An Economic Approach to Business Rules Normalization

Martijn Zoet  
Optimizing Knowledge-Intensive Business Processes  
Zuyd University of Applied Sciences  
Sittard, the Netherlands  
martijn.zoet@zuyd.nl

Koen Smit  
Digital Smart Services  
HU University of Applied Sciences Utrecht  
Utrecht, the Netherlands  
koen.smit@hu.nl

Abstract—This paper presents a cost/benefit analysis method for the normalization of business rules. To determine the economic benefit of business rules normalization three variables are addressed: 1) the number of anomalies a rule set endures, 2) the storage space a rule set requires and the 3) deterioration of rules in response time. The approach is evaluated by means of an experiment, based on mortgage data of an international bank. Results show that the method is useful for determining when to normalize business rule sets; the method enables business rules analysts to produce more cost-effective business rules architectures.

Keywords—Business Rules; Decision Management; Normalization; Cost-Benefit Analysis

I. INTRODUCTION

Good decision making is a key denominator for a corporation’s competitiveness [2]. Therefore, organizations are increasingly urged to make fast and accurate decisions. At the same time, decisions are becoming more and more complex affecting maintainability and transparency. Decisions can be formulated by means of business rules [22]. A business rule is defined by Morgan [13] as: “a statement that defines or constrains some aspects of the business intending to assert business structure or to control the behavior of the business.”

To realize changes within an organization’s decision-making process, an organization should be able to maintain the aforementioned asserts and it should be able to adapt its business rules efficiently and effectively to realize changes within its decision-making process. In order to realize this, information systems, such as expert systems, knowledge management systems, case based reasoning systems, fuzzy expert systems and business rules management systems have been built for and adopted by organizations [12].

Research on the management of business rules has been conducted since the mid-1960’s [12]. Distinct research streams have emerged, focusing on the following three subjects: 1) subject transformation, 2) platform transformation, and 3) business rule model transformation [21]. Subject transformation research focuses on processes, methods and information systems used for mining and cleansing decision sources, such as regulations, organizational policies, laws, documents and databases. The second stream focuses on the use of information technology for the deployment, execution and monitoring of business rules. Important research topics are: 1) algorithms for faster and easier execution, 2) business rules architectures, and 3) business rules engines [1][6][15]. Business rule model transformation research focuses on verification, validation and improvement of existing business rules. To verify business rules, a formal grammar notation and/or a set of constructs is applied. A grammar notation describes how a business rule should be constructed or formulated. An example of a standardized business rules grammar is the Semantics of Business Vocabulary and Business Rules [16].

Despite the accumulation of literature, there is a surprisingly scarce amount of research that examines methods and processes to factor business rules [22]. Factoring entails the process of dividing business rules, and therefore decisions, in more comprehensible structural elements to increase maintainability and transparency. Research that has focused on this subject is “single language oriented” [21][22][23]. Since a relatively high number of business rules modelling languages exist within scientific and professional literature, a factoring procedure per language is not desired from the viewpoint of the authors. Furthermore, current research does not provide guidelines to financially quantify the value of factoring business rules. As far as the authors are aware, no method exists that is business rules modelling language-independent in combination with quantifying the financial benefits of factoring given business rules. An example is the work of [23] which solely focuses on achieving the third normal form while factoring business rules, without investigating whether this is financially optimal. Given the fact that organizations invest large amounts of money for implicitly managing business rules, a valid question is whether and when an explicit factoring procedure is economically beneficial. For example, a business rule set, which only changes or is executed twice a year might, from an economic perspective, be better off in an un-factored form. Taken previous statements into account, the following research question arose: “How can business rules be factored such that economic beneficial manageability is realized?” Following Van Thienen and Snoek’s [18] research on factoring decision tables and Zoet’s [22] research on factoring business rules, we adopt relational theory to factor business rules.

The current study extends previous research by developing a factoring method that incorporates mainstream rule modeling languages and guidelines to determine the cost and revenue of
(re-)factored business rules. We developed a factoring method and validated it by means of an experiment based on case study data at a large international bank. The results showed that our method is effective in determining the economic costs and benefits.

In section two, we provide a discussion on the theoretical foundations of factoring business rules in terms of relational theory, normalization and economic factors. This is followed by the construction of the method in section three. In section four we demonstrate the application of the method on mortgage decision making at a large international bank. We conclude this paper, in section five, with the study’s core findings, contributions as well as its limitations.

II. BACKGROUND AND RELATED WORK

There are few methods available to (re-)factor business rules [22]. Currently, two different methods are described: one by Van Thienen and Snoeck [18] and one by [22]. Van Thienen and Snoeck’s [18] method has two underlying assumptions (1) business rules are specified in decision tables and (2) relational theory is the basis for normalizing business rules. Guidelines are proposed to factor decision tables, thereby improving maintainability. However, instead of formulating one common procedure they proposed multiple exceptions to the normal form. These exceptions have to be formulated, which is an implicit result of the foundation of their research namely the use of decision tables. The second method proposed by [22] also takes relation theory into account. Moreover, this method distinguishes itself by applying one common procedure, which can be used for several languages.

The definition of the term relational as used in this paper is adopted from the mathematical domain, more specifically from the relational algebra theory [4]. Relational algebra theory has received a lot of attention during the last four decades, since it is popularized by Codd [4] for database normalization. The basic idea of the relational algebra theory involves that a relationship (R) can exist of a given set of elements (Sn), visualized as follows: R = (S1, S2, ..., Sn) [4]. The elements (Sn) can be condition- or conclusion-facts. Most authors [4][9] represent element sets by applying two-dimensional arrays. In order to apply relational theory on business rules, one must be able to translate business rules to sets of relationships. Previous research has answered the question [22] whether current business rule modelling languages can be translated to unified views by applying relational algebra theory. Based on representational difference analysis, the authors show that the six most common business rules languages can be transformed to sets of relations. Representational difference analysis is a technique, which is used to identify differences and overlap between concepts or constructs in ontology’s, languages and visual syntax [8] zur Muehlen and Indulska [20]. The six languages which were examined during this study are: If-Then business rules [17], Decision Tables [10] Van Thienen and Snoeck [18], Decision Trees [3], Score Cards [14], Event, Condition & Action Business Rules [5], and Event Condition Action Alternative Business Rules [7]. By translating business rules to relations between specific sets of elements, normalization is made possible. Normalization is the process of removing partial dependencies and transitive dependencies [4][9].

III. METHOD CONSTRUCTION

A detailed explanation of the business rules normalization procedure can be found in [22]. However, to ground our research, a summary of the normalization procedure is provided in sub-section A. Subsequently, in sub-section B, we described the cost reduction analysis method for business rules normalization.

A. Business Rules Normalization Procedure

The process for business rules normalization consists of three activities. The results of these activities are (1) the transformation of business rules to the proper relational structure, and (2) the removal of partial and (3) the removal of transitive dependencies. The latter is realized by applying the third normal form, while the second normal form deals with partial dependencies and the 1st normal form deals with achieving the proper structure for business rules.

The first normal form is realized by duplicating the original business rules equally often as the amount of conclusion-facts that exist. In other words, all of the duplicated rules exist of all condition- and conclusion-fields. The difference between the original and new tables is that only one of the original conclusion-fields is now still a conclusion-field while the others are condition-fields. In order for a relation to be in the second normal form, all condition-facts must be functionally dependent on a conclusion-fact and adhere to the first normal form. Condition-facts, which are not fully dependent on the conclusion-fact must be deleted or added to another relationship. The second normal form reveals whether condition-facts are included that actually do not contribute to a conclusion. To realize the third normal form in business rules, condition-facts that are not fully dependent on the conclusion-fact (but on another condition fact) should be removed and added to a new relation. The new relation contains the removed condition-facts, as well as the conclusion-fact to which they are related. A relationship is established between two sets of relations by means of a secondary decision. After applying the third normal form, all specified relations do not contain any repeating groups, partial dependencies and transitive dependencies anymore.

To visualize the normalization procedure a decision tree can be used [19]. A decision tree consists of two types of nodes: 1) normalization decision nodes (squares) and 2) end nodes (circles), for example see Fig. 1. A normalization decision node represents the decision to further normalize the relationship. From a normalization decision node, two types of branches can emerge: 1) a stop branch, and 2) a normalization branch. A stop branch emerges when further normalization is not needed, consequently leading to an end node. When further normalization is needed, two or more normalization branches emerge from the decision node. These branches lead to other decision nodes representing the newly normalized relationships.

Copyright (c) IARIA, 2017. ISBN: 978-1-61208-542-5
End nodes do not have further identification information, whereas normalization decision nodes do. Each node starts with the capital letter R, which is an abbreviation for relationship. The digit before the decimal point shows the number of the relationship. In case two digits are included before the comma, it designates a relationship resulting from another relationship. Furthermore, the digit after the decimal point indicates in what normalization form the relationship resides. In our example (see Fig. 1), the node R1,2 means that relationship 1 is in the second normal form. Moreover, the nodes R11,3 and R12,3 are both in the third normal form and are a relationship resulting from R1,2.

![Decision Tree for Normalization](image)

**Figure 1. Decision Tree for Normalization**

### B. Cost Reduction Analysis Method for Business Rules Normalization

Currently, in most normalization procedures the decision to normalize is generally based on intuitive flair. It remains uncertain whether the normalization effort is economically beneficial. For example, from an economic perspective, a business rule set, which only changes twice a year may not be beneficial to normalize.

Lee [11] and Westland [19] have conducted research towards the cost reduction of database normalization, which is based on relational theory. Cost reductions realized by database normalization are 1) decreased machine time, and 2) decreased data-inconsistencies (avoiding loss of business). The three main drivers of cost reduction are a) reduced anomalies, b) reduced storage requirements, and c) deteriorated response time. Anomalies that occur to data are: update-anomalies, insert-anomalies and deletion-anomalies [4]. Previous research has shown that database normalization principles can be applied to business rule sets [22]. Taken previous statement into account, the following question arose: Can the cost reduction model from database normalization be adopted as well?

Before adopting and adapting the model for business rules normalization, first the fit between the database determinants and business rules determinants has to be investigated. First, both the relations of data and business rules need to be updated and deleted, and new data or business rules have to be inserted. Second, previous research [11] has shown that business rules normalization can also lead to fewer storage requirements, such as the case is with database normalization. Thirdly, deteriorated response time is an important issue since decision making in organizations is increasingly complex with for example predictive analytics. As such, we can adopt the formulas proposed by Lee [11]. However, before the formulas can be used, the variables need to be adapted towards business rules.

The remainder of this section will discuss the formulas provided by Lee altered towards business rules.

The cost reduction realized by normalization is calculated in four phases 1) cost reduction due to reduced anomalies, 2) cost reduction due to reduced storage space, 3) cost increase due to increased join processing, and 4) comparing cost reduction due to reduced anomalies and cost reduction due to reduced storage space with the cost increase due to increased join processing.

Let \( \phi \) be the cost reduction due to reduced anomalies, see also equation 1. We define \( \phi \) as:

\[
\phi = \sum_{M=1}^{N}\alpha_{M}^{u}\lambda_{M}^{u}\omega_{M}^{u} + \sum_{M=1}^{N}\alpha_{M}^{l}\lambda_{M}^{l}\omega_{M}^{l} + \sum_{M=1}^{N}\alpha_{M}^{d}\lambda_{M}^{d}\omega_{M}^{d}
\]

**Equation 1. Cost reduction due to reduced anomalies**

Where \( N_u, N_l, \) and \( N_d \) are the number of updates, number of inserts and number of deletions, respectively. \( \lambda_{M}^{u}, \lambda_{M}^{l} \) and \( \lambda_{M}^{d} \) denote the frequency of the \( m \)th update, the \( m \)th insertion and the \( m \)th deletion. The average number of business rules affected by the update, insertion and deletion are denoted by \( \omega_{M}^{u}, \omega_{M}^{l} \) and \( \omega_{M}^{d} \). Furthermore, \( \alpha_{M}^{u}, \alpha_{M}^{l} \) and \( \alpha_{M}^{d} \) denote the cost for each insert, update and deletion.

Let \( \psi \) be the cost reduction due to reduced storage space, see also equation 2. We define \( \psi \) as:

\[
\psi = B_{X}\omega_{x} - B_{Y}\omega_{y}
\]

**Equation 2. Cost reduction due to reduced storage space**

Where \( B_x \) and \( B_y \) denote the storage cost per business rule in the current normalized situation. \( B_x \) and \( B_y \) denote the storage cost per business rule in the normalized situation + 1. The number of business rules stored in the current normalization situation is depicted by \( \omega_{x} \), while the normalized situation + 1 is depicted by \( \omega_{y} \).

Let \( \Omega \) be the cost increase due to increased join processing, see also equation 3. We define \( \Omega \) as:

\[
\Omega = \sum_{M=1}^{N}\gamma_{M}^{\mu_{m}}\omega_{x}\omega_{y}
\]

**Equation 3. Cost increase due to increased joint processing**

Where \( \gamma_{M}^{\mu_{m}} \) represents the number of joins required to determine the conclusion of a specific decision. \( \gamma_{M} \) denotes the cost per execution per business rule for join \( M \). Moreover, \( \mu_{m} \) represents the frequency of join \( M \). The time to realize the join is denoted by \( \omega_{x} \) and \( \omega_{y} \). The business rule sets \( (x, y) \) between which the join \( M \) is realized, is denoted by \( x, y, \in \mathbb{M} \). Let \( O \) be the cost reduction from normalization form R (R1,2) to normalization form R+1 (R11,3). We define \( O = \phi + \psi \geq \Omega \). \( O \) can be either positive or negative. If \( O \) is positive, then normalization should be applied.

### IV. Experiment Setup

In our validation, we apply an experiment on case study data. This allows us to use data from an actual case while fully controlling the execution of the method and input variables. The method is applied to a mortgage decision of an Anonymous International Bank (AIB). Our choice to select this case study setting was based on two theoretical criteria. Firstly, the case had to provide a proper amount of business rules used to take a
decision. The mortgage decision at AIB consisted of 665 facts (conditions and conclusions), and 1479 individual business rules. Secondly, the organization had to be willing to provide the financial details needed to perform the calculations. AIB agreed to this, however, with two demands. The first demand implied that their name and financial data were altered when it would be published. The second demand entailed that the applied business rule sets were not published. Since space limitations do not allow to walk through the entire mortgage decision and normalization procedure, both demands are met.

**Figure 2.** Photo impression 1 of normalized business rules

**Figure 3.** Photo impression 2 of normalized business rules

The evaluation, by means of conducting an experiment, was divided into three phases. Phase one was used to make the researchers familiar with the case parameters, by analyzing 133 pages with descriptions of decisions for completeness and accuracy. This phase resulted in the identification of multiple gaps. With the help of additional documentation and experts these gaps have been fixed. During the second phase, the business rules have been normalized according to our method. This normalization was done on paper after which the results were presented on a big wall (see Fig. 2 and Fig. 3). During the normalization, additional gaps were identified. These gaps have been marked with “post-its”, see Fig. 2 and Fig. 3. Again, with the help of additional documentation and experts, these gaps have been filled.

**Figure 4.** Decision tables to determine judgment personal situation

To ground our method, we explain the determination of the cost reduction from normalization form R to normalization form R+1 for the business rule set “personal situation of applicant” from the case described in the previous section. The business rule set consists of 10 facts, 1 conclusion fact and 8 condition facts; see left side of Fig. 4. The question that needs to be answered before normalizing this business rule set is: “Does normalizing the business rule set from R to R+1 realize a cost reduction?”

The decision personal situation is mainly affected by update and insert anomalies. For example, the facts “judgment age” and “judgment age savings” are updated regularly. Insert anomalies occur when new type of rules for age determination are inserted. The application of the method exist out of four phases 1) determine benefits in terms of reduced anomalies, 2) determine savings of storage requirements and 3) determine effect on response time, and 4) comparing cost reduction due to reduced anomalies and cost reduction due to reduced storage space with the cost increase due to increased join processing.

During phase one, three steps can be distinguished. **Step one:** determine the type of update, insert and deletion operations on a specific business rule set. In our case, “update judgment age” and “insert age determination rule”. For each identified operation type, it should be determined if the operation is affected by anomalies. If anomalies do not occur, normalization is not needed at all. If anomalies do occur, the frequency of each operation type and the number of business rules that are affected should be determined, this corresponds to step **two**. In this specific case \( \lambda_U^1 = 7 \) (per 2 weeks), and \( \lambda_U^2 = 6 \) (per 2 weeks). Additionally, the number of business rules affected by each update needs to be determined. In this specific case \( \omega_U^1 = 2 \) and \( \omega_U^2 = 1.5 \). During step **three**, the cost of an anomaly should be determined. In this case, the cost of a person
that adjusts the specific business rules $\alpha_2 = \€35.00$ per instance and $\alpha_3 = \€52.50$ per instance, see also equation 4. So, the total benefit due to reduced number of anomalies is:

$$\psi = (35 \times 7 \times 2) + (52.5 \times 6 \times 1.5) = \€962.50$$

Equation 4. Total benefit due to reduced anomalies

The first step of phase two is to determine the results of the transformation in terms of business rule sets. In this case, one business rule set (personal situation) is divided into three business rule sets namely 1) judgment personal situation, 2) judgment age, and 3) judgment internationality. The results of the normalization are shown in Fig. 3. For each business rule set, the number of business rules must also be determined, in this case, respectively, $\omega = 20, \omega_2 = 3, \omega_3 = 6$. During the second step, the cost per stored business rule must be determined. This needs to be determined for the current situation as well as for the post normalization situation. This information was retrieved from the information technology department, in this case, respectively, $B = \€4, B_x = \€0.5, B_y = \€0.5$ and $B_z = \€0.75$. Duplications are removed, thereby decreasing the number of individual business rules, see also equation 5. The total benefit due to reduced number of anomalies is:

$$\gamma = 20 \times 4 - 2 \times 0.5 - 3 \times 0.5 - 4 \times 0.75 = \€73.00$$

Equation 5. Total benefit due to reduced number of anomalies

To form a decision, two joins are required in the new situation, so $\emptyset = 2$. The cost for each join is $\€0.015$. The execution frequency of the join is 4000 per two weeks ($\mu_m$), see also equation 6. The additional cost due to additional join operations ($\Omega$) is therefore:

$$\Omega = 0.015 \times 4000 \times (2 + 3 + 6) = \€660.00$$

Equation 6. Total additional cost due to additional join operations

In conclusion, further normalization for the decision personal situation is recommended since $(962.50 + 73.00) > \€660.00$. Assume a situation where $\lambda_2 = 7/(\text{per 2 weeks}), \lambda_3 = 6/(\text{per 2 weeks})$ are decreased to $\lambda_1 = 2/(\text{per 2 weeks}), \lambda_2 = 2/(\text{per 2 weeks})$. Applying these changes reduces $\phi$ from $\€962.50$ to $\€446.25$, which changes $\Omega$ from $(962.50 + 73.00) > \€660.00$ to $(446.25 + 73.00) < \€660.00$ in which case further normalization would not realize a cost reduction.

The above example has shown a situation in which normalization leads to cost reduction and therefore the normalization should occur. By changing two parameters, we showed that normalization would lead to a negative cost reduction therefore an increase in cost and normalization should not be performed.

VI. EXPERIMENT VALIDITY

Internal validity threats, when conducting controlled experiments, can be classified into nine categories: 1) ambiguous temporal precedence, 2) selection, 3) history, 4) maturation, 5) regression, 6) attrition, 7) testing, 8) instrumentation, and 9) additive and interactive effect of threats to internal validity (Shadish et al., 2002). Ambiguous temporal precedence indicates a lack of clarity of variable occurrence, thereby influencing the cause and effect relation. In our research, temporal precedence occurs when decisions are transformed from source code to business rules management systems. The cost to realize an anomaly within the source code is higher compared to changing a business rule in a business rules management system. To reduce the temporal precedence, the source code was first transformed to be applicable for the business rules management systems, after which normalization took place. We can ensure that the learning effect was not present during our case. Given the fact that all four subjects who have participated in the experiment, already had executed the business normalization procedure before. Furthermore, the economical beneficial calculation itself was made explicit in Excel and required the respondents only to enter the variables. We cannot exclude learning during the transformation of the case information to the relational representation. Selection, history, maturation, attrition, instrumentation and additive and interactive effects of threats to internal validity are excluded due to the experiment setup.

Outcomes of an experiment can vary when subjects, tasks or the environment changes. External validity is concerned with the extension of variations on such changes (Shadish et al., 2002). Our results were obtained from one decision: a mortgage decision. Therefore, we cannot claim that our conclusions are generally applicable. However, the answer to the research question itself is not influenced by the fact that only one case has been analyzed. Our experiment has been applied outside the project life cycle of AIB. We do not consider this as a threat to environmental validity since the entire procedure can be repeated during normal project life cycles.

VII. CONCLUSION

Business rules are a key denominator for a corporation’s competitiveness. Thereby, the management of such business rules is increasingly becoming more important. However, business rules are becoming more and more complex affecting maintainability and transparency. In order to properly structure business rules, normalization is applied. Normalization increases control over insertion, update and deletion anomalies affecting storage requirements and response time. Currently, the normalization procedure does not take the costs and benefits of normalization into account but is based on intuitive flair. Therefore, we defined the research question: How can business rules guiding decisions be factored such that economic beneficial manageability is realized?

We presented a cost/benefit formula that provides guidelines for normalizing business rules. To determine the normalization business case, three variables were addressed 1) the number of anomalies a business rule set endures, 2) the storage space a business rule set requires, and the 3) deterioration in response time. By means of an experiment based on case study data from an international bank, we have shown the applicability of the model. Results show the importance of properly normalized decisions and what role the
cost and benefit analysis plays in this. On the one hand, modelers should attempt to properly factor business rules. To achieve this factoring, the three normalization forms can be applied. On the other hand, practitioners should take cost and benefits of the organization into account when applying such normalizations forms. Currently, the transformation of the business rules is performed manually. However, in future research we aim to develop an approach which applies an algorithm to re-write (transform) business rules for applying the method presented in this paper. Furthermore, future research should also focus on further validating the method presented in this paper using more cases, and ideally, cases from different industries in various sizes to improve its generalizability.

From a practical perspective, our study provides product engineers, business rules modelers and decision modelers with a method that can be used to normalize business rules based on an economic rationale. This rationale comprises the ideal fit between storage space utilization, anomaly management and execution costs. The method will enable organizations to guard, on the one hand, execution costs and, on the other hand, performance of business rules.

REFERENCES