Volume 3, Issue 7, July 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Identifying the Vague Regions by Using Covering Based Rough Sets

Y.SaiSandeep , P. Vidya Sagar Reddy, C.Manoj, K. Ajanthaa Lakkshmanan
School of Computer Science
VIT University, India

Abstract— Spatial database is a collection of data concerning objects. The real world is bounded in uncertainties and any aspect to model the world should include some mechanisms to handle Uncertainties .so for management of this uncertainty we use the Rough Set Theory .In Rough Sets we consider equivalence or indiscernibility relations and approximation regions for handling these uncertainties. A slight improvement to rough sets is covering based rough sets to find more rough in the given vague data. Both normal and covering based rough sets uses lower and upper approximation to find the rough in the data. But the covering based rough sets uses, covering to find the upper approximation region which possibly holds more rough. So it is more powerful mathematical tool compared to rough sets.

Keywords— Buffer Spatial Databases, uncertainties, lower and upper approximation regions, equivalence relation, covering, roughness

I. INTRODUCTION

Spatial Database'sacquisition and reasoning about spatial data are current active researching area. The spatial database comprises of data related to space, which can be represented by points, lines and regions. Rapid growth in the development of the technology makes spatial database complex and big. Due to the conditions on the current technologies there are uncertainties in most of the spatial data. These uncertainties may take different forms like understanding the enterprise or in the quality or understanding the meaning of the data. These uncertainties may directly affect quality of the spatial data mining or indirectly affect. If there is uncertainty in model, which affects the uncertainty in entities or attributes describing them.So, there is a need to manage these uncertainties in databases in real world applicationssince the real world bounded with uncertainties. If we develop a mechanism, that should include a model to solvethese uncertainties. One way toreducethese uncertainties is by using "Rough sets".

ROUGH SET BASICS

The concept "Rough sets" was introduced by Pawlak in 1970. It is the mathematical technique to handle uncertainties and to identify cause- effect relationships, whichhas applications on several research areaslike logic of knowledge discovery. It converts uncertain data into certain and definable data. Rough Sets providemathematical tools to discover hidden patterns in data.It is also used for feature selection, feature Extraction, generation and pattern extraction. It is also used to identify the partial or total dependencies in data. Rough sets also eliminate redundancies.

It is based on two concepts.

1) Indiscernible Relation or Equivalence relation.
2) Approximation regions.

Indiscernible relation used to partition the universe into equivalence classes. Lower and upper approximation regions distinguish between certain and possible or partial regions. Indiscernibility is the inability to distinguish two or more values which can arise from lack of exactness in measurement.

Following are the terms that we come across this paper.

U- A non-empty universe

R- Indiscernible Relation or Equivalence relation

A binary relation R (subset of)X*X which satisfies

1) Reflexive XRX X belongs U
2) Symmetric If XRY then YRX X,Y belongs to U
3) Transitive If XRY,YRZ then XRZ X,Y,Z belongs to U

A= (U,R) is an ordered pair, called approximation space

$[x]_R$ is an equivalence class of R containing x

The equivalence relation R partitions the universe into equivalence classes called elementary sets. So the by union of all equivalence classes we can again retrieve the universe.

Given the X "subset of" U, X can be defined as definable sets, So the

Lower approximation region of X in A- which holds the certain values of data set and is denoted by $\underline{R}$X.

Upper approximation region of X in A- which holds the possible or partial values of data set and is denoted by$\overline{R}$ X.

If lower and upper approximation regions are not equal, then X is rough with respect to R

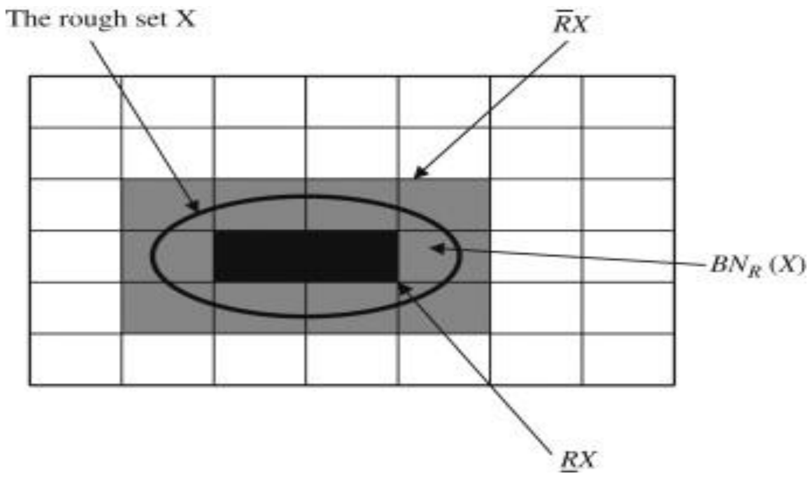

Figure1 Representation of Rough set model

If lower and upper approximations are equal then X is called R definable.
Given lower and upper approximation regions we can define,
R-positive region of X as ($POS_R(X)$) $\underline{R}$X.

R-negative region of X as ($NEG_R X$) U-$\overline{R}$X.

R-Boundary region of X as (BNR(X)) $\overline{R}$X - $\underline{R}$X.

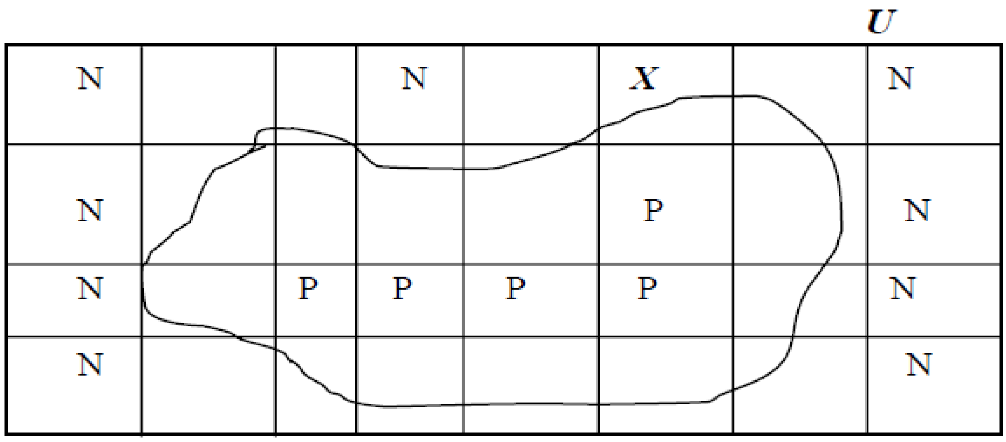

Figure 2 Representation of approximation regions.

The above figure indicates the negative and positive regions. The equivalence classes which are represented as squares. The elements in the lower approximation region of X, $POS_R(X)$, are denoted by P, and the elements in the negative region $NEG_R X$ are represented by N. The remaining elements belong to universe.

COVERING BASED ROUGH SETS BASICS

The covering based rough sets are identical to rough sets; the only difference is covering based rough sets uses covering of the universe instead of partitions or equivalence classes in rough sets. Covering refers to family of subsets in the universe. The distinctness of covering and partition is, in partition the sets are disjoint, but this may not be in case of covering. The data objects in one subset may repeat in other subset of the covering. The covering based rough sets also uses upper and lower approximation regions to find the rough in the given data. It is applied to multivariate attributes in the databases. There are about eight types of covering based rough sets are proposed, in which the lower approximation region is same but there is change in the upper approximation region. First type of covering generalized rough sets uses reduct, second type uses exclusion and third type uses minimal description. Here we mainly focusing on second type generalized rough sets.

II. Literature Survey

1. ZdzislawPawlak -proposed the basic concepts of rough sets, lower and upper approximation regions, the indescribability relation and the applications of rough sets to draw the conclusions like in telecommunications.
2. William Zhu- described the properties of rough sets to find out the constraints under which the second type covering based rough sets will satisfy these properties.

3. Jianguo Tang, Kun She, William Zhu- proposed that in the first and third type covering based rough sets the lower approximation regions gives more data than original lower approximation regions and upper approximation regions gives less data than original upper approximation regions in some special situations.
4. ChengyiYu,Fan Min, William Zhu- described the free matroidal structure of covering based rough sets and the relationships between the reducible elements of covering based rough sets to the reducible matroid in matroid theory.
5. Jianguo Tang, Kun She, William Zhu – described the refinement of covering based rough sets by reducing the size of the covering elements in the covering, so that the description of the object is more accurate.

III. IMPLEMENTATION

ROUGH SET METHODOLOGY

Rough set is a mathematical tool to find the vagueness in the given data or region.The intersection of lower approximation region and upper approximation region gives the roughness.Lower approximation region is defined as

$$\underline{R}X = \{x \in U \mid [x]_R \subseteq X\}$$

Upper Approximation Region is defined as

$$\overline{R}\,X = \{x \in U \mid [x]_R \cap X \neq \emptyset\}.$$

Mathematically the rough can be defined as

Example for Roughsets

Table I : Information System

Objects	S1	S2	S3	S4	S5
X1	1	0	1	0	0
X2	0	1	0	1	0
X3	2	0	1	0	1
X4	1	1	1	1	0
X5	1	0	1	0	0
X6	2	0	1	0	1
X7	1	1	1	1	0
X8	1	2	0	1	2
X9	2	0	1	0	1

Equivalent-Classes={{X1,X5},{X2},{X3,X9},{X4,X7},{X6},{X8}}

The Target class={X1,X2,X3,X4,X5,X9,X10}

Lower approximation Region for the above equivalence classes from the eq

$\underline{R}$X ={{X1,X5},{X3,X9},{X2}}

={X1,X2,X3,X5,X9}

Upper approximation Region for the above equivalence classes from the eq

$\overline{R}$ X ={{X1,X5},{X2},{X3,X9},{X4,X7}}

={X1,X2,X3,X4,X5,X7,X9}

COVERING BASED ROUGH SET METHODOLOGY

Covering based rough set uses covering instead of partitions. A covering is a family of subsets.

C = {k1,k2,……,kn}

Where k1,k2,…,kn are the subsets of universe U such that k1 U k2 U……..U kn=U(universe).

The covering elements contains at least one data object should be common in any one of the subsets of C. If the subsets are disjoint then covering becomes partitions.

Let C ={{X1,X5,X3},{X2},{X3,X7,X9},{X4,X7,X6},{X6},{X8}}

From the above covering C

k1={X1,X3,X5},k2={X2},k3={X3,X7,X9},k4={X4,X6,X7},k5={X6},k6={X8}

The lower approximation region for the above covering elements for the target set given above is

$\underline{R}$X ={{X1,X3,X5}U{X2}}

={X1,X2,X3,X5}

The upper approximation region for the above covering elements for the target set give above is

$\overline{R}$ X ={{X1,X3,X5}U{X2}U{X3,X7,X9}U{X4,X6,X7}}

={X1,X2,X3,X4,X5,X6,X7,X9}

IV. EXPERIMENTAL RESULTS

From the above two methodologies discussed above, the identification of roughness or the vagueness in the given data increases from rough set to covering based rough set.

From the above example

Roughness according to rough sets $\overline{R}$ X - $\underline{R}$X = {X4,X7}

Roughness according to covering based rough sets $\overline{R}$ X - $\underline{R}$X = {X4,X6,X7}

The two results clearly describes that roughness in covering based rough sets having more elements in the set compared to the normal rough sets. So the identification of rough is more in covering based rough sets. By these observations we can say that covering based rough sets is a powerful tool compared to normal rough sets

V. PERFORMANCE EVALUATION

The covering based rough sets is an extension to rough sets. Both methods use lower and upper approximation regions. the second type generalized rough sets which we mainly focused here uses covering to find the lower and upper approximation regions, through which identifies the more rough than normal rough sets. So we can conclude that covering based rough sets is a powerful tool. There is a future extension to refine the covering based rough sets by reducing the number of elements in the covering subsets so that the object can be accurately described.

REFERENCES

[1] Z. Bonikowski, "Algebraic Structures of Rough Sets," Proc. Int'l Workshop Rough Sets and Knowledge Discovery: Rough Sets, Fuzzy Sets and Knowledge Discovery, W. Ziarko, ed., pp. 243-247, 1994.

[2] .W. Zhu and F.-Y. Wang, "Covering based granular computing for conflict analysis," in IEEE International Conference on Intelligence and Security Informatics, vol. 3975 LNCS, San Diego, CA, United states, 2006, pp. 566–571.

[3] .Y.-L. Zhang, J. Li, and W.-Z. Wu, "On axiomatic characterizations of three pairs of covering based approximation operators," Information Sciences, vol. 180, no. 2, pp. 274– 287, 2010.

[4] Z. Pawlak, "Rough sets," International Journal of Computer Information and Science, vol. 11, pp. 341–356, 1982.

[5] W. Zhu and F. Wang, "Reduction and axiomization of covering generalized rough sets," Information Sciences, vol. 152, no. 1, pp. 217–230, 2003.

[6] W. Zhu, "Relationship between generalized rough sets based on binary relation and covering," Information Sciences, vol. 179, no. 3, pp. 210–225, 2009.

[7] K. Qin, Y. Gao, and Z. Pei, "On covering rough sets," inRSKT 2007, ser. LNCS, 2007, pp. 34–41.

[8] W. Zhu and F.-Y. Wang. Binary relation based rough set. In IEEE FSKD 2006, volume 4223 of LNCS, pages 276–285, 2006.

[9] W. Zhu and F.-Y. Wang. Covering based granular computing for conflict analysis. In IEEE ISI 2006, volume 3975 of LNCS, pages 566–571, 2006.

[10] Z. Bonikowski, E. Bryniarski, and U. Wybraniec-Skardowska, "Extensions and Intentions in the Rough Set Theory," Information Sciences, vol. 107, pp. 149-167, 1998.

