

ROUGH SETS ANALYSIS WITH ANTISYMMETRIC AND INTRANSITIVE ATTRIBUTES: CLASSIFICATION OF BRAZILIAN SOCCER CLUBS

Annibal Parracho Sant'Anna

Universidade Federal Fluminense (UFF)

Niterói – RJ

tppaps@vm.uff.br

Recebido em 09/2006; aceito em 04/2008 após 1 revisão

Received September 2006; accepted April 2008 after one revision

Abstract

This work aims to develop alternative classifications for teams in a Championship. Data from the 2005 Brazilian National Soccer Championship are analyzed. Rough Sets Theory (RST) is employed in this analysis. By evaluating the quality of the approximation in terms of probabilities of concordance and discordance between the classification by the set of decision attributes and by the set of condition attributes of a randomly chosen pair of objects as discernible or indiscernible, the modification of RST employed allows to consider antisymmetric and intransitive relations. The balance between the numbers of goals scored by the pairs of clubs in direct confrontations is one such relation.

Keywords: rough sets; quality of approximation; soccer.

Resumo

Este trabalho visa a desenvolver classificações alternativas para times de futebol em um campeonato. Dados do Campeonato Brasileiro de Futebol de 2005 são analisados. A Teoria dos Conjuntos Aproximativos (TCA) é empregada nesta análise. Por avaliar a qualidade da aproximação em termos de probabilidades de concordância e de discordância entre a classificação pelo conjunto de atributos de decisão e pelo conjunto de atributos de condição de um par de objetos escolhidos aleatoriamente como discerníveis ou indiscerníveis, a modificação da TCA empregada permite considerar relações intransitivas e anti-simétricas. A diferença entre o número de gols marcados pelos pares de clubes em confrontos diretos é uma relação desse tipo.

Palavras-chave: conjuntos aproximativos; qualidade da aproximação; futebol.

1. Introduction

This work aims to develop a framework to generate alternative classifications for teams in a championship. A motivation for that is provided by the need to separate classes of clubs to more important tournaments or to relegation to lower leagues. A more reliable classification for that end is derived from combining the rank on points officially won in the pairwise confrontations throughout the championship with alternative ranks generated by taking into account other criteria of performance evaluation.

The alternative classification framework is developed in such a way as to allow for different attributes becoming more important each year, so that the selection procedure, though precisely defined, does not allow to anticipate which rules will be finally applied to select the clubs for promotion or relegation. This aspect is important to avoid that clubs early doomed to relegation become less motivated in their last matches. The declining motivation of some clubs may rouse an undesirable effect of the order of the matches on the final championship result.

Rough Sets Theory (RST) constitutes a theoretical basis suitable for the development of such a framework. As developed by Pawlak (1982), RST provides the elements to uncover classification rules hidden in data sets containing information about distinct attributes of the objects being classified. These objects may be thought as options to choose among and their attributes may be preference evaluations according to different criteria.

The main instrument of rough sets analysis is the measurement of the similarity between the classification of the objects according to different sets of attributes, by one side the decision attributes and by another side the condition attributes. The decision attributes determine a partition to be approximated and the condition attributes determine approximations to it. Measuring the quality of approximation of the decision partition by different subsets of condition attributes one may identify which condition attributes are relevant to that decision classification, in such a way as to rank the condition attributes and eventually choose among them a small number to keep and discard the information about the others. After that, decision rules based on the relevant condition attributes may be extracted from the data set.

The classical measure of quality of approximation is the index γ of Pawlak, which estimates the probability of a randomly chosen object being unambiguously classified. In the computation of the index γ , an object is considered ambiguously classified if it is indiscernible according to the set of condition attributes from another object that the decision attributes place in a different class.

RST may be easily extended to deal with attributes that naturally determine order instead of equivalence relations. Such an extension has been made by Sant'Anna & Sant'Anna (2006) and, within the classical framework, by Greco *et alii* (1999). In fact, complete order relations induce a partition in the universe of objects as well as equivalence relations.

Discernibility by a set of attributes tends to increase with the number of attributes in the set. Usually, there is only one decision attribute or a small number of decision attributes and there is a large number of condition attributes. In such context the applicability of RST is impaired by the fact that, if the measurement is accurate enough, boundaries tend to vanish and the index γ tends to 1.

Another aspect of the classical index of approximation is its asymmetry, favoring discernibility in the decision partition and indiscernibility by the condition attributes. Such

asymmetry penalizes the rougher condition attributes and inhibits the reduction of the number of attributes. This may be overcome by disregarding the differences below some threshold, as in the variable-precision rough sets (Ziarkos, 1993). The hybridization with fuzzy sets in the fuzzy-rough set models as, for instance, in Dubois (1990) or Greco *et alii* (2000) is another development to deal with this difficulty. Sant'Anna (2007) takes a different approach, measuring the quality of the approximation by the probability of a randomly chosen pair of objects being concordantly classified as discernible or indiscernible by the set of decision attributes and the set of condition attributes.

For the case of order relations this approach is here complemented with a new rule to combine the evaluations according to the different attributes. In the partitions determined by order relations, the problem of vanishing boundaries occurs again. Now, instead of indiscernibility it is dominance that becomes rare if the same ordering according to all the condition attributes is required. The new way to deal with this difficulty here proposed explores the possibility of ranking the condition attributes. The values of the indices of quality of approximation of each isolate attribute automatically provide one such ranking.

Dealing only with the relations between the objects determined by the attributes and not with the partitions determined in the universe of such objects, the approach developed in Sant'Anna (2007) allows for extending RST not only to asymmetric but also to intransitive relations. In the present application, two condition attributes evaluate offensive and defensive power of the teams along the championship, respectively, by the number of goals scored and the number of games in which the club was defeated. Scoring goals and not being defeated are the main objectives that a club tries to achieve in any match and conquer the championship should result from the efficiency on accomplishing that. A third condition, intransitive, is added to these two, taking as an attribute of each pair of clubs the result of their direct confrontation. By the presence of this attribute in the rules, every match may become important and the motivation of all clubs is kept high throughout the whole championship.

The next section contains a brief presentation of the basic concepts of rough sets, a description of the new approximation approach and a forward procedure to rank the attributes. Section 3 presents the championship data and the results of the rough set analysis, showing the effect of the level of the granulation applied to the decision classification on the indices of quality of approximation. Final comments are presented in Section 4.

2. Rough Sets Approximations

Formally, any classification is characterized by a quadruple (U, Q, V, f) , where U is a non-empty set, the universe of the objects evaluated, Q is a set of attributes of the objects in U , each attribute $q \in Q$ being able to receive evaluations in a set V_q so that V is the union of the V_q and, finally, f is an application of the Cartesian product of U and Q in V such that, for each $o \in U$ and $q \in Q$, $f(o, q) \in V_q$.

To each $P \subseteq Q$, non-empty, is associated a relation of indiscernibility in U , denoted I_P , and a partition O_P , formed by the I_P equivalence classes, the sets of objects identically evaluated by all the attributes in P . If $(o_1, o_2) \in I_P$, we say that o_1 and o_2 are P -indiscernible. The equivalence class in O_P containing object o_1 is denoted $I_P(o_1)$ and is called a P -elementary set.

The set Q of attributes is divided into two parts, the set D , of decision attributes, responsible for the classification that must be explained using the attributes of the complementary set

$C = Q-D$, called condition attributes. The measurement of the quality of the approximation of the partition O_D by the partition O_C is a key tool of RST. This concept is originally measured by the index of quality of approximation γ of Pawlak (1982).

The index γ is based on evaluating the approximation to each decision class. It counts the proportion of objects indiscernible according to C that belong to distinct classes according to D . It does not take into account the probability of objects discernible according to C being indiscernible according to D . Variants have been proposed for this index, always preserving this point of view (Gediga & Düntsch, 2003).

The asymmetry in the definition of γ , makes this index assume its maximum anytime all options are discernible by the set of the condition attributes, independent of which is the partition determined by the decision attributes. Sant'Anna (2007) proposes an index of the quality of approximation based on determining the proportion of cases of simultaneous discernibility or indiscernibility according to D and C . More precisely, this index gauges the quality of approximation dividing by the total number of pairs of objects in the universe U the sum of the number of pairs of indiscernible objects according to C that belong to the same class according to D and the number of pairs discernible according to C that belong to distinct classes according to D . Depending only on the knowledge of the results of pairwise comparisons, this computation may be done for intransitive relations, which occur in many real-world instances. Other advantages of depending only on pairwise comparison are found, in diverse contexts, by Saaty (1980) and Lootsma (1999), for instance.

The value of this index may be reduced by the possible presence of pairs of objects incomparable to each other, which are counted in the denominator. A variant whose symmetry is not affected by incomparability passing unnoticed is proposed here. This new index is given by the difference between the proportions of concordant and discordant pairs. More precisely, it is determined dividing by the total number of pairs of objects in the universe U the difference between the sum of the number of pairs of indiscernible objects according to C that belong to the same class according to D and the number of pairs discernible according to C that belong to distinct classes according to D and the sum of the number of pairs of indiscernible objects according to C that belong to distinct classes according to D and the number of pairs discernible according to C that belong to the same class according to D .

This index may be thought in statistical terms as an estimator of the correlation coefficient between random variables summarizing the discernibility by decision and condition attributes. For each set $P \subseteq Q$, discernibility may be represented by a function G_P defined in the Cartesian product $U \times U$ by $G_P(o_1, o_2) = 1$ if o_1 and o_2 are indiscernible according to P , $G_P(o_1, o_2) = -1$ if o_1 and o_2 are discernible according to P and $G_P(o_1, o_2) = 0$ if o_1 and o_2 are incomparable according to P . The new index of quality of approximation of D by C is then defined by the expected value of the product of G_D and G_C , assuming a uniform distribution in $U \times U$.

This index will be denoted by ρ . It is an estimator of the correlation coefficient and of the covariance between G_D and G_C if these random variables are assumed to have a null expected value and maximal variance. These last assumptions, if not satisfied, will reduce the value of the index for condition attributes with high rates of discernibility or indiscernibility that would artificially increase the correlation coefficient.

2.1 The case of dominance

Greco *et alii* (1999) adapted Rough Sets Theory to take into account partitions determined by order relations. The attributes are then preference criteria, in such a way that the set of decision attributes and the set of condition attributes determine each an antisymmetric relation, i. e., a dominance relation such that, given two objects o_1 and o_2 , if o_1 and o_2 are discernible, then ' o_1 dominates o_2 ' implies ' o_2 does not dominate o_1 ' and ' o_2 dominates o_1 ' implies ' o_1 does not dominate o_2 '. For such a relation, we say that o_1 strictly dominates o_2 if o_1 dominates o_2 and o_1 and o_2 are discernible.

In the same way, the index ρ above defined may be extended to order relations. The quality of approximation of the set of decision criteria D by the set of condition criteria C will then be evaluated dividing, by the total number of pairs of different objects of the universe of objects U , the difference between the number of pairs with strict dominance in the same direction according to C and D and the number of pairs with strict dominance in opposite directions according to C and D .

This definition may still be formulated in terms of graph theory. The sets of criteria D and C determine two oriented graphs, with directed edges if there is strict dominance and simple edges if there is indiscernibility. Depending on the definition of dominance, there may be the case of objects not comparable, represented by unconnected nodes. The quality of approximation index is determined dividing by the total number of pairs the difference between the number of pairs of different objects with links directed in the same direction according to both sets of criteria and with links directed in opposite directions.

2.2 Joint Dominance

In classical RST, for a pair of objects (o_1, o_2) be discernible according to a set of attributes P , it is enough that, for some $p \in P$, $f(o_1, p) \neq f(o_2, p)$. As the number of attributes grows, it becomes difficult to find indiscernible pairs. This suggests replacing the requirement of unanimity by majority. Extending the RST approach to the case of order relations, Greco *et alii* (1999) require, for an object o_1 strictly dominate another object o_2 according to a set of attributes P , the existence of at least one attribute $p \in P$ for which o_1 strictly dominates o_2 according to p and the absence of attributes in P for which o_2 strictly dominates o_1 . With this definition, it is strict dominance that becomes difficult to establish as the number of attributes in P grows. Analogously to the discernibility case, a majority relaxation would require that, for the strict dominance of o_2 by o_1 , the number of attributes in P for which o_2 dominates o_1 be strictly smaller than the number of attributes for which o_1 dominates o_2 . If these numbers are equal than the objects will be indiscernible.

This advantage of the majority rule over the unanimity rule may not be enough because both these definitions raise the possibility of intransitivity cases, with o_1 and o_2 indiscernible, o_2 and o_3 indiscernible and o_1 strictly dominating o_3 . If one wishes to accept ranking the condition attributes by means of a complete order relation, then another definition of dominance that is free of intransitivity may be employed. Such ranking of attributes may always be derived from their isolated indices of quality of approximation.

According to this new definition, there is strict dominance of object o_1 over object o_2 if o_1 is preferable to o_2 according to some criterion and there is no criterion ranked best than that according to which o_2 is preferable to o_1 ; again, if there is neither strict dominance of o_1 over

o_2 nor of o_2 over o_1 , then the two objects are considered indiscernible, so that every pair of different objects can be compared according to this definition too. This approach goes in the same direction of the priority heuristics, shown by Brandstater *et alii* (2006) to be a reasonable prescriptive approach.

2.3 A Forward Reduction Procedure

A procedure to compute the index of quality of approximation for all subsets of condition attributes would have an exponential complexity. To avoid that, a fast forward procedure to select a minimal set of condition attributes approaching a given set of decision attributes D and, simultaneously, rank their attributes was developed in Sant'Anna (2007). It has the following steps.

1. Start ranking the unitary sets of condition attributes according to its individual quality of approximation. Select first and make definitively enter the set of chosen condition attributes that condition attribute A_1 with the highest value of the index of quality of approximation.
2. Rank the pairs of attributes (A_1, A_j) with $j \neq 1$ according to the value of that index. Exclude from the set of attributes those A_j whose addition to A_1 does not increase the value of the index.
3. Among the attributes not yet excluded, select to enter the set of chosen condition attributes and rank as second, that A_j with the pair (A_1, A_j) presenting the highest value of the index in the ranking of the previous step.
4. Rank the triples formed by adding to that pair (A_1, A_j) one attribute still not excluded. Exclude from the set of attributes any attribute forming a triple with a value of the index smaller or equal to that of the starting pair.
5. Proceed to the quadruples that include the triple with the highest values of the index among those ranked in the previous step.
6. Proceed increasing the size of the set in the same way until there are no more attributes to be added.

To speed up convergence, the exclusion steps may withdraw more attributes from the set of condition attributes. For instance, may be excluded all those attributes that do not increase the value of the index more than a given δ larger than zero, instead of only those that do not raise it at all.

These rules suppose absence of ties in the values of the index of quality of approximation. If ties are observed, then all untying alternatives should be compared. It may be useful to explore every set of attributes thus generated and to investigate the differences eventually appearing in the final classifications.

2.4 Inner and Outer Classes

These new definitions do not depend on the concepts of lower and upper approximations and of boundaries of the rough sets. But they can be used to determine such approximations. The lower approximation of a decision class is formed by those objects in it that are indiscernible also according to the condition attributes of those not in it. The upper approximation is formed by those objects not indiscernible according to the condition attributes of some object in the decision class.

Inside the lower approximations we may still identify smaller classes, that will be here called inner classes, maximal sets of objects indiscernible from any object within the set according to both condition and decision attributes. If we deal with antisymmetric relations we may have objects that strictly dominate another according to the condition attributes and are indiscernible according to the decision attributes or vice-versa. Thus, an inner class within a decision class may strictly dominate another according to the condition attributes.

By the other side, in the case of intransitive relations we may have two levels of outer classes, the first determined by considering in the boundary only those objects that are indiscernible of the objects inside the decision class. These first classes may be involved by classes of another kind, including those objects not necessarily indiscernible themselves of any object inside the decision class, but indiscernible of objects of the first involving class, i. e., of objects that are indiscernible of objects inside the decision class.

3. Analysis of the Classification in the Soccer Championship

In this section, the quality of approximation evaluation and attributes ranking procedure above described are applied to analyze a case of dominance relations involving an intransitive attribute. The objective of the analysis, is employing attributes that determine the performance of the teams in terms of general attack and defense performance and of particular behavior in direct confrontations, explain the final ranking of the clubs that entered a soccer championship, in such a way as to provide alternative final classifications.

3.1 The Data Set

Table 1 presents final number of points scored, games won, tied and lost, goals scored and goals taken by the 22 clubs in the 2005 Brazilian Soccer Championship.

The championship rules assign to the club three points by game won, one point for a tie and zero points in case of loss. Besides, the number of wins and the balance of goals are employed as untying criteria. Looking for factors that affect the classification in the championship, we follow the point of view taken, for instance, by Souza Jr & Gamerman (2004) and Carmichael & Thomas (2006), trying to find such factors in separate offensive and defensive skills.

The offensive power is naturally measured by the number of goals scored. The defensive skill is more difficult to access. To take a symmetric point of view, we might consider with a negative sign the number of goals suffered, but a favorable balance of offensive and defensive power may lead to large numbers of goals pro and against the team. The strength of the defense is forged to assure that the team will not be beaten when the offensive power is not enough. Thus the ability of avoiding defeats is a more natural way to access it than the number of goals suffered. For this reason, maybe, the typical club fan leaves the stadium proud of goals scored and not caring about goals taken if the final result is not a defeat.

If there were proportionality between the vectors of wins and ties or if one of these vectors were constant, there would be precise correlation between the number of games lost and the number of points. However this is not the case. It will be seen in the following that to explain the final ranking the number of games lost leaves considerable space to be filled by the other variables in the model proposed.

Table 1 – Final Standings.

Clubs	Points	Wins	Ties	Losses	Goals scored	Goals taken
Corinthians	81	24	9	9	87	59
Internacional	78	23	9	10	72	49
Goiás	74	22	8	12	68	51
Palmeiras	70	20	10	12	81	65
Fluminense	68	19	11	12	79	70
Atlético-PR	61	18	7	17	76	67
Paraná	61	17	10	15	59	51
Cruzeiro	60	17	9	16	73	72
Botafogo	59	17	8	17	57	56
Santos	59	16	11	15	68	71
S. Paulo	58	16	10	16	77	67
Vasco da Gama	56	15	11	16	74	84
Fortaleza	55	16	7	19	58	64
Juventude	55	15	10	17	66	72
Flamengo	55	14	13	15	56	60
Figueirense	53	14	11	17	65	72
S. Caetano	52	14	10	18	54	60
Ponte Preta	51	15	6	21	63	80
Coritiba	49	13	10	19	51	60
Atlético-MG	47	13	8	21	54	59
Paysandu	41	12	5	25	63	92
Brasiliense	41	10	11	21	47	67

Direct confrontation appears as a complement for these two factors by bringing into consideration the aspects of the two teams structure that affect each game particularly. Table 2 presents the results of direct confrontation. The value 1 means that the team with the acronym heading the column has scored more goals than that with the acronym in the beginning of the row in the two games between these clubs. Zero means same number of goals scored. And -1 means that the team in the row has scored more than the opponent in the two games of their direct confrontation.

Table 2 – Direct Confrontation.

Club	AP	BG	BS	CN	CT	CZ	FI	FM	FS	FT	GO	IT	JV	PL	PR	PS	PT	SA	SC	SP	VG
AM	1	1	0	1	0	1	-1	-1	-1	-1	1	0	-1	1	1	-1	-1	0	1	1	1
AP		-1	-1	1	-1	-1	1	-1	0	-1	1	1	1	-1	1	0	-1	1	1	-1	-1
BG			0	-1	-1	-1	1	1	1	0	1	-1	-1	1	0	0	-1	1	-1	1	-1
BS				1	0	1	1	0	1	1	-1	1	0	0	1	1	-1	-1	1	-1	1
CN					-1	0	-1	-1	-1	-1	1	0	1	-1	-1	-1	-1	-1	1	1	-1
CT						1	-1	0	-1	-1	1	0	1	-1	0	-1	-1	1	0	1	-1
CZ							-1	0	1	-1	-1	1	-1	-1	1	1	0	-1	1	1	-1
FI								1	1	1	-1	1	-1	1	1	1	-1	-1	-1	0	-1
FM									0	-1	1	1	1	-1	-1	-1	-1	-1	-1	1	0
FS										1	-1	-1	1	1	1	-1	0	-1	1	-1	1
FT											1	1	0	-1	1	0	0	0	-1	0	-1
GO												0	0	1	-1	-1	-1	1	-1	1	-1
IT													1	0	1	-1	-1	-1	-1	-1	-1
JV														1	-1	1	1	-1	-1	1	0
PL															-1	-1	-1	0	-1	-1	-1
PR																-1	1	0	1	1	1
OS																	1	1	-1	1	1
PT																		-1	1	-1	-1
SA																			-1	-1	-1
SC																				0	0
SP																					0

3.2 Application of the Selection Procedure

Granulation is a key concept in rough sets theory. The accuracy in the measurements and the consequent level of distinction between the objects determine the roughness in the sets and the quality of the approximation. Three different granulations are explored for the clubs classification. Initially, the decision variable is the exact number of points of the club in the championship final standings. This results in a thin granulation that puts more than a half of the clubs in unitary sets, resulting in a total of 17 classes.

More complex situations occur when, requiring an empty space of one point to separate decision classes, the six clubs with 58 to 61 points are considered tied in one decision class, the four clubs with 55 and 56 points in another decision and the three clubs with 51 to 53 points in a third. The final result of this granulation is a set of 11 decision classes.

An even thicker granulation is considered, derived from the requirement of an empty space of two points to separate decision classes. With this requirement, all clubs with 49 to 61 points are in the same class. And, above these, Palmeiras and Fluminense, the fourth and fifth classified, are also put together. The final result of this granulation is a set of only 7 decision classes.

According to number of games lost and number of goals scored, two granulations are employed. First the exact observed values are employed. A second granulation is employed with the clubs divided into six thicker classes according to each attribute. According to goals scored, these classes are separated by spaces of at least two goals. The six classes generated are of 87 goals, 72 to 81 goals, 63 to 68 goals, 54 to 59 goals, 51 goals and 47 goals. According to games lost the classes are of 9, 12, 15, 18, 21 or 24 losses, plus or minus 1.

Table 3 presents the values of the indices γ and ρ for the decision classification of the three condition attributes isolated, each line corresponding to a different combination of granulations for the two kinds of attributes. In the successive pairs of rows corresponding to the same number of classes in the decision partition, the different granulations of the condition attributes are determined, above, by their exact observed values and, below, by their representations in six classes. For the last attribute, only the observed values were considered. For this attribute, the index γ cannot be determined. The computation of the index γ employed 4emka (2000).

Table 3 – Quality of Approximation of Isolated Attributes.

Granulation		games lost		goals scored		direct confrontation	
decision	condition	γ	ρ	γ	ρ		ρ
17	thin	0.14	0.76	0.09	0.59		0.20
17	thick	0.04	0.71	0.09	0.47		0.20
11	thin	0.14	0.74	0.09	0.57		0.21
11	thick	0.04	0.72	0.09	0.46		0.21
7	thin	0.73	0.57	0.09	0.41		0.19
7	thick	0.64	0.57	0.09	0.35		0.19

The values of the index ρ in Table 3, being all positive, show that the three condition attributes are positively correlated to the decision attribute. They also confirm that the number of games lost is the condition criterion to be ranked first and to start with in the forward selection procedure. The second largest quality of approximation according to this index is the number of goals scored. Direct confrontation presents the lowest value of ρ .

The value of the index ρ is slightly reduced, in the case of goals scored, by the reduction in the accuracy of measurement of the condition attributes to six classes. For the other two condition attributes it presents very close values in the two cases. It is almost the same for every condition criterion as we pass from 17 to 11 decision classes and only falls a little more sharply as the indiscernibility in the decision classification increases with the reduction of the number of decision classes from 11 to 7. This suggests that this last aggregation to reduce the number of decision classes may be excessive.

The index γ is not affected by the granulation, except for the number of games lost, when the number of classes in the decision partition is reduced to 7. In this case, it jumps from 0.04 or 0.14 to 0.64 and 0.73, respectively. For the other decision classes, its values are always close to its minimum possible value of zero, with the evaluation of the number of games lost even lower than that of the number of goals scored in some cases. The quality of approximation of games lost and goals scored taken together, as measured by the index γ also presents large

variation, with values of 0.18 and 0.32 when the exact numbers of points in the official classification are confronted, respectively, to exact and granulated values of the condition attributes and of 0.86 when the decision partition in 7 classes is confronted to the condition attributes in the exact as well as in the granulated form. For the case of 11 decision classes, the values of γ are 0.59 and 0.32, respectively, for exact and granulated values of the condition attributes.

Table 4 presents the values of ρ for the composed sets of condition attributes. Three different definitions of joint dominance were applied: unanimity, majority and complementarity. In this last case, only the order determined by the quality of approximation of the condition attribute isolated was considered: priority for number of games lost, followed by the number of goals scored and, in the last priority, the result of direct confrontation. Different columns for unanimity and majority are presented only for the case of all conditions attributes since these different definitions lead to the same result in the case of two attributes.

Table 4 – Joint Quality of Approximation.

Granulation		goals+confront	losses+confront	losses+goals		All		
decision	condition			maj./unan.	priority	unan.	maj.	priority
17	thin	0.45	0.53	0.69	0.78	0.48	0.71	0.78
17	thick	0.40	0.51	0.70	0.78	0.50	0.63	0.78
11	thin	0.45	0.53	0.70	0.78	0.49	0.7	0.78
11	thick	0.40	0.51	0.72	0.79	0.52	0.64	0.79
7	thin	0.40	0.47	0.60	0.68	0.43	0.62	0.68
7	thick	0.37	0.47	0.57	0.69	0.46	0.57	0.69

Table 4 confirms the information in Table 3 about the limited influence of the granulation on the index ρ and about the order on the quality of approximation, higher for games lost and lower for the result of direct confrontations. Besides, it shows that the majority rule for joint dominance results in better approximations than the unanimity rule and that the best results are obtained as secondary attributes enter only to complement the priority attributes in the case of a tie.

3.3 Classification and Decision Rules

Finally, classification rules are derived. In this section are presented, as an example, the classifications for the case of the decision classes directly determined by the exact number of points attained, i. e., the case of 17 classes. The other cases present similar results.

If the attributes are not ranked, the lower classes coincide with the unitary classes in the case of the clubs in the two first positions. All other inner approximations result in empty classes. Analogously, two upper classes coincide with the unitary classes in these same cases and, for the other decision classes, there is only one upper class, formed by all the remaining clubs.

Under the hypothesis of identical strengths, the distribution of the sum of points would be approximately normal, in such a way that the rarefaction of the tails of the distribution would allow for the separation of small classes at both ends, not only among the best classified. The

asymmetry detected may be revealing an aspect of the sports culture. The clubs aim the title of champion in such a way that distinctions between clubs better qualified to reach that title may be observed and the other clubs are indiscernible.

If the attributes are ranked, then other lower classes emerge, in the opposite end of the list. The new lower classes are the unitary classes of S. Caetano and Atlético-MG and the last decision class. The two objects of this last class may still be separated into two elementary inner classes, since Brasiense has fewer games lost than Paysandu. On the other side, the first unitary classes according to the decision attribute are again upper classes. But now the same happens to the binary class formed by the two last clubs and, besides those, three new upper classes appear. The first involves the three clubs in the third, fourth and fifth positions in the final standings. The second is a big class involving the clubs classified above S. Caetano according to the decision variable except those in the three first classes at the upper end. The other is formed by the five clubs in the lower end of the decision partition.

It is interesting to notice that the need to keep together the objects of the decision classes in the upper classes generate some large classes, but no division of upper classes due to the presence of an intransitive relation among the condition attributes is observed.

In terms of decision rules based only on the condition attributes, the classes above referred may be determined as:

1. Corinthians, the champion: 10 or less games lost, 87 goals scored;
2. Internacional, the champion runner up: 10 or less games lost, more than 70 and less than 87 goals scored;
3. Third class: 12 games lost;
4. Median class: around 15 games lost or from 60 to 70 goals scored or, yet, around 18 games lost, from 54 to 60 goals scored and win over S. Caetano in direct confrontation;
5. S. Caetano: around 18 losses, from 54 to 60 goals scored and defeat in direct confrontation to the other clubs in the same range of games lost and goals scored;
6. Coritiba and Ponte Preta: one characterized by the position around 18 games lost and less than 54 goals scored and the other around 21 games lost and with more than 60 goals scored;
7. Atlético-MG: around 21 games lost and 54 to 60 goals scored;
8. Brasiense: around 21 games lost and less than 54 goals scored;
9. Paysandu: more than 22 games lost.

4. Conclusion

A variant of rough sets analysis based on comparing the links produced by decision and condition attributes between pairs of objects was here employed. It has been used to evaluate forms to reduce the set of condition attributes and to rank the attributes.

This ranking ability was explored in a strategy to combine preference criteria into global dominance relations that avoids intransitivity cycles. This approach was successfully applied to create alternative classifications for clubs in a championship.

The evaluation of the approximation in terms of relations between pairs of objects encompasses the cases where instead of equivalence or order relations determining partitions in the universe of objects we deal with intransitive relations. This occurs in the championship

situation studied. One attribute found significant was the result of the direct confrontation between the clubs.

It was finally shown how a description of the rough decision classes in terms of decision rules based on the condition attributes can be derived from this new kind of approximation evaluation. A feature of this approach is the ability to identify inner classes inside the decision classes. The distinct inner classes that it allows to determine in the clubs classification were clearly described in terms of only the values of the condition attributes. Similar results were obtained for involving classes.

This approach may be extended to more general classifications. For instance, racing cars championships or the ATP ranking of players participating in different tournaments. By treating missing matches as incomparable pairs, the quality of approximation evaluation procedures here developed may be adapted to such cases.

A more general extension to be considered is that of the index of quality of approximation here proposed. Without weights to combine the variables and employing only the sign of the differences, it will be more robust than canonical correlation coefficients in applications involving observations subject to mixture of heavy tailed disturbances.

Acknowledgements

This work was partially supported by a CNPq grant. I am grateful to Pesquisa Operacional anonymous referees for important improvement recommendations.

References

- (1) 4eMka System. (2000). *A rule system for multicriteria decision support integrating dominance relation with rough approximation*. Laboratory of Intelligent Decision Support Systems, Institute of Computing Science, Poznan University of Technology, Poznan.
- (2) Brandstater, E.; Gigerenzer, G. & Hertwig, R. (2006). The Priority Heuristic: Making Choices without Trade-Offs. *Psychological Review*, **113**, 409-432.
- (3) Carmichael, F. & Thomas, D. (2006). What Makes for Winning Team Performances? Evidence from EURO 2004. **In: An Amalgam of Sports & Exercise Research** [edited by G.T. Papanikos], ATINER, Atenas, 5-35.
- (4) Dubois, D. (1990). Rough fuzzy sets and fuzzy rough sets. *International Journal of General Systems*, **17**, 191-209.
- (5) Gediga, G. & Duntsch, I. (2003). On model Evaluation, Indices of Importance and Interaction Values in Rough Set Analysis. **In: Rough-Neuro Computing: A way for Computing with Words** [edited by S.K. Pal, L. Polkowski and A. Skowron], Physica-Verlag, Heidelberg, 251-276.
- (6) Greco, S.; Matarazzo, B. & Slowinski, R. (1999). Rough Approximation of a Preference Relation by Dominance Relations. *European Journal of Operational Research*, **117**, 63-83.

- (7) Greco, S.; Matarazzo, B. & Slowinski, R. (2000). Fuzzy extension of the rough set approach to multicriteria and multiattribute sorting. **In:** *Preferences and Decisions under Incomplete Knowledge* [edited by J. Fodor, B. De Baets and P. Perny], Physica-Verlag, Heidelberg, 131-151.
- (8) Lootsma, F.A. (1999). *Multicriteria Decision Analysis via Ratio and Difference Judgement*. Kluwer, Dordrecht.
- (9) Pawlak, Z. (1982). Rough Sets. *International Journal of Computer and Information Sciences*, **11**, 341-356.
- (10) Saaty, T.L. (1980). *The Analytic Hierarchy Process*. McGraw-Hill, New York, NY.
- (11) Sant'Anna, A.P. (2007). Probabilistic Indices of Quality of Approximation. **In:** *Rough Computing: Theories, Technologies and Applications* [edited by A.E. Hassanien, Z. Suraj, D. Slezak & P. Lingras], IGI, New York, 162-174.
- (12) Sant'Anna, A.P. & Sant'Anna, L.A.F.P. (2006). A Probabilistic Approach to Evaluate the Exploitation of the Geographic Situation of Hydroelectric Plants. *Proceedings of ORMMES 2006*, 1-17.
- (13) Souza Jr, O.G. & Gamerman, D. (2004). Previsão de Partidas de Futebol usando Modelos Dinâmicos. *Anais do XXXVI SBPO*, S. João Del Rey, 649-659.
- (14) Ziarko, W. (1993). Variable precision rough set model. *Journal of Computer and System Sciences*, **46**, 39-59.