solution of Fredholm integral equations," *American Journal of Computational Mathematics*, vol. 2, pp. 114-117, 2012.
 [11] Z. Pawlak, "Rough sets and intelligent data analysis," *Information Sciences*, vol. 147, pp. 1-12, 2002.



En-Bing Lin received his Ph. D. in mathematics from Johns Hopkins University. Currently, he is the chair and a professor of mathematics at Central Michigan University, USA. He has taught and visited at several institutions including Massachusetts Institute of Technology, University of Wisconsin-Milwaukee, University of California, Riverside, University of Toledo, UCLA, and University of Illinois at Chicago.

His research interests include data analysis, image processing, applied and computational mathematics, wavelet analysis and applications, and mathematical physics. He has supervised a number of graduate and undergraduate students. He serves on the editorial boards of several mathematics journals and several academic committees of regional and national associations. He has organized several special sessions at regional IEEE conference and American mathematical society national and regional meetings.



Yu-Ru Syau received her PhD degree from the University of Illinois at Chicago, and now is the dean of the College of Management and a professor of Information Management at the National Formosa University, Taiwan. She also served as the chair of the Department of Information Management. Her teaching and research interests are in granular computing and the general areas of operations research.