A Generalized Fuzzy-Rough Set Application for Forest Fire Risk Estimation Feature Reduction

T. Tsataltzinos, L. Iliadis, S. Spartalis

HAL Id: hal-01571458
https://hal.inria.fr/hal-01571458

Submitted on 2 Aug 2017

A generalized fuzzy-rough set application for forest fire risk estimation feature reduction.

Tsataltzinos T[1], Iliadis L[2], Spartalis S[3]

[1]PhD candidate Democritus University of Thrace, Greece, tsataltzinos@yahoo.gr

[2]Associate Professor Democritus University of Thrace, Greece, liliadis@fmenr.duth.gr

[3]Professor Democritus University of Thrace, Greece, sspart@pme.duth.gr

Abstract. This paper aims in the reduction of data attributes of a fuzzy-set based system for the estimation of forest fire risk in Greece, with the use of rough-set theory. The aim is to get as good results as possible with the use of the minimum amount of data attributes possible. Data manipulation for this project is done in MS-Access. The resulting data table is inserted into Matlab in order to be fuzzified. The final result of this clustering is inserted into Rossetta, which is a Rough set exploration software, in order to estimate the reducts. The risk estimation is recalculated with the use of the reduct set in order to measure the accuracy of the final minimum attribute set. Nine forest fire risk factors were taken into consideration for the purpose of this paper and the Greek terrain was separated into smaller areas, each concerning a different Greek forest department.

Keywords: fuzzy sets, rough sets, forest, fire, risk, attribute reduction

1   Introduction

Forest fire risk estimation is one of the major assessments of present and future forestry. The number of fire instances is growing annually and one of the most basic concerns of every modern country is to find a way to estimate where and when the next forest fire incident will occur. Fire risk evaluation is an important part of fire prevention, since the planning that needs to be done before a fire instance occurs, requires tools to be able to monitor where and when a fire is more prone to ignite, or when it will cause the most negative effects. The most commonly used definition of Fire risk (danger) is the one that is expressed as an index of both constant and variable risk factors which affect the ignition, spread and difficulty of control of fires and the damage they cause [15].

A lot of theories have been applied but all of them luck in precision. Traditional fire danger systems rely on meteorological indices, based on atmospheric conditions which are not the only factors that compose the overall risk of fire. Human caused factors, fuel loads, and any other non atmospherical factor needs to be taken under consideration. Modeling fuel trends have been proposed to analyze spatial and temporal changes in fire risk conditions [14]. Many methods from various countries such as USA, Canada, Australia and France take into consideration weather factors, soil water and fuel loads. Most of those methods use statistical analysis of past and present data in order to evaluate a short term fire risk.

Many other methods were used to evaluate forest fire risk. Each one of these methods is adapted to the data available, or even needs specific data in order to produce results. The kind of data indices used in every one of these occasions is different, according to the objectives of the forest management planning. Many countries have greatly invested in fire risk evaluation and data collection because risk estimation produces better results, when a great amount of data is available. On the other hand, when data entries are missing, imprecise, cannot be measured or when specific data is not available at all, risk estimation becomes a very difficult task.

One of the most interesting methods that helps in overcoming the problem of luck of accurate data, concerns the use of fuzzy sets theory, introduced by Lotfi A. Zadeh in 1965, in order to present vagueness in everyday life as a mathematical way. This theory is widely used for everyday appliances and helps into depicting real life situations and problems, such as uncertainty and luck of precise information into computer data, available for use in order to estimate the risk. Despite its positive points, fuzzy set theory has a basic backdrop. In order to increase the accuracy of each fuzzy set system, the user has to increase the available data, which results in problems like difficulty in obtaining this amount of data and the increase in rules that need to be inserted to the system. With the use of bigger data sets the total number of rules becomes the basic problem of the user.

This paper concentrates in the application of a rough-set based method in order to reduce the total number of data entries needed. Rough set theory was developed by Zdzislaw Pawlak in the early 80's, it uses data that can be acquired from measurements or human experts and it deals with the classification analysis of data tables [16]. The concepts in rough set theory can be used to define the necessity of features, which in this case are the factors that influence the forest fire risk index. The basic concept is to produce as good results as possible, with a minimum data set. Data is not always easy to find in time and sometimes might not be available. The final goal is to be able to produce a flexible system that will be able to use more than one data set, from any kind of available data, in order to produce results that will be helpful to the expert, in order to deploy human and vehicle power in the right place at the right time.

2 Research Area and Software

The research area of the DSS used for the assessment of the risk index is the Greek terrain. Data has been gathered from the Greek public services that are responsible for

the meteorological, morphological and statistical data. The initial data table have been processed through MS Access so that the results become more meaningful, but due to the lack of resent detailed information from the public services, the studies concerns a past period of time between 1983 and 1993. The processed data were inserted into Matlab in order to apply the fuzzy rule based decision support system. Reducts were calculated with the use of rough set theory in the Rosetta software [1], [2], [3], [4], [5]. The final results were imported into MS Excel, so that comparisons with reality are possible.

The Greek terrain has been separated into smaller areas, each one concerning a different forest department. The final results do not produce a specific risk index, but a ranking between each and every forest department under study, according to their overall forest fire risk.

3   Fuzzy set and Rough set theory

The use of computers for assigning risk degrees has its positives due to ease of computation but comes with a prize. The huge amount of data that is needed cannot always be described with classical mathematics and crisp sets. At this point, fuzzy set theory can be used to describe almost every situation. The fact that in fuzzy sets there exists a degree of membership (DOM) μs(X) that is mapped on [0,1] for every data entry helps in the production of rational and sensible clustering [9] and any element of the real world can belong to each set with a different degree of membership.

Fuzzy Sets use membership functions μs(x) which are mappings of the universe of discourse X on the closed interval [0,1]. That is, the degree of belonging of some element x to the universe X, can be any number μs(x), where $0 \leq \mu_s(x) \leq 1$. Below are the five major steps that should be followed when designing Fuzzy Models.

Step 1: Define the Universe of discourse (e.g. domains, input and output parameters)
Step 2: Specify the Fuzzy Membership Functions
Step 3: Define the Fuzzy Rules
Step 4: Perform the numerical part, like T-Norms, S-Norms or other custom inference
Step 5: If not-Normal Fuzzy sets exist, perform Defuzzification

The 3rd step in which the fuzzy rules must be defined, is actually the basic backdrop of any fuzzy – rule based system. The more the data attributes increase, the more accurate the system is, but the complexity of creating the rule set becomes very difficult. In an effort to reduce the data attributes as much possible various methods can be used varying from simple statistical ones, to more complex neural network systems.

One of the most promising methods, introduced by Zdzisław Pawlak in 1991 [13], which can be used for attribute reduction, is the Rough-set theory. Rough set theory is still another approach to vagueness. Similarly to fuzzy set theory it is not an alternative to classical set theory but it is embedded in it. Rough set theory can be viewed as a specific implementation of Frege's idea of vagueness, i.e., imprecision in this approach is expressed by a boundary region of a set, and not by a partial

membership, like in fuzzy set theory. Rough set concept can be defined quite generally by means of topological operations, interior and closure, called approximations. Given a set of objects U called the universe and an indiscernibility relation $R \subseteq U \times U$, representing our lack of knowledge about elements of U, we assume that R is an equivalence relation. Let X be a subset of U. We want to characterize the set X with respect to R. To this end we will need the basic concepts of rough set theory given below.

- The *lower approximation* of a set X with respect to R is the set of all objects, which can be for certain classified as X with respect to R (are certainly X with respect to R).
- The *upper approximation* of a set X with respect to R is the set of all objects which can be possibly classified as X with respect to R (are possibly X in view of R).
- The *boundary region* of a set X with respect to R is the set of all objects, which can be classified neither as X nor as not-X with respect to R.

Now we are ready to give the definition of rough sets.

Set X is crisp (exact with respect to R), if the boundary region of X is empty.

Set X is rough (inexact with respect to R), if the boundary region of X is nonempty.

4   The case of forest fire risk assessment.

The first step in assigning forest fire risk is related to the determination of the main n risk factors (RF) affecting this specific risk. The risk factors are distinguished in two basic categories; Human factors and Natural ones [11]. Another issue is that forest fire risk is a general concept. In order to be more precise, the expert needs to concentrate in a more specific kind of risk. In this case, there exist to basic risks:

1. A risk concerning the speed with which each fire is spread
2. A risk concerning the total amount of fire incidents that occur in a specific area during a given period of time

Of course these two different risk types are influenced in a different way by numerous factors that most of them are difficult to measure and also need to be measured for a long period of time. This project is going to concentrate on the first kind of risk, concerning the speed with which each fire incident is spread.

Forest fires can have various factors that can influence both of the above risk indicators, either long term or short term. The long term forest fire risk factors are distinguished in two basic categories; Human factors and Natural ones [11]. Each of these general risk types consists of several other sub-factors that influence in their own way the overall risk degree (ORD). Some of the major factors can be seen in the table below.

Of course there are multiple other factors that can influence the forest fire risk but it is not this paper's purpose to explore them in depth. Further separation of the "Natural" factors into two more sub-groups, the "Weather" and the "Landscape" factors is being made, so that it becomes easier to apply the rule set of the fuzzy decision support

system. The same applies for the "Human caused" factors that are being separated into "Measurable" and "Immeasurable".

Table 1 Factors used by the fuzzy rule based system

Level 1	Natural factors		Human caused factors	
Level 2	Weather factors	Landscape factors	Measurable	Immeasurable
Level 3	Temperature	Forest cover	Tourism	Land value
	Humidity	Altitude	Population	Other
	Wind speed			

5   The decision support system

The system was designed and developed under the Matlab platform. With the use of fuzzy logic and fuzzy sets, the system assigns a degree of the long term forest fire risk degree to each area by using Matlab's fuzzy toolbox. The functions used to define the DOM are called fuzzy membership functions (FMF) and in this project the triangular FMF (TRIAMF) and the semi-triangular FMF (semi-TRIAMF) were applied [10]. Functions 1 and 2 below represent the TRIAMF and semi-TRIAMF which are implemented in Matlab.

$$\mu_s(X) = \begin{cases} 0 \text{ if } X < a \\ (X - a)/(c - a) \text{ if } X \in [a, c] \\ (b - X)/(b - c) \text{ if } X \in [c, b] \\ 0 \text{ if } X > b \end{cases} \tag{1}$$

$$\mu_s(X) = \begin{cases} 0 \text{ if } X < a \\ (X - a)/(b - a) \text{ if } X \in [a, b] \end{cases} \tag{2}$$

The system assigns each forest department three Partial Risk Indices (PRI), for every one of the nine factors that are taken under consideration, as it is shown below:
- Low Danger due to each factor
- Medium Danger due to each factor
- High Danger due to each factor

This method allows the use of any kind of data and does not need specific metrics for every factor. Due to this fact, there was no need to do any changes in the row data provided by the Greek national services. The above steps resulted in having 27 different PRIs. The more detailed the linguistics become the greater the number of PRIs.

In order to unify those partial risk indices in one Unified Risk Index (URI), the system has to take into consideration the human experience and to apply the rules that a human expert would use. These rules are distinct for each risk factor and most of them are specified in bibliography [11]. The total amount of rules that has to be applied is 3^9=19683, but with the use of sub groups as shown in the table above, this number reduces to 33+6*32=81. This is accomplished by creating partial rule sets in order to combine the factors as shown in figure 1 below
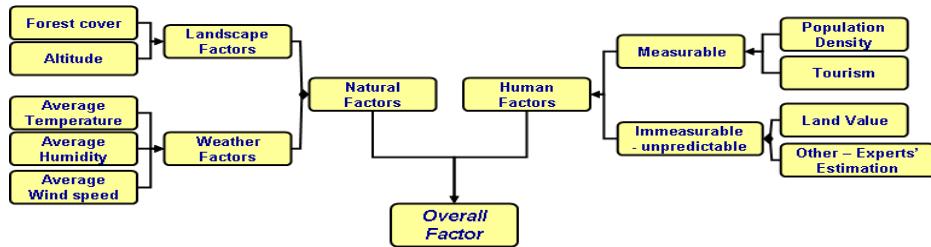
Figure 1 Categorization of risk factors

Previous studies [7], [8] have shown that this way the system retains its simplicity, while still producing the same results. This way, the system is able to provide a ranking of the Greek forest departments and the comparison of these results to reality gives the user a degree of "compatibility" with the actual ranking of these forest departments according to each year's total burnt area.

6   Attribute reduction with the use of rough sets

Despite the fact that the accuracy of the system was good, the fact that 9 factors had to be combined, resulted in a rather complicated amount of data sets. Another issue was also the fact that in order to find the optimal data set that could produce the best result would need a huge amount of tests. Each test would require the built of a new rule set. Building 511rule sets, which is the total number of combination of 9 attributes and testing each one of them for every year of the data available would be time consuming.

Rough set theory is a good tool for attribute reduction and with the help of the Rosetta software, the process of finding the reducts is much easier. The whole dataset, including the FFR-FIS' final risk degree index, was inserted into Rosetta and the reducts were calculated with the help of RSES dynamic reducer. The algorithm used was the generic RSES. The values used for the rest of the parameters are shown in the Screen 1 below.

This resulted in the production of 85 possible reducts. The 3 reducts with the best results in Rosetta were used in order to create 3 new different fuzzy rule based systems, with new rule sets. The basic idea was to compare these results with the first result that was produced from the whole dataset.
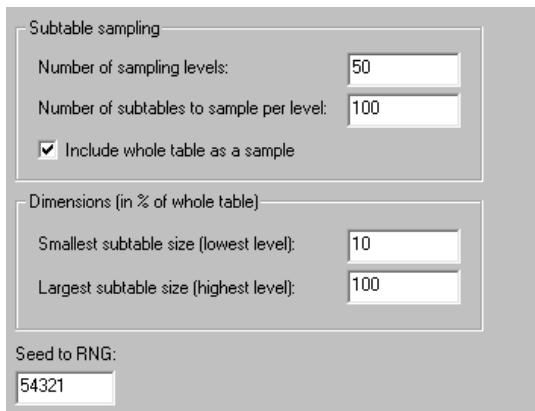
Screen 1 Values for the RSES dynamic reducer parameters

7 Fuzzy C-means clustering

Providing a ranking of the forest departments is a good way to compare the each other, but the actual ranking of those departments according to their burnt areas is not possible to be accurately found. This is due to the fact that the fire incidents are influenced by a great number of factors that are unpredictable. The speed of expansion on the other hand is influenced by more "predictable" factors but still the ranking does not provide a sharp image of reality. For this reason it is more helpful to cluster the forest departments and compare the results afterwards.

Table 2 Example of actual data set

Forest Department	Temperature	Humidity	Burnt area	Risk degree
A	14	0,1	1500	0.7785
B	17	0,4	1100	0.3722
C	22	0,7	200	0.2162

Table 3 Example of clustered data set

Forest Department	Temperature	Humidity	Burnt area	Risk Degree
A	Low	Low	Vast	High
B	Average	Low	Vast	Average
C	High	High	Small	Small

In this paper, the whole data set that is under study has been clustered with the use of fuzzy c-means clustering method. Each factor, the results of the forest fire risk fuzzy inference system and the "Burnt area" attribute have been clustered into three different clusters. This way each data entry can be described as shown in the example of tables 2 and 3 above.

This makes the comparison between results and reality easier. The user can compare whether a forest department that has a "vast" burnt area, like forest departments A and B, has been indexed as a "high" or "average" risk department. Comparison between these two clustered attributes gives an index of quality, about the final results of the fuzzy inference system.

8   Results and discussion

Fuzzy rule based systems can be helpful in many various occasions and forest fire risk assessment is one of these occasions. The built of a system that could utilize as many attributes as possible was the first stage. Results from this system varied from 25% to 70%. During the built of this system, it became obvious that the luck of data is a frequent problem. The accuracy gets smaller, when more risk attributes are missing and reaches its peak with a full data set. The creation of smaller systems that could produce acceptable risk indices was necessary. The utilization of a rough set attribute reduction system in order to find possible reducts, gave some interesting results. The 10 best reducts are shown in the table 4 below.

Table 4 The reducts that produce the best results

	Reduct	Support
1	{Altitude, Forest Cover}	5093
2	{Humidity, Altitude}	5093
3	{Temperature, Altitude}	5093
4	{Humidity, Wind Speed, Population}	4497
5	{Temperature, Forest Cover, Wind Speed, Other (expert's evaluation)}	3430
6	{Temperature, Forest Cover, Wind Speed, Tourism}	3264
7	{Altitude, LandValue}	2813
8	{Altitude, Population}	2813
9	{Humidity, Wind Speed, Tourism}	2728
10	{Temperature, Wind speed, Land Value}	2688

Human factors are known to be the most possible cause for forest fires, but they are not the basic cause for each forest fire incidents expansion. According to these results, only in 5 out of 10 reducts include human caused factors.

Careful study of these results gives some interesting discovering about forest fire risk in Greece. It is obvious it possible to study forest fire risk with just the use of long term factors alone. This shows that forest fires, during the period of time that is under study, which expand the most rapidly and cause the most damage are most of the times (almost 100%) in the same forest departments. 5093 cases out of the 5100 can be described with Average Altitude and Forest Cover as only knowledge about the terrain. It wouldn't require an "expert" to prove that almost all cases of big fire incidents were near the sea, or near cities that are by the sea, that also had great forest cover and unprotected forests.

Another interesting point of view is that the attribute "Altitude" is found in 6 of those 10 reducts. This also proves that the altitude of a forest department is one of the most basic factors that influence the risk indices. All the forest departments that are far away from the sea have greater altitude in Greece and none of them seems to have major fire incidents. Knowing the altitude of a forest department and with the use of just the humidity or the temperature, it is safe to say that this department is a risky one or not.

An interesting aspect of this study is the fact that the best reducts that occur are small, composed by only 2 or 3 attributes. For the first 3 reducts, a fuzzy rule based system was built and new indices were extracted. The results were not that accurate, reaching a maximum similarity with the real ranking of 43%. On the other side the application fuzzy c-means clustering on the results and the total burnt area of each department, and further comparison between those two, showed that it is possible to cluster the departments into "Dangerous" and "non-Dangerous" with great accuracy, reaching even 90% in some cases, with the use of just a couple of attributes.

The target of future studies is to find the optimum attributes set for the Greek terrain, that will provide results, as close to reality as possible. A basic categorization of the Greek forest departments due to their long term factors that influence the forest fires, in order to provide a standard map of the dangerous areas, with the combination of short term factors in order to rank the dangerous areas in daily basis is a future thought.

References

1. Øhrn, J. Komorowski (1997), ROSETTA: A Rough Set Toolkit for Analysis of Data, Proc. Third International Joint Conference on Information Sciences, Fifth International Workshop on Rough Sets and Soft Computing (RSSC'97), Durham, NC, USA, March 1-5, Vol. 3, pp. 403-407.
2. Øhrn, J. Komorowski, A. Skowron, P. Synak (1998), The Design and Implementation of a Knowledge Discovery Toolkit Based on Rough Sets: The ROSETTA System, In Rough Sets in Knowledge Discovery 1: Methodology and Applications, L. Polkowski and A. Skowron (eds.), Studies in Fuzziness and Soft Computing, Vol. 18, Chapter 19, pp. 376-399, Physica-Verlag. ISBN 3-7908-1119-X.
3. Øhrn, J. Komorowski, A. Skowron, P. Synak (1998), The ROSETTA Software System, In Rough Sets in Knowledge Discovery 2: Applications, Case Studies and Software Systems, L. Polkowski and A. Skowron (eds.), Studies in Fuzziness and Soft Computing, Vol. 19, pp. 572-576, Physica-Verlag. ISBN 3-7908-1130-3.

4. Øhrn (2000), ROSETTA Technical Reference Manual, Department of Computer and Information Science, Norwegian University of Science and Technology (NTNU), Trondheim, Norway.
5. Øhrn (2000), The ROSETTA C++ Library: Overview of Files and Classes, Department of Computer and Information Science, Norwegian University of Science and Technology (NTNU), Trondheim, Norway.
6. J. Komorowski, A. Øhrn, A. Skowron (2002), The ROSETTA Rough Set Software System, In Handbook of Data Mining and Knowledge Discovery, W. Klösgen and J. Zytkow (eds.), ch. D.2.3, Oxford University Press. ISBN 0-19-511831-6.
7. Tsataltzinos T. (2007) "A fuzzy decision support system evaluating qualitative attributes towards forest fire risk estimation", 10th International Conference on Engineering Applications of Neural Networks, Thessaloniki, Hellas, August 29-31, 2007.
8. Tsataltzinos, L.Iliadis,S.Spartalis,"An intelligent Fuzzy Inference System for risk estimation using Matlab platform: The case of forest fires in Greece", BOOK CHAPTER (PROCEEDINGS): "Artificial Intelligence Applications and Innovations III",5th IFIP Conference(AIAI'2009), Springer (ISBN 978-1-4419-0220-7), p. 303 -310, 2009.
9. Kandel A., 1992, Fuzzy Expert Systems. CRC Press. USA
10. Iliadis L., Spartalis S., Maris F., Marinos D. 2004 "A Decision Support System Unifying Trapezoidal Function Membership Values using T-Norms". Proceedings of ICNAAM (International Conference in Numerical Analysis and Applied Mathematics) J. Wiley-VCH Verlag GmbH Publishing co.Weinheim Germany.
11. Kailidis, D. (1990). Forest Fires.
12. E A Johnson, K. M. (2001). Forest Fire: Behavior and Ecological Effects.
13. Zdzisław, P. (1991) Rough sets: theoretical aspects of reasoning about data
14. He, H.S., Shang, B.Z., Crow, T.R., Gustafson, E.J., Shifley, S.R., 2004. Simulating forest fuel and fire risk dynamics across landscapes – LANDIS fuel module design. Ecological Modeling 180, p 135-151
15. Chandler, Craig C. 1983 Fire in forestry / Craig Chandler ... [et al.] Wiley, New York
16. Jan Komorowski , Zdzislaw Pawlak , Lech Polkowski , Andrzej Skowron (1998), Rough sets: A tutorial