The application of Rough Sets Theory as a data-mining tool to classify complex functions in safety management

Hussein SLIM, Sylvie NADEAU

Department of Mechanical Engineering, École de technologie supérieure Montréal, Québec, Canada

Abstract. In recent years, considerable research efforts in safety management were directed at proposing innovative methodological frameworks to address the complexity of modern sociotechnical systems. The significance of results in such endeavors, whether quantitative or qualitative, relies largely on the quality of input data and the validity of the implemented methods to model such systems. To provide more objective and valid results, new protocols and tools for data processing are needed as well. An interesting data-mining tool for computing with incomplete and uncertain information is Rough Set Theory (RST). In this study, we propose the application of RST to generate comprehensible IF-THEN rule bases for classifying outcomes within the framework of the Functional Resonance Analysis Method (FRAM). The steps for the integration process of both frameworks are introduced in this paper and an illustrative example is consequently provided to demonstrate a possible approach for realizing the combination. Such an approach could allow for an efficient rule generation and data classification process, which could aid in addressing classification challenges and input data limitations in safety management. The model however still requires further optimization and validation using expert's input data in future applications.

Keywords: FRAM, rough sets, safety management, sociotechnical system, functional resonance, decision making

1. Introduction

Technology in recent years has been making huge leaps and several game-changing applications were introduced lately reshaping how systems function and behave. This evolution can pose challenges for system analysis and safety management in the years to come. The field of safety management in recent years has been witnessing significant research efforts as well emphasizing the need to adopt additional perspectives. Consequently, innovative tools such as the Functional Resonance Analysis Method (FRAM) (Hollnagel, 2004) were introduced to address the objectives of safety analysis from a new systemic perspective. FRAM, as a resilience engineering tool, adopts new concepts for redefining safety such as SAFETY-II and the distinction between work as imagined (WAI) and work as done (WAD). Performance variability in FRAM is a natural characteristic of any sociotechnical system and is even considered necessary and beneficial. The principles of FRAM allow for characterizing complex and dynamic relationships using qualitative scales expressed in natural language. However, several limitations can be observed in such analyses such as lack of data, uncertain information, and classification problems. Therefore, the adoption and standardization of such innovative tools is faced with many challenges, which require further research to provide more representative and reliable results.

In our project, we directed our efforts at exploring new approaches to address safety and performance challenges in the field of aircraft deicing. To this end, we applied the Functional Resonance Analysis Method (FRAM) at a previous stage in conjunction with fuzzy logic. The objective thereby was to introduce a systematic methodology to account for internal variability considering present performance conditions and generate a quantified and more precise representation of the output's variability. The prototyping model was however still faced with several limitations. The application with a high number of variables, associated phenotypes and classes translated into a significant number of rules. To avoid the rules' explosion problem and construct an efficient model, the number of variables and associated classes was limited, and the impact of the phenotypes was simplified. Additionally, the consequent part or the decision class was not always easily identifiable using qualitative scales. The vagueness of the provided input information could affect the decision-making process and make the assignment of a decision class a difficult task to achieve in many cases. In a predictive assessment, it would not be always possible using a qualitative scale to determine whether the output would be variable and to which extent. Differences between expert judgements to conclude a definitive decision can often be faced in such assessments. Therefore, we aimed in the next step at proposing a possible solution to address these limitations. In the third phase of this project, we proposed the application of Rough Sets Theory (RST) as a data-mining tool to generate a more efficient rule base and classify outcome relying on historical and recorded data. The integration process shall be presented in a simplified way in this paper.

2. Rough Set Theory (RST)

The concept of Rough Sets was introduced by Zdzislaw Pawlak in 1982 (Pawlak, 1982). RST provides mathematical tools for processing imperfect information, classify data sets, and generate comprehensible conditional rules in the presence of uncertainty, inconsistency and incompleteness of information. The RST approach has proven to be helpful so far in solving problems of data mining and classification in several fields such as machine learning, decision analysis, expert systems, and pattern recognition. It can therefore offer promising solutions in the field of safety management generally and specifically to recently emerging concepts and tools such as FRAM and the discipline of Resilience Engineering. Complex analysis tools as FRAM rely mainly on natural language to characterize variables in question, for which RST could be helpful. The generated rules are easily understandable and offer a straightforward interpretation of the obtained outcomes.

An information system (**IS**) in RST is a two-dimensional matrix or data table consisting of a pair of sets (**U, A**), namely a finite non-empty set of objects (**U**) and a finite non-empty set of attributes (**A**) so that: $\mathbf{a: U \rightarrow V_a}$ for every $\mathbf{a \in A}$, where $\mathbf{V_a}$ is the value set of **a** with respect to each object (**U**). A decision system (**DS**) consists accordingly of the **IS** adding a decision set **D** ($\mathbf{d \notin A}$) such that $\mathbf{DS = (U, A \cup (D))}$.

A rough set has a boundary region, which contains objects that cannot be classified with certainty as members of the set or of its complement. This means that the available information is not sufficient to definitively classify these elements. In RST, any set

Table 1. *A decision system in RST*

Set of Objects U	Set of Attributes A				Decision D
	A_1	A_2	.....	A_n	
U_1	V_{11}	V_{12}	.....	V_{1n}	D_1
U_2	V_{21}	V_{22}	.....	V_{2n}	D_2
U_3	V_{31}	V_{32}	.....	V_{3n}	D_3
.....	......	......	.....	......	.....
U_m	V_{m1}	V_{m2}	.....	V_{mn}	D_m

of objects is replaced by a pair of precise sets, called the lower and the upper approximations. The lower approximation consists of all objects that are certain members of the original set, and the upper approximation contains all objects that could possibly belong to the original set (Pawlak, 1982). The difference between the upper and the lower approximation constitutes the boundary region. Approximations are two basic operations in rough set theory. The principle of indiscernibility forms the basis, which is utilized to identify equivalence classes. The indiscernibility relation is a binary relation, which represents the sets of objects for which a decision cannot be discerned given a specific array of values of the attributes. The set of indiscernible objects form an equivalence class. A discernibility matrix is then constructed for the equivalence classes to determine the respective discernibility functions and reducts. To this end, efficient algorithms can be utilized for identifying hidden patterns in data tables to produce minimal sets of data (data reduction), evaluating the significance of data, and generating representative sets of decision rules.

3. Proposed Approach: Rough FRAM

To integrate RST with FRAM, the FRAM function would be redefined as an RST decision system in the form of a table consisting of objects, which represent the many iterations of the function recorded over time, and the set of respective attributes, which would then represent the functional aspects as defined in FRAM. The values assigned to the objects with regard to each attribute would therefore consist of the classes that each functional aspect can assume: {dampening, variable or unpredictable}. Accordingly, the five steps of FRAM would be structured as follows:

Step Zero: The start is with defining the objective of analysis and what is to be achieved. This step is unchanged and would define the context and type of application needed to achieve the defined objective.

Step One: Next, the set of functions that constitute the system is defined and characterized specifying the aspects for each function and accordingly the relationship to other functions since the output of upstream functions serves as an input or an incoming aspect for downstream ones. For each function, there are six aspects: input, preconditions, time, control, resources, and output (Hollnagel, 2012). The characterization of the functions is decisive to determine the type of data needed for constructing the data tables. The recorded data can be generated relying on expert input and thereafter entered into the RST information system. Using the principles of approximation and indiscernibility, the set of equivalence classes can be identified, and a discernibility matrix is then built. The discernibility matrix is then completed to identify the set of reducts and accordingly generate a reduced set of rules, which could maintain the same accuracy of the original set of attributes.

Step Two: Performance variability is classified usually in basic FRAM using two phenotypes, each with a qualitative three-point scale, namely precision and time. In our prototyping model, we simplified the scale further to one phenotype to minimize the number of generated rules and since it made more sense in a predictive assessment. Each aspect is classified accordingly as: {dampening (or non-variable), variable and unpredictable (or highly variable)}. The data table is then fed to the RST software to compute reducts using a genetic algorithm and consequently generate the reduced set of conditional If-then rules.

Step Three: A specific analysis scenario is selected for running instantiations of the developed model with specified performance conditions. A list of performance conditions can be used to anticipate the potential of each function to produce a variable output due to the internal variability coming from within. Depending on the present conditions, the internal variability for each function can be determined and thereafter, the output's variability and its resonance and impact on other functions can be tracked within the system using the graphical representation as a map of the dominant relationships within the studied system.

Step Four: For variability management, the results of the FRAM analysis can be used as indicators to point to possibilities of variable outputs and accordingly measures can be implemented to ensure preferable performance conditions that promote a resilient systemic behavior. For further details on the developed model, the reader is advised to consult Slim & Nadeau (2020).

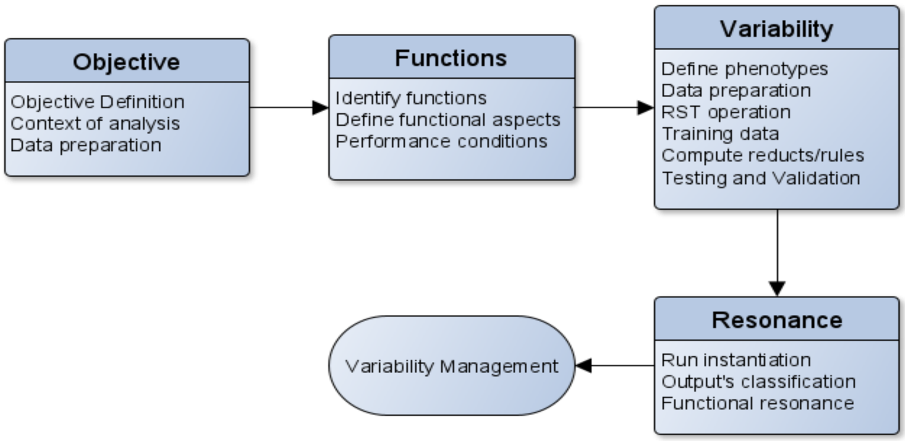

Figure 1. *A graphical illustration of the steps of Rough FRAM.*

4. An illustrative example: A FRAM function as an RST decision system

In this section, an illustrative example is presented to demonstrate how a FRAM function can be constituted as an RST decision system. To this end, we assume that we have a FRAM function with four connected incoming aspects: {Input 1, Input 2, Resource and Control} and one {Output}. The four ingoing aspects are accordingly defined as attributes {A1, A2, A3, A4} and the output as the decision class {D} of the function, which can possess one of three possible qualities: {dampening, variable and unpredictable}. Twenty-six random instances of the function are recorded, each of which represents an object in the decision system. We can accordingly construct our FRAM decision system as shown in Table 2.

Table 2. *A data table showing the example described in this section*

Func-tion	Input (A1)	Input (A2)	Resource (A3)	Control (A4)	Output (D)
1	unpredictable	unpredictable	unpredictable	variable	unpredictable
2	variable	variable	variable	variable	unpredictable
3	variable	unpredictable	dampening	unpredictable	unpredictable
4	variable	variable	variable	unpredictable	unpredictable
5	unpredictable	unpredictable	dampening	unpredictable	unpredictable
6	unpredictable	dampening	variable	unpredictable	unpredictable
7	dampening	dampening	variable	dampening	dampening
8	unpredictable	variable	dampening	variable	unpredictable
9	unpredictable	variable	dampening	dampening	variable
10	dampening	dampening	unpredictable	unpredictable	unpredictable
11	dampening	dampening	unpredictable	variable	variable
12	dampening	dampening	unpredictable	dampening	variable
13	variable	variable	variable	variable	variable
14	variable	unpredictable	dampening	dampening	variable
15	variable	variable	unpredictable	unpredictable	unpredictable
16	variable	variable	unpredictable	variable	unpredictable
17	variable	dampening	dampening	variable	variable
18	variable	dampening	dampening	dampening	dampening
19	unpredictable	unpredictable	unpredictable	dampening	unpredictable
20	dampening	variable	variable	dampening	variable
21	dampening	variable	dampening	unpredictable	variable
22	dampening	variable	dampening	variable	variable
23	dampening	variable	dampening	dampening	dampening
24	dampening	dampening	variable	unpredictable	variable
25	dampening	dampening	variable	variable	variable
26	dampening	variable	dampening	dampening	variable

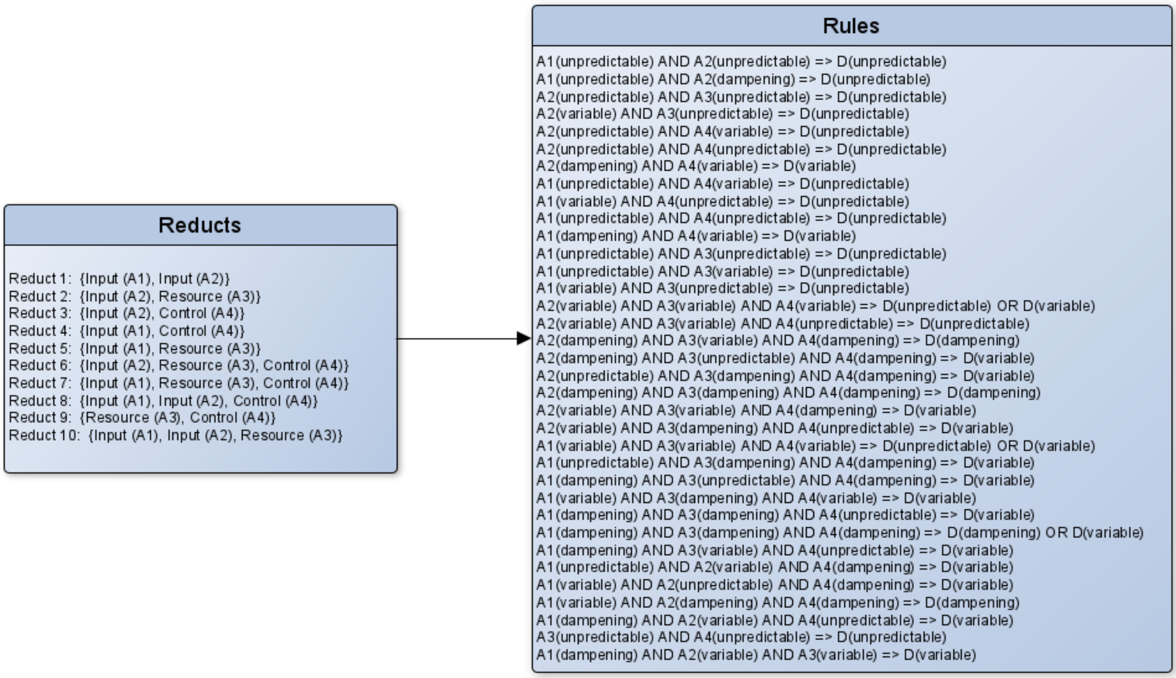

Figure 2. *The computed set of reducts and the consequent reduced set of conditional rules*

The data table can be constructed in a real-world setting using historical data or data collected from field observations. The table above is used as the training set, which is processed using a searching algorithm to identify patterns and determine equivalence classes, and consequently construct the discernibility matrix. The set of reducts and the rule base are determined by simplifying the discernibility function. The rules can be tested running many instantiations of the FRAM model. The reduced set of rules generated considering the set of reducts in exchange for the original set allows for a more efficient rule base. The accuracy depends significantly on the quality of provided data concerning size, consistency and completeness. A threshold of accuracy can be defined to consider only reliable rules and discard of insufficiently accurate ones (Figure 2).

5. Conclusions

Innovative solutions are needed in safety management to address challenges associated with modern sociotechnical systems. Tools as RST can be helpful in this regard offering tools to address uncertain information. Systemic methods as FRAM combined with such approaches can offer promising and practical solutions. This paper shows how the basic FRAM model can be combined with the RST method by providing a simplified example of a FRAM function redefined as an RST decision system. The combined approach can be helpful to address limitations concerning limited input data, inconsistencies, incompleteness, and output classification. The RST approach allows as well to produce reduced and efficient rule bases, which can be used in conjunction with fuzzy logic to provide a more intersubjective representation of performance variability. The conditional IF-THEN rules are more comprehensible and can be automatically deduced from the provided data table recorded from field observations or retrieved from archived data. The FRAM framework with its principles rooted in Resilience Engineering allows for characterizing complex relationships and interdependencies within the system in question. The phenotypes were simplified here to facilitate the integration with RST and allow for an easier predictive assessment at this stage. Going forward, the model would require further optimization and validation providing a full-fledged model and building on a real-world case study. The approaches discussed in this paper are solely for illustrative purposes and provide a mere skeleton of the model, which should be further developed and improved in future projects.

6. References

Hollnagel E (2004) Barriers and Accident Prevention. Aldershot, UK: Ashgate Publishing.

Hollnagel E (2012) FRAM, the functional resonance analysis method: modelling complex socio-technical systems. Ashgate Publishing, Ltd.

Pawlak Z (1982) Rough sets. International Journal of Computer & Information Sciences, 11(5), 341-356.

Slim H, Nadeau S (2020) A mixed rough sets/fuzzy logic approach for modelling systemic performance variability with FRAM. *Sustainability*, *12*(5), 1918.

Acknowledgement. The authors warmly thank the Arbour Foundation, the National Sciences and Engineering Council of Canada (NSERC) & École de technologie supérieure (ÉTS) for funding this project.

Gesellschaft für
Arbeitswissenschaft e.V.

Arbeit HUMAINE gestalten

67. Kongress der
Gesellschaft für Arbeitswissenschaft

Lehrstuhl Wirtschaftspsychologie (WiPs)
Ruhr-Universität Bochum

Institut für Arbeitswissenschaft (IAW)
Ruhr-Universität Bochum

3. - 5. März 2021

GfA-Press

Bericht zum 67. Arbeitswissenschaftlichen Kongress vom 3. - 5. März 2021

Lehrstuhl Wirtschaftspsychologie, Ruhr-Universität Bochum
Institut für Arbeitswissenschaft, Ruhr-Universität Bochum

Herausgegeben von der Gesellschaft für Arbeitswissenschaft e.V.
Dortmund: GfA-Press, 2021
ISBN 978-3-936804-29-4

NE: Gesellschaft für Arbeitswissenschaft: Jahresdokumentation

Als Manuskript zusammengestellt. Diese Jahresdokumentation ist nur in der Geschäftsstelle erhältlich.

Schriftleitung: Matthias Jäger
im Auftrag der Gesellschaft für Arbeitswissenschaft e.V.

Screen design und Umsetzung
© 2021 fröse multimedia, Frank Fröse
office@internetkundenservice.de · www.internetkundenservice.de