7th PACIFIC RIM REAL ESTATE SOCIETY MEETING
ADELAIDE – 21 - 24 January 2001 AUSTRALIA

Paper Pubblicato sul CD-Rom of Paper

Appraising Properties with Rough Set Theory[i]

By

Maurizio d'Amato
Researcher
1st School of Engineering
Polytechnic of Bari

Home Address: 272, Calefati St., 70122 Bari, Italy
Home Telephone: 39 080 5216009
E-mail : **madamato@yahoo.com**

Abstract

It is well known that the main approaches to property appraisal are income, market and cost. The work is focused on a methodology to analyze unprecise and multiattribute information defined rough set (Pawlak, 1982), that will be proposed in this work as a further method inside market approach. The theory will be analyzed showing a practical application on a group of property transactions obtained from a local real estate broker. The application will show interesting results. A comparison between this method and more widespread statistical tools will be highlighted.

INTRODUCTION

The traditional approaches of Market Cost and Income, are the bases of property appraisal. This work tries to show how is possible to use market data to appraise properties without defining a model. In fact, Rough set reaches the knowledge through a simple knowledge of data.

The Rough Set Theory has been applied to several fields[ii]. The work is organized as follows: in the following paragraph will be offered a brief history and a general introduction to this technique. The second paragraph will show an application of this method to residential property market of Bari, a southern Italian town. The last paragraph will offer some final remarks and future directions of research.

1.Rough Set Theory: a brief overview

There are several tools developed in order to analyze uncertain problems and imprecise information. Zdzislaw Pawlak introduced this theory in two famous works[iii]. It is assumed that, each element (or object) of a universe could be associated to some information which are referred to several attributes describing the elements. If a real property is considered as an element, the information available will be the specific features (attributes) of the real property sold. If the real properties have the same attributes, then they will be considered *indiscernible* at a certain information level. It has been highlighted that "…*indiscernible relationship…is the mathematics foundation of rough sets theory, it is the brick on which is built the building of knowledge of reality.*"[iv].

An object or element has a feature Certainly, Possibly and Certainly not. For this reason it has been defined that rough set logic correspond to a fuzzy logic with a three-value membership function[v].

Rough Set Data Analysis are based on internal knowledge originated by data

Each indiscernible element is defined as an *elementary set.* A subset of universe can be defined more precisely in terms of granularity or as a union of elementary sets approximately. In the latter case this subset will be defined through two ordinary sets. The former will be called positive region or lower approximation, while the second will be defined as " *possible region*" or upward approximation. The rough set will be defined through these approximations[vi], and the differences between positive and possible regions will be represented by " *boundary region*" of rough set.

Positive of Y set is composed by all the elements included in Y. On the other side upward approximation is defined by the elementary sets with a no-empty intersection with Y. The elements could belong or not to Y. According to a level of information it is not certain if some elements belong or not to Y. An imprecise concept can be described with a couple of precise concepts: positive or possible region.

This seems to be useful in the comparison process inside market approach. The property to be estimated is described through lower or upward approximations.

In fact the objects "…*belonging to the same category are not distinguishable, which means that their membership status with respect to an arbitrary subset of the domain may not be clearly definable. This fact leads to the definition of a set in terms of lower and upper approximation…*"[vii]

Furthermore, it is possible to highlight the causal relationship between the available data. It is possible to use both qualitative and quantitative data without controlling their consistence. For this reason it is not necessary to analyze and to eliminate some information because both bad and good data are useful. The importance of the attributes is revealed by an analysis of the problem. The results will be defined through decision rules like "if…then" based.

The information on the elements, which compose the universe, will be offered in an "information table". In the row there will be several units and in the column will be listed the different attributes. For example in the row will be considered one only property similar to the one to be estimated. In the column there will be the attributes that could be considered important to explain the property value. Each cell will have the quantitative or qualitative evaluations of the attribute for an element. For example the presence of air condition system could be estimated with economic age life method and put inside a cell. The information table S is

$$\mathbf{S = < U, Q, V_q, f >} \quad \textit{(a)}$$

Where U is the universe or a finite set of element (or properties), Q is a finite set of attributes or features (characteristic of real properties), V_q is the domain of attribute q (if we consider the variable parking a dummy variable could be considered) and f is the information function[viii] that could be described as:

$$f : U \times Q \rightarrow V \text{ and } f(x, q) \in V_q \quad \forall\, q \in Q \text{ and } x \in U \quad \textit{(b)}$$

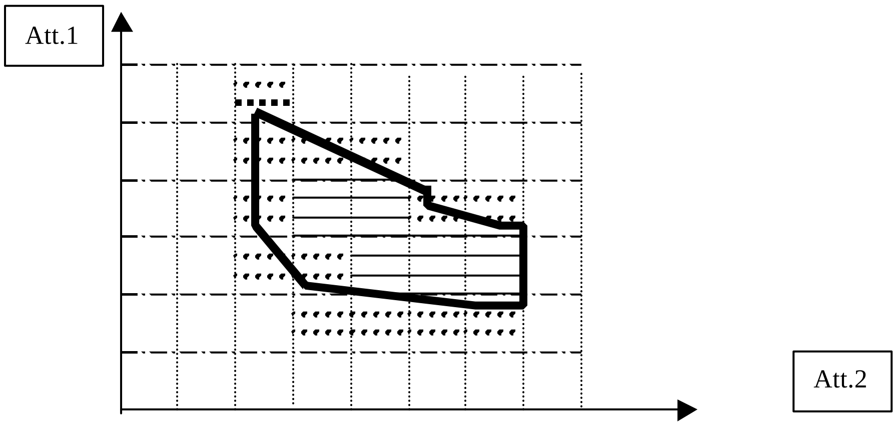
Vectors will describe all the elements of U. This vector, also called description, will show the value that an attribute assume for x inside Q set and can be defined as $Des_Q(x)$.
The object $x \in U$ will be described using a no empty subset $P \subseteq Q$. For each subset of features P there is a indiscernible relation on U that could be indicated as Ip, where

$$Ip = \{(x, y) \in UxU: fq(x) = fq(y), q \in P\}$$

This binary indiscernment[ix] is an equivalent relationship. The pair (x, Ip) defines an *approximation space*. If $(x, y) \in Ip$ then it will be possible to say that x and y are P-indiscernible. Furthermore, if P= Q the elementary Q sets are called atoms.
Another important concept is union. It is possible to define *upward approximation* or P_UX the subset of U composed by the elements belonging to P that have one element at list similar to X set. *Downward approximation* of X or P_LX is the subset of U whose elements belong to P elementary set *included in X set,* and only them.
The difference between these sets is defined as X boundary defined as $BN_p(X)$ and could be mathematically described as

$$BN_p(X) = P_UX - P_LX$$

If the frontier is empty then X is the union of several ordinary sets defined through the union of several elementary P sets.
« *The lower approximation is a description of the domain objects which are known with certainty to the belong to the subset of interest, whereas the upper approximations is called a rough set.* »[x]
It is possible to draw a graphical representation of lower and upper approximations. In the graphic below the geometric figure is described through two different features 1 and 2.



Describes Boundary Region

Describes Positive Region or Lower Approximation $P_L X$

Describes Differences between Possible and Positive regions or between the Upward $P_U X$ and the Lower Approximation

If $X \subseteq U$ is given by a predicate (attribute) P and $x \in U$ then

a. $x \in P_L X$ means that x certainly has property P

b. $x \in P_u X$ means that x possibly has property P

c. $x \in U/P_u X$ means that definitely does not have property P

The area of uncertainty extends over $P_u X / P_L X$ while the area of certainty is $P_L X \cup - P_u X$.

If the boundary is not empty X is a rough set on P that can be defined through P upward union and downward union. In this way we define a non-perfect reality. Information seems not to be perfect and could be represented by the P-elementary sets, whose elements are P-indiscernible because they have the same description in term of P attributes.

Many dimensions influence the granularity of information as quality of attributes, number of attributes, domains of each attribute. It is obvious that this procedure is strongly dependent on the quality of information, on the capability to classify the information and the level of confidence and knowledge of the problem.

The information used in the appraisal process could be correct or not. It is possible to define a minimum subset of attribution (called reduct) which allows the evaluator to have the same approximation of U of the complete set of features of P.

If $P \subseteq Q$ e $p \in P$ a feature is not important in P if $Ip = I_p - \{p\}$; otherwise it is in P. P is defined orthogonal if all attributes are important.

P set is independent if all the attributes are all important. The subset P’ is a reduct of P if P’ is independent and I_p’ = I_p

In an information table there could be more than one reduct of P and *core* of P is the set containing all the indispensable attributes of P. The core is inside each reduct of P then is considered as the most important subset of attributes of Q. No elements of this sub set can be removed without diminishing information quality.

An information table become a decision table if the attribute are divided in conditional attribute (C set) and decisional attribute (D set)

This table highlights the causal relationship between conditional attribute and decisional attribute. Lowest numbers of conditional attribution are chosen to take a decision (to appraise) with little information.

Decision rules are based on logic preposition as “ if..then” . The former part of preposition is referred to one or more conditional attributes and the second part of preposition is represented by the set of decisional features.

There are two general kinds of decision rules. The former is the “exact decisional rule” or deterministic where decisional set contains conditional attribute, the latter is the “approximated decision rule” in which some conditional attributes are contained inside decisional set.

The logic prepositions if then allow the appraiser to build a preference system based on property market data.

The granularity of the system become higher and could be due to a chance when information is based on few observations.

As it is possible to see “…*Using this attribute one can build a rule that classifies a given training set 100% correct; need less to say, the rule will not perform on an independent test set…*”[xi]. For this reason specific significance tests have been developed essentially based on randomization technique[xii]. Furthermore a criterion for model selection based on minimum description length principal[xiii] define the better selection of the model to explain d attribute.

It is always possible to control and improve the appraisal processing each phase,

2. Appraising Properties with Rough Set Theory

The property data considered in the example are essentially continuous scale based. It has been highlighted how this procedure works with continuous data.[xiv]

In order to show how this methodology works thirty transactions of residential properties[xv] have been considered in the town of Bari. They have been observed and organized in ten classes. Property price has been "approximated" through three attributes. In this way U set will be composed by the ten "*classes*", and the Q set is defined through the most important attributes influencing house price in that market. These transactions are referred to properties similar to the one to be estimated and will allow the appraiser to define the most important attribute and the "*decision rule*» to appraise a residential property.

It must be highlighted that "*operationalisation*" of *Object* **->** *Attribute* assumes that each object has exactly one value of each attribute at a given time and the observation of this value is without error[xvi]. This method rely, also, on the principle of indifference[xvii] which states that in the absence of further knowledge all basic events are "assumed to be equally likely"

The appraiser does the delicate choice of features and their measure and if one or more important attributes are not considered the results of the appraisal process will be not reliable.

A great help come from several informatic tools which make easy the calculation and the definition of rules using a great amount of data[xviii]. Information table could be defined similar to "sales summary grid" table of MCA. The presence or not of the parking and the commercial sqm of the properties have been taken in account. The first is considered as a dummy variable and the second in a cardinal scale.

Properties	**SQM**	**PARKING**	**PRICE**
1	90	NO(or 0)	A
2	90	YES (or 1)	B
3	90	YES (or 1)	A
4	100	NO(or 0)	B
5	100	NO (or 0)	B
6	90	NO (or 0)	A
7	110	YES (or 1)	D
8	110	NO (or 0)	B
9	110	NO (or 0)	B
10	110	YES (or 1)	D

Table 1- Information Table Referred to 7 Real Property Transactions

In this informative table it is possible to observe U = {1,2,3,4,5,6,7,8,9,10} Q = { SQM, PARKING, PRICE} and the table represents the information function f (x, q). For example if it is considered f (5,PARKING) = NO or 0

The value of the attributes V_q will vary depending on several different scales. The example shows a dichotomy variable and a cardinal scale for the square meters. The attribute "property prices" are assumed to be defined in interval as in the following table:

Property Prices	**Intervals**
200 – 205	**A**
205 – 210	**B**
210 – 215	**C**
215 – 220	**D**

Table 2- Interval Scale for Property Prices

These intervals are originated by the previous defined thirty data concerning property transactions analyzed. The intervals depend on the data and the necessity of appraisal process. In an appraisal process in which there is the necessity to define a closer rang of value intervals will have a little difference between minimum and maximum value. This will be possible according to the data that will be available.

As the appraiser defines the relationship between the features of a property and its price, the *informative table* will be rewrite obtaining a *decision table* in the following way.

Objects (U – SET)	Conditional Attribute (Q SET)		Decisional Attribute (d SET)
Properties	**SQM**	**PARKING**	**PRICE**
1	90	NO(or 0)	A
2	90	YES (or 1)	B
3	90	YES (or 1)	A
4	100	NO(or 0)	B
5	100	NO (or 0)	B
6	90	NO (or 0)	A
7	110	YES (or 1)	D
8	110	NO (or 0)	B
9	110	NO (or 0)	B
10	110	YES (or 1)	D

Table 3- Decision Table

Starting from the data, without any models assumption, it is possible to define the causal relationship between the price and the attribute through if then rules. Using the indiscernible relationship defined in the previous paragraph, several different equivalence *classes* will be built. Inside the set the behavior of the attributes is always the same. The table n.4 indicated below shows several groups dividing the conditional attributes from the decisional attribute (property price)

Conditional Features - Q	**Classes of Equivalence Relationship-Ip**
í SQMý	{1,2,3,6} {4,5} {7,8,9,10}
íPARKINGý	{1,4,5,6,8,9,} {2,3,7,10}
íSQM, PARKINGý	{2,3} {4,5} {1,6}{8,9}{7,10}
Decisional Feature – Property Price - d	**Classes of Equivalence Relationship-Id**
íPRICEý	{1,3,6}{2,4,5,8,9}{7,10}

Table 4- Defining classes of equivalence

Now it will be easier to approximate the set of property prices through the attributes Parking and Square meters. It should be found the rules if.. Then that allows the appraiser to represent the causal relationship between attributes and price through the concept of upper and lower approximations. The rule Q -> d will be defined if

(X, d) Î Q ->d ⇔ x Ç d ¹ 0

Consequently it will be possible to define the following rule:

íSQMý → Price

{1,2,3,6}{1,3,6}

{1,2,3,6}{2,4,5,8,9}

í4,5ý í2,4,5,8,9ý

{7,8,9,10}{7,10}

{7,8,9,10}{2,4,5,8,9}

íPARKINGý → Price

{1,4,5,6,8,9}{1,3,6}

{1,4,5,6,8,9}{2,4,5,8,9}

íSQM PARKINGý → Price

{2,3}{1,3,6}

{2,3}{2,4,5,8,9}

í4,5ýí2,4,5,8,9ý

í1,6ýí1,3,6ý

í8,9ýí2,4,5,8,9ý

í7,10ýí7,10ý

There are several kinds of rule as it has been defined in the previous paragraph. Appraiser is interested in dealing with those rules defined “deterministic” which allow the appraiser to define “strong” relationship between property features and prices. They rely in graphical term inside the lower approximation of the concept. In the previous list they can be pointed out if it is considered that a deterministic relationship Q->d is defined in the following way:

Q => d if and only if I_q Í I_d

This relationship (downward approximation) is verified in the following cases:

Q => d	Deterministic Rules
í SQMý → íPRICEý	**í4,5ý í2,4,5,8,9ý**
íPARKINGý → íPRICEý	**0**
íSQM, PARKINGý → íPRICEý	**í4,5ý í2,4,5,8,9ý**
	í1,6ý í1,3,6ý
	í8,9ý í2,4,5,8,9ý
	í7,10ý í7,10ý

It will be possible to create the following rule based on the observation of the data:

SQM → Price

IF SQM = 100 => PRICE = B

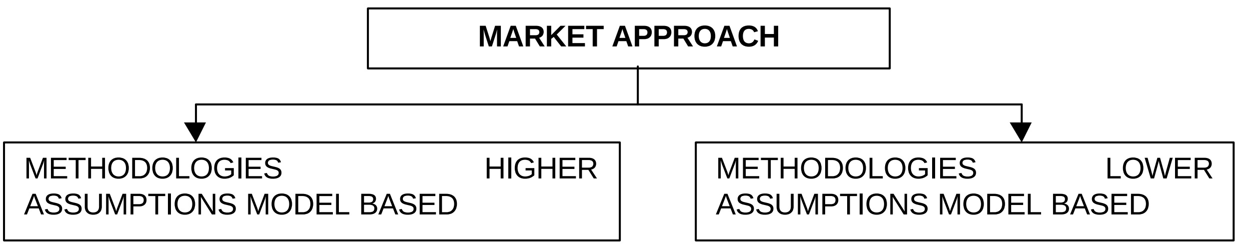
PARKING → Price

NO DETERMINISTIC RULES

íSQM PARKINGý → Price

IF SQM=90 AND NO PARKING => PROPERTY PRICE = A
IF SQM=100 AND NO PARKING => PROPERTY PRICE = B
IF SQM=110 AND NO PARKING => PROPERTY PRICE = B
IF SQM=110 AND YES PARKING => PROPERTY PRICE = D

Establishing causal rules between property prices or rents and its features allow the evaluator to appraise the property similar to the group of property considered in the table on information and decision. This methodology could be defined " rough set approach" (RSA) and could be put inside the market approach. Inside this approach there are procedures as MCA and simple or multiple regression analysis, which are based on a higher number of assumptions. RSA could be considered a new kind of market approach lower assumption based as indicated in the exhibit 1 below:



REGRESSION ANALYSIS
MARKET COMPARISON APPROACH

ROUGH SET APPROACH

Exhibit 1 – Market Approach and Rough Set Theory

In fact, while in the regression analysis a model is predefined, while in the MCA the consumer behavior is supposed to be consistent in order to define an adjusted price, in this method data are read and property features analyzed relying only on the principle of indifference. There is a top down process starting from the full attribute set trying to reduce it in a few deterministic rules. Statistical method tries to introduce new variable in a defined model. While in statistical model there are few variables for many observations requested in Rough set theory it is possible to consider a huge number of features with few data[xix].

The previous example shows as rough set could permit to analyze and define some rules. Without defining a specific model it is possible to define some rules which explain the link between property value and its attributes. This procedure seems to be defined useful in several specific contexts. The former is mass appraisal. In fact if the appraiser can analyze a huge amount of data and features will be able to define rules (deterministic) more and more sophisticated.

Another possible application is to apprise real property. In fact the appraisal value could be reached following the rule developed on a number of real transaction analyzed through this method. Also income-producing properties could be appraised through this method. In fact income is an economic feature of property to be estimated.

Furthermore, defining rule allow the appraiser to foresee the future behavior of property market and to make clear useful relationships between property features and price or rent.

FINAL REMARK AND FUTURE DIRECTIONS OF RESEARCH

At the end of this work is possible to highlight some results. Formerly the contribution of Rough Set appraisal could improve the appraisal process in that market where it is not easy to appraise a property because the data are less than the observations. RSA allows the appraiser to define deterministic rules that link features of a property to its price. Property appraisal is a consequence of these rules. This methodology rely on one only assumption and do not need of a previous analysis on the quality of the data and variables considered. Informatic software makes the application of these methodologies not difficult also to a complex appraisal problems.

In Rough Set appraiser does not need a specific number of transactions. For this reason this method can be used both with few and with a huge number of data. Probably the better application could be mass appraisal because of the necessity of data.

After this contribution could be interesting compare the results coming from regression analysis and RSA in order to define if there are not significant differences. For little amount of data it is possible to compare the results of RSA with Market Comparison

Approach. Another interesting field is forecast. The best attribute to predict the decisional features can be defined, as well as, in this way, in a more complex context, the outliers of a property price and the deterministic rules that link property price to the attributes of a property can be focused.

REFERENCES

---

[i] The work has been developed with a grant of Italian National Council of Research in September 2000 inside the University of Alicante in Spain. I am in debt with Prof. Marco Simonotti and Prof.Halbert C.Smith for their useful suggestions

[ii] See Lin, T.Y. and Cercone N. (1997), *Rough Set and Data Mining*, Dordrecht, Kluwer; Orlowska E. editor, (1997) *Incomplete Information-Rough Set Analysis*, Physica-verlag Heidelberg; Pal. S. and Skowron A. editors (1999), *Rough Fuzzy Hybridization*. Springer – Verlag, Polkowski L. and Tsumoto S. and Lin, T.Y. (2000) editors, *Rough Set Theoryand Applications: New Development*s, Physica Verlag,Heidelberg, to appear, Zong N., Skowron A. and Oshuga S. editors (1999), *New Directions in Rough Sets, Data Mining and Granular Soft Computing* Vol. 1711 of *Lecture Notes in Artificial Intelligence*, Berlin, Springer-Verlag

[iii] See Pawlak Z. (1982), *Rough Sets*, International Journal of Information and Computer Sciences, 11,341-356 and Pawlak Z. (1991), *Rough Sets. Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publisher, Dordecht

[iv] See Matarazzo B. (1997), *L'Approccio dei Rough Set all'Analisi delle Decisioni*, Proceedings XXI Annual Meeting A.M.A.S.E..S., Rome , 10-13 September

[v] Pagliani P. (1997), *Rough Sets Theory and Logic-Algebraic Structures*, In E.Orlowska (Ed.) *Incomplete Information-Rough Set Analysis*, 109-190 Heidelberg: Physica-Verlag. For an analysis between Rough Set Theory and Fuzzy Sets see Dubois D. and Prade H. (1992) *Putting Rough Sets and Fuzzy Sets together*, in Slowinski R. editor (1992) *Intelligent Decision Support: Handbook of Applications and Advances of Rough Set Theory*, Vol.11 of System Theory*, Knowledge Engineering and Problem Solving*, Kluwer Dordrecht.pages 203 - 232

[vi] Note that these approximations sets will be one in the case of ordinary sets

[vii] W.Ziarko (1993), *A Brief Introduction to Rough Sets, The First International Workshopon Rough Sets: States of Art and Perspectives*, University of ReginaSaskatchewan

[viii] See Pawlak, Z. (1991) *Rough Sets. Theoretical Aspects of Reasoning About Data*. Dordrecht: Kluwer Academic Publishers

[ix] In some later versions of this methodology it has been used a dominance relationship that allow to use attributes with preference-ordered scales or criteria. See Greco S., Matarazzo B. and Slowinsky R. (1998) " *A new rough set approach to evaluation of bankruptcy risk*" in Operational Tools in Management of Financial Risk, edited by C. Zopounidis, pp.121-136, Dordrecht, Boston, Kluwer Academic Publishers

[x] W.Ziarko (1993), Ibidem,

[xi] Holte R.C. (1993) *Very Simple Classification Rules Perform Well On Most Commonly Used Datasets,* Machine Learning, 11:63-91

[xii] Duntsch I and Gediga G. (1997) *Statistical Evaluation of Rough Set Dependency Analysis*, International Journal of Human-Computer Studies, 46 589-604

[xiii] For everything see the two works of Rissanen J. (1985), *Minimum Description Length Principle*, in S. Koltzand N.L. Johnson editors Encyclopedia of Statistical Sciences,523-527 N.Y. Wiley; Rissanen J. (1978) *Modeling by the Shortest Data Description*, Automatica, 14, 465-471

[xiv] Browne C. Duntsch I and Gediga G. (1998). *IRIS revisited: A comparison of Discriminant and Enhanced Rough Set Data Analysis* in Slowinski R. editor (1992) *Intelligent Decision Support: handbook of Applications and Advances of Rough Set Theory* Volume 11 System Theory, Knowledge Engineering and Problem Solving, Kluwer Dordrecht pages 345-368

[xv] The transactions have been obtained from RUBINO Immobiliare in Bari. I would like to thank this real estate broker for his cooperation in this work.

[xvi] This is the so called "*nominal scale restriction*"

[xvii] See Bacchus F, Grove A. J. Halpern J.Y., Koller D. (1994) *From Statistical Knowledge Bases to Degrees of Belief.* Technical Report 9855, IBM

[xviii] See for example ROSETTA software inside the following website : **www.idt.unit.no/~aleks/rosetta/rosetta.html**

[xix] For an analysis of the differences between statistical model and Rough Set see Duntsch I. & Gediga G. (1997), *Roughian – Rough Set Information Analysis*, In Sydow A. Editor (1997), Proc.15th IMACS World Congress Vol.4, Berlin, Wissenschaft und Technik Verlag