Comput Econ
DOI 10.1007/s10614-017-9683-7

Multi Criteria Decision Making in Financial Risk Management with a Multi-objective Genetic Algorithm

Sujatha Srinivasan[1] · T. Kamalakannan[1]

Accepted: 12 April 2017

Abstract A huge amount of data is being collected and stored by financial institutions like banks during their operations. These data contain the most important facts about the institutions and its customers. A good and efficient data analytics system can find patterns in this huge data source that can be used in actionable knowledge creation. Actionable knowledge is the knowledge that can be put to decision making and take some positive action towards better performance of organizations. This actionable knowledge is termed Business Intelligence by data scientists. Business Intelligence and Analytics is the process of applying data mining techniques to organizational or corporate data to discover patterns. Business Intelligence and Business Analytics are emerging as important and essential fields both for data scientists and organizations. Risk analysis, fraud detection, customer retention, customer satisfaction analysis and actuarial analysis are some of the areas of application of business intelligence and analytics. Credit risk analysis is an important part of a successful financial institution particularly in the banking sector. The current study takes this risk analysis in financial institutions and reviews the state of the art in using data analytics or data mining techniques for financial risk analysis. The analysis of risk from financial data depends on several factors that are both objective and subjective. Hence it is a multi-criteria decision problem. The study also proposes a multi-objective genetic algorithm (MOGA) for analyzing financial data for risk analysis and prediction. The proposed MOGA is different from other evolutionary systems in that a memory component to hold the rules is added to the system while other systems

✉ Sujatha Srinivasan
ashoksuja03@yahoo.co.in; drsujatha.scs@velsuniv.ac.in

T. Kamalakannan
tkannan722003@gmail.com

1 Department of Computer Applications and Information Technology,
School of Computing Sciences, VELS University, Chennai, Tamil Nadu, India

Published online: 02 May 2017

in the literature are memory less. The algorithm is applied to bench mark data sets for predicting the decision on credit card and credit applications. The preliminary results are encouraging and show light towards better decision making in reducing risks.

Keywords Business analytics · Business intelligence · Big data · Enterprise risk management · Credit risk

1 Introduction

Organizations form the back bone of every country siphoning money out and into the system. Financial institutions are organizations that provide services including lending money to individual persons as well as to huge organizations. Therefore they are directly involved in the financial strength of a country. All types of organization whether product or service have to deal with several risks arising due to a variety of reasons. Financial institutions are prone to more risks and they cannot operate without taking risks. Risk causes a great deal of potential damage and inconvenience for the enterprise stakeholders (Wu and Birge 2016). Thus organizations need to use different strategies to manage or avoid risks. Therefore it becomes important for financial institutions to model risks using historical data in order to gain insight into the risk patterns, so that it falls under their acceptable thresholds.

Financial institutions collect a large amount of data and this data is potentially underutilized. To determine the risk patterns more accurately this data should be integrated within a model. (Katal et al. 2013). Insurance companies regularly extract facts from text gathered by using text analytics to parse the mountains of text that result from the claims process, turn text into structured records, then add that data to the samples studied via data mining or statistical tools for risk, fraud, and actuarial analysis (Russom 2011).

Business analytics is the process of applying data mining techniques to the huge volume of data collected by organizations and the output of business analytics in business intelligence. Business Intelligence is actionable knowledge that can be used in decision making so as to avoid risks or deal with them in a better manner. Thus business analytics has become mandatory for all organizations for dealing with risks. The proposed study is carried out to gain more insight into the risk analysis process, especially in the financial sector. The rest of the paper is organized as follows. Section 2 gives an overview of the types of risks involved in the financial sector. Section 3 reviews the state of art in using Information and Computing Technologies (ICT) and business analytics for risk analysis and management. Section 4 defines the problem to be studied and a multi-objective genetic algorithm for solving risk analysis as a multi-criteria decision problem. Section 5 explains the experiments carried out on bench mark data sets from the UCI machine learning repository to test the proposed algorithm, the results followed by discussion on the observations obtained from the experiments and lists the contributions of the study. Section 6 concludes with a summary and future research directions.

2 Overview of the Types of Risks Involved in Financial Sector

Enterprise risk management (ERM) has emerged as a systematic and integrated approach to efficiently manage risks faced by an organization and it is now emerging as a new discipline in data science and engineering (Wu and Birge 2016). Big data business analytics has the potential to help financial institutions and insurance companies to both describe the past disasters and also predict how such catastrophic disasters can be prevented in the future. It is very critical for the companies to understand and identify fully the impact of operational risks by means of both descriptive and predictive analytics (He 2014).

Wu et al. (2014) provide a review of the state-of-the-art research in business intelligence in risk management, and further propose general classification of risks as Field-based that include financial and non financial risks while Property-based risks are said to have certain properties like uncertainty, dynamics, interconnection and dependence, and complexity. Financial risks include market risk, credit risk, operational risk and liquidity risk. It is observed that the properties of interconnection and dependence and complexity are the properties studied under financial risk assessment.

Beasley et al. (2015) have conducted a study on different perspectives of risk managers and executives on the different kinds of risks involved in organizations. The authors conclude that risk assessment is very important to an organization and it is an on-going process. The authors have used data collected from executives from various levels over a 3 year period to assess different kinds of risks. They have identified 20 different kinds of risk under 3 main categories namely macro-economic risks, strategic risks and operational risks. Out of the 20 risks, customer acquisition and retention by the organization stands in the fourth place. This observation brings out the importance of customers and their behavior in influencing risk patterns to an organization. Credit risk of an organization is greatly influenced by the characteristics or attributes of customers. Hence our study aims at gaining more insight into customer characteristics and their relationship to credit risk.

Iyer et al. (2016) examine heterogeneity in depositor responses to solvency risk using depositor level data in their empirical study. The authors define financial crisis as shock and have reported the following important observations from their empirical study—(i) depositors with loans and bank staff are in a low solvency risk shock, less likely than others to run, but, in a high solvency risk shock, more likely to run, (ii) uninsured depositors are also sensitive to bank solvency and likely to attrite, while (iii) depositors with older or aged accounts run less, and those with frequent past transactions run more, irrespective of the underlying risk. Their results show that the fragility of a bank depends on the composition of its deposit base or rather their type of customers. Empirical studies are carried out using statistical analysis tools which have limited capabilities and can deal only with small data sets. For dealing with big data sets which is the norm of the day, it is necessary to have automated and efficient Information and Computing Technology (ICT) with an efficient data analytics system.

ICT enabled tools can handle large data sets efficiently and can reduce the time required for processing the data. Data analytics techniques further can find useful, meaningful and actionable patterns and knowledge for better decision making which in turn gives organizations an upper hand in risk management.

3 State of Art in Using Business Analytics and ICT for Risk Analysis and Management

Izvercian and Ivascu (2014) list and discuss types of risks associated with sustainable enterprise. They have integrated the most common risk assessment methods of check-lists and instruments of resource mobilization into a platform for risk evaluation. They have developed a Web 2.0 solution called "ONRisk" to assess enterprise risk in various fields and applied their solution to multi dimensional companies from different fields. Their system for risk assessment has the capability to identify, treat, communicate and develop long term strategies for managing risks. As future work the authors wish to add user behavior to include individual factors influencing the risk assessment. As further improvement the authors suggest a semantic risk assessment within business entities as future work. The system proposed in their study is based on data provided by individuals acting as assessor of risks and do not consider customer data collected from previous experiences. Historical data provide better insight into the institutions under various conditions and provide better solutions to handle different risks.

The study by Weng et al. (2016) uses decision trees and logistic regression to model classification rules for the prediction of business intelligence in order to test the effectiveness of BI systems. Data collected from executives and top officials from Taiwanese industries have been analyzed using Decision trees and logistic regression analysis to create classification rules for business intelligence system effectiveness. However their study does not study enterprise risk management.

Enterprises need to apply risk assessment system so as to ensure that correct and timely measures are taken to manage and avoid risks. Fang (2016) have proposed a network marketing performance evaluation index using interval-valued intuitionistic fuzzy sets and generalized mixed weighted aggregation operator to evaluate the risk of an enterprise objectively and effectively. The corresponding strategy to manage the risk is decided by comparing the evaluation value. The author has used empirical analysis to verify the model and have observed that the method for enterprise risk evaluation is feasible and effective.

Data visualization for analytics of risk communication forms the focus of the work by Sarlin (2016). A platform named VisRisk which is comprised of three modules namely, Plots—for interactive interfaces, Maps that provides a three dimensional data cube and Network that deals with data cube dimension and linkages is proposed.

Banks and financial institutions rely on probability of default (PD) model for credit decision making. However Sousa et al. (2016) argue that traditional systems that are one-shot, fixed memory-based, trained from fixed training sets, and static settings are not prepared to process the evolving data. With this context the authors propose a new dynamic modeling framework for credit risk assessment that extends the prevailing credit scoring models built upon historical data static settings. The model has been applied to a real-world financial dataset of credit cards from a financial institution in Brazil. They have observed that their model is able to consistently outperform the static modeling schema.

Enterprises and banks analyze their historic data for credit risk using credit risk rating method. Darwish and Abdelghany (2016) propose a fuzzy logic based model to predict credit risk to analyze the data of Egyptian commercial banks. They have considered attributes of banks like their profitability, debt paying ability, operation ability and liquidity to predict the credit risk.

Öner (2014) has reviewed literature on different Financial models including asset-liability management, credit risk, bank rating, securities trading, risk model comparison, financial early warning system, financial decision making, financial risk management and financial risk measuring. Out of these, machine learning techniques have been used only in two categories namely credit risk and financial early warning system. Moreover the studies found in the literature have taken organizations as a whole and carried out studies on risk management strategies for various risks taken all together. But from most of the studies found in the literature, an important observation is that, the customers form the back-bone of any financial institutions and their behavior and characteristics greatly influence the credit risk of organizations. Hence it is high time that data mining techniques be used to analyze the vast amount of customer data generated by financial institutions to find patterns and gain business intelligence to avoid financial risk in organizations.

Iyer et al. (2016) put forward some questions from their study namely: First whether the characteristics of depositors like education and financial literacy which have been omitted in their study have any influences on the depositor's attrition behavior. Second what will be the response of depositors to different levels of shock (Financial crisis)? Therefore there are many features of customers that directly or indirectly influence customer attrition rates. These questions are the motivation behind the current study where credit risk management in financial sectors like bank and insurance organizations is taken as a multi-criteria decision making problem. To solve this multi-criteria decision making problem, we propose a multi-objective genetic algorithm to find patterns in the form of classification "If-Then" rules discovered from customer data.

Research gap identified from the literature review:

(i) First most of the studies carried out for risk analysis are empirical studies and use statistical tools for analyzing the data. But statistical tools are limited to small data sets and cannot deal with big data efficiently. However data generated by organizations is huge and multi-dimensional. Hence there is a requirement for data analytics system to deal with such big data.

(ii) Secondly customer data is comprised of a number of characteristics like financial status, educational status, loan paying capability and many other attributes. The problem of finding accurate and interesting patterns from this multi-dimensional data is a multi-objective optimization problem. But the systems proposed in the literature take the organization and the different risks faced by them as a whole and the customers only as a small part. The problem of risk assessment considering customer data as a multi-criteria and multi-objective problem is lacking in the literature.

(iii) Thirdly, evolutionary systems deal better with multi-criteria decision problems and multi-objective problems which is lacking in the literature of business analytics for enterprise risk management.

Thus the observations from the literature survey on enterprise risk assessment in general and credit risk from customer data in particular is that: a data analytics system which can deal with big data is the call of the day and hardly any such system is found in the literature.

This is the motivation behind the current study which aims to apply data analytics techniques to financial data, in particular to gain more insight into the patterns of customer characteristics that describe credit risk assessment, taking it as a multi-objective problem.

4 The Problem and a Multi-objective Genetic Algorithm for Risk Analysis

Credit risk is defined as the risk that borrowers will fail to pay its loan obligations. (Abdelmoula 2015). Credit risk analysis is a multi-criteria decision problem since the decision taken depends on various factors that define a customer. Wu et al. (2014), have observed that key approaches in financial analysis are computational intelligence, evolutionary computation and optimization approaches. Evolutionary systems deal better with multi-criteria problems where the multiple criteria are represented as vectors in the problem space. Data analytics techniques are used to discover patterns from these vectors and these patterns define a solution space. These patterns are used to model actionable knowledge that is used in decision making. The process of using data analytics techniques to mine actionable knowledge from business data is known as Business analytics and the output of such a system is Business Intelligence.

Problem definition

> Given a data source of financial customer data, the problem is applying business analytics techniques in a multi-objective evolutionary computational environment for discovering patterns or business intelligence that models credit risk and use these credit risk models to enable better decision making to avoid or better manage future financial risks.

4.1 Multi-objective Genetic Algorithm for Risk Analysis

Evolutionary computation applies natural and biologically inspired evolution techniques like selection, crossover and/or mutation, and works on the strategy of survival of the fittest. The potential solutions to a given problem form the population space and a fitness function determines their survival in the environment namely the solution space. Evolutionary techniques can deal with large and multi-dimensional population spaces in an efficient manner. GAs work with a population of points, and provides a set of Pareto-optimal solutions making it a very powerful tool especially for Multi-objective optimization problems (MOOP) (Govindan et al. 2016**)**. The current study

Table 1 Multi-objective genetic algorithm for risk analysis

```
Algorithm MOGA
Input: Financial data Source, Evaluation Metrics, Parameters for MOGA,
termination condition
Output  : Financial risk models as "If-Then" rules, metrics and accuracy
statistics
Start
Step 1  Convert data to Chromosome
Step 2  Create initial Population
Step 3  Evaluate individuals in initial population using metrics
        and select best individuals for next generation
Step 4  While termination condition = false
           Select individuals for reproduction
           Reproduce
           Update Population

Step 4.1 For each individual in Population
             Evaluate individuals in Population
             Select individuals based on evaluation metrics and update
Population
         End for
        End while
Step 5.  Evaluate the classifier on test set
Step 6.  Output rules, metrics and accuracy statistics
Stop.
```

proposes a multi-objective genetic algorithm for discovering classification "If-Then" rules from financial data. The steps in the algorithm are given in Table 1.

4.2 Methodology

The multi-objective genetic algorithm works as follows. The data set is divided into train data (2/3 rd of data) and test data (1/3 rd of data). The data source and parameters for the genetic algorithm namely crossover rate, mutation rate, population size, number of generations and the stopping criteria, the metrics for optimization namely threshold values for the evaluation metrics of confidence and coverage are given as input. The algorithm takes the train data set and converts them to chromosome by converting the records of the data set into vectors. The algorithm then chooses a set of initial seeds from the chromosome set at random and produces the initial population by applying the reproduction operators of cross-over, mutation and selection. The selection is based on the threshold values for the rule metrics. Coverage and confidence have been taken as the metrics in the proposed study. The best rules above the threshold values of the evaluation metrics are selected and go to the next generation while other rules are stored in a rule base for further exploration. This is in contrast to the other proposed algorithm found in the literature where the rules which do not qualify are dropped and deleted from the system. The rule base is maintained until the end of the algorithm providing memory to the otherwise memory-less evolutionary systems. However only the best rules go to the next generation and the algorithm may choose the rules both from the rule base (10% probability) as well as the best rules for reproduction (90%

probability). The number of generations is taken as the stopping criteria. When the specified number of generations is reached, the algorithm tests these rules on the test data. Then the rules, the values of rule metrics for that rule and the number of test data instances that have been misclassified are given as output from which the classification accuracy is calculated. The system has been developed in Java. Table 1 gives the algorithm of the proposed multi-objective genetic algorithm with memory for risk analysis.

4.3 Objectives of the Study

The objectives of the study are listed below:

(i) To take risk analysis and management as a multi-criteria decision making problem and analyze the customer data to discover patterns and relationships among the features that acts as deciding factors in customer attrition decisions.
(ii) To apply multi-objective evolutionary systems that deal with multi-objective problems better to the problem of risk analysis and management for discovering Business Intelligence from the huge volume of data collected from financial institutions.
(iii) To study the data collected by financial institutions to gain more insight into the customer behavior to better deal with and avoid risks in the future.

4.4 Research Questions

The study aims at answering the following questions:
Research Question 1: Can the proposed data analytics algorithm using evolutionary computation technique be able to discover useful, meaningful and actionable patterns and relationships from the data?
Research Question 2: Can we model and predict the risk beforehand from the data collected so that decisions can be taken to avoid such risks before hand?
Research Question 3: What are the features of a customer that decides the behavior of customers for taking decision on their credit card or credit application?

5 Experiments, Results and Discussion

5.1 Experiments

Experiments were carried out on two bench mark data sets from the UCI machine learning repository namely credit card approval (CRX) data and Australian Credit approval data sets (Bache and Lichman 2013) to test the proposed algorithm. These data sets concerns credit card and credit applications respectively. All attribute names and values have been changed to meaningless symbols to protect confidentiality of the data. The algorithm was applied to the data sets ten times each and the classification accuracy measures of confidence and coverage have been observed. Tables 2 and 3 give the data set information for Credit card approval (CRX) data and Australian Credit

Table 2 Data set and attribute information—credit card approval (CRX)

Attribute	Values of attributes and algorithm representation
A1	b, a. (1, 2)
A2	Continuous. (−, 38.96):1, (38.96, −):2
A3	Continuous. (−, 4.20):1, (4.20, −):2
A4	u, y, l, t.—1, 2, 3, 4
A5	g, p, gg—1, 2, 3
A6	c, d, cc, i, j, k, m, r, q, w, x, e, aa, ff—1–14
A7	v, h, bb, j, n, z, dd, ff, o—1–9
A8	Continuous (−, 1.27):1, (1.27, −):2
A9	t, f—1, 2
A10	t, f—1, 2
A11:	Continuous (−, 5):1, (5, −):2
A12:	t, f—1, 2
A13:	g, p, s—1, 2, 3
A14:	Continuous (−, 105):1, (105, 2889):2
A15:	Continuous (−, 492):1, (492, −):2
A16: +, − (class attribute)	1, 2
Data set information	
Number of instances	653 (After removal of records with missing attributes)
Number of attributes	15 + class attribute
Class distribution	
1	300
2	366

approval data sets respectively. There are certain parameters that are to be input for the multi-objective genetic algorithm. Table 4 summarizes these values.

5.2 Results

A classifier system is evaluated using its ability to classify unknown test data instances. The classification accuracy of the classifier returned by any algorithm is calculated as the percentage of test data instances correctly predicted by the returned classifier. Different sets of experiments each consisting of ten runs of the algorithm have been conducted for each data set using different combinations of metrics. Table 5 shows the summary of the ten runs along with minimum, maximum, average and standard deviation of accuracies obtained over ten runs of the algorithm for CRX data set. Coverage and confidence that are accuracy metrics have been taken as the rule evaluation metrics for choosing individuals to the next generation in the current study. Also the accuracy of the algorithm in classifying the test data is presented in Table 5 for credit card approval data set and Table 6 gives a couple of sample rules. Table 7 summarizes the results of the set of experiments run on Australian credit approval data set.

Table 3 Data set and attribute information (Australian credit approval)

Attribute	Values of attributes and algorithm representation
A1	a, b—0, 1
A2	Continuous
A3	Continuous
A4	p, g, gg—1, 2, 3
A5	ff, d, i, k, j, aa, m, c, w, e, q, r, cc, x—1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14
A6	ff, dd, j, bb, v, n, o, h, z—1, 2, 3, 4, 5, 6, 7, 8, 9
A7	Continuous
A8	t, f—1, 0
A9	t, f—1, 0
A10	Continuous
A11	t, f—1, 0
A12	s, g, p—1, 2, 3
A13	Continuous
A14	Continuous
A15	+,− − 1, 2 (class attribute)
Number of instances	690
Number of attributes	14 + class attribute
Class distribution	
+	307 (44.5%) CLASS 2
−	383 (55.5%) CLASS 1

Table 4 Parameter settings for the multi-objective genetic algorithm for mining rules

Parameters	Values
Population size	100
Number of generations	25
Crossover rate	80%
Mutation rate	20%
Type of crossover	Uniform multi-point crossover
Threshold values to choose best rules	0.80 (both for coverage and confidence)

5.3 Discussion

The results of the experiments conducted are presented in Table 5 for credit card application (CRX) data set and a set of sample rules is presented in Table 6. From the results on CRX data it is observed that the algorithm is able to produce on an average of 414 unique rules with a maximum of 449 and minimum of 396 rules. The accuracy of the classifier on test data ranges from 60.55 to 77.06%, with an average accuracy of 67.77%. The number of best rules above the specified threshold values for the chosen metrics of coverage and confidence of 0.80 ranges from a minimum of 29 rules to a maximum of 300 rules and with an average of 78 rules. As for the

Table 5 Results of the experiments (over 10 runs) for credit card approval (CRX) data set

Iteration no.	No. of unique rules	Total rules in the classifier	Time (sec)	Accuracy%
1	399	29	3.869	66
2	411	67	16.923	77.06 (Best value)
3	449	300	50.598	60.55
4	417	58	5.062	72.48
5	396	40	9.670	66.97
6	410	52	4.820	61.47
7	418	66	9.639	61.93
8	418	83	6.099	70.18
9	406	57	10.556	66.51
10	418	72	7.026	74.56
Avg	414.2	82.4	12.426	67.77
SD	14.60	78.02	13.947	5.69
Min	396	29	3.869	60.55
Max	449	300	50.598	77.06

Australian credit approval data set from Table 7, the maximum obtained accuracy is 76.52%, with an average of 70.65% on the test data set. On an average the algorithm is able to create 401.6 unique rules and out of these 40.5 rules are observed to be above the specified threshold value for the chosen metrics. The observations suggest that the algorithm is able to produce a large number of unique rules and also best rules above the specified threshold values. However it is necessary to reduce the number of best rules to reduce the complexity of the solution set and increase the understandability of the classifier. The time taken by the algorithm to produce and select the set of rules ranges from 3.869 to 50.598 with an average of 12.426 for the CRX data set with 653 instances and 16 attributes while it ranges from 11.31 to 21.68 with an average of 17.23 for the Australian credit approval data set which has 690 instances and 15 attributes. That is the algorithm is able to create and select unique and best rules in less than a minute for both the data sets. However the accuracy of the set of rules on test data seems to be less compared to leading algorithms in the literature. This may be due to the fact that chosen threshold value might be too low or too high, or the algorithm needs to be fine tuned for different parameters of the evolutionary system. Kliegr et al. (2014) state that “The confidence threshold can be used to control the quality of the resulting classifier”. Moreover association rules that have high confidence are known as “hard” association rules and may not be interesting to the user and will have strong correlation between the data instances. (Dimitrijevic and Bosnjak 2010). The high threshold values for the rule indices of confidence and coverage might have resulted in hard core rules that might have high classification accuracy on a particular class and low on the other class. This has been observed in imbalanced data sets. Thus choosing the threshold value for the rule indices plays a vital role in deciding the resulting classifier and its classification performance. Thus more experiments need to be done in choosing better threshold values for rule indices. The occurrence of a

Table 6 Sample rules and metric values output by the algorithm

Rules	Confidence	Coverage
(IF ((A1 IS 1) AND (A2 IS 1) AND (A3 IS 1) AND (A4 IS 1) AND (A5 IS 1) AND (A6 IS 10) AND (A7 IS 1) AND (A8 IS 1) AND (A9 IS 1) AND (A10 IS 1) AND (A11 IS 2) AND (A12 IS 1) AND (A13 IS 1) AND (A14 IS 2) AND (A15 IS 1) THEN Class IS 1	0.9444444444444444	0.9444444444444444
(IF ((A1 IS 1) AND (A2 IS 1) AND (A3 IS 1) AND (A4 IS 1) AND (A5 IS 1) AND (A6 IS 2) AND (A7 IS 1) AND (A8 IS 1) AND (A9 IS 1) AND (A10 IS 2) AND (A11 IS 1) AND (A12 IS 1) AND (A13 IS 1) AND (A14 IS 1) AND (A15 IS 1) THEN Class IS 2	0.9016949152542372	0.9971910112359551

Table 7 Results of the experiments (over 10 runs) for Australian credit approval data set

Iteration no.	No. of unique rules	Total rules in the classifier	Time (s)	Accuracy%
1	387	45	12.43	74.35
2	389	50	16.78	65.65
3	408	42	17.67	68.26
4	406	62	18.69	76.08
5	377	40	11.31	70.86
6	394	27	19.55	69.13
7	397	21	17.91	76.52 (Best value)
8	411	71	21.68	69.13
9	434	25	15.11	66.52
10	413	22	21.20	70.00
Avg	401.6	40.5	17.23	70.65
Stdev	16.28	17.16	34.39	3.81
Min	377	21	11.31	65.65
Max	434	71	21.68	76.52

high standard deviation in all the observations suggests that the consistency of the algorithm also needs fine tuning. Therefore more experiments needs to be conducted to do sensitivity analysis for fine tuning the different parameters of the algorithm and for calibration of the algorithm using different metrics for rule evaluation and different threshold values. Moreover as stated in (Nath et al. 2013) “Single objective function (i.e. only frequency of occurrence) based rule generation cannot generate frequent as well as rare rules simultaneously”. Therefore a better optimization strategy like Pareto optimal optimization which considers all the values of metrics together and compares them as a whole for choosing better rules needs to be incorporated. Use of a better optimization strategy and other rule indices might tend to reduce the number of rules in the classifier and decrease complexity of the rule set. The type of crossover which is used in exploitation of the solution space and mutation which is used for exploration of the solution space for finding new solutions are also features of evolutionary systems that influence the outcome which also needs to be explored. These results of fine tuning the algorithm we propose to report in the future.

5.4 Contributions of the Study

(i) A novel multi-objective genetic algorithm is proposed for mining business intelligence considering risk analysis as a multi-criteria decision making problem and presenting users the mined knowledge in the form of user friendly “If-Then” rules.

(ii) The proposed genetic algorithm extends other evolutionary algorithms found in the literature by adding a memory component to hold the intermediate rules produced unlike the memory less algorithms proposed so far in the literature.

(iii) The algorithm has been tested on bench mark data sets and the insights obtained through the experiments and results have been discussed throwing light in the direction for future research.

6 Conclusion and Future Work

The study undertaken has considered risk assessment in financial institutions as a multi-criteria decision making problem. Since evolutionary systems are best at dealing with multi-criteria and multi-objective problems, a genetic algorithm is proposed and applied on credit card application and Australian credit approval data sets which are bench mark data sets from the UCI machine learning repository concerning credit risks. The results are encouraging in that the algorithm discovers valid patterns from the data. Moreover the algorithm is able to create a large set of unique rules and best rules. However, the algorithm uses threshold values for the evaluation metrics to choose rules for the next generation rather than an optimization strategy like Pareto optimization. Therefore as future work we propose to use a Pareto optimization strategy to improve both the convergence of the solution and to improve the classification accuracy and decrease the complexity of the classifier. The study also proposes to use other rule evaluation metrics like interestingness of the rules to observe their influence on the convergence of solutions and accuracy.

Further the proposed algorithm is tested on bench mark data sets available in bench marking repositories. The financial data available for experimentation is such that the real information is hidden using discretization techniques to preserve privacy of customer data. Therefore it is not possible to infer meaningful patterns to be applied in decision making in the real world. Hence initially it is proposed to use these available data sets for validating and calibrating the proposed system and then applying it to a real data set to test if it discovers interesting and actionable business intelligence. Hence this is also proposed as future research challenge.

The data set used in the current study although multi-dimensional cannot fit into big data category. However this is justified from the observation from the review of customer churn prediction models by Verbeke et al. (2011). The authors have reviewed the literature on customer churn prediction models using rule induction. An important observation from the review is that, most data sets concerning customers are private and are not publicly available for research. In particular 14 out of the 18 articles taken by them for review deals with private data. This shows the difficulty in obtaining business data even data sets in which private data are decoded or hidden using some form of discretization technique. Hence application of the proposed algorithm on real world big data is also in the agenda for future direction of the study. For this purpose, a questionnaire to collect details of customers of a private bank is being developed with the help of banking professionals and insurance professionals using their expertise. The improved version of the proposed algorithm will be tested on the collected data set and will be reported.

References

Abdelmoula, A. K. (2015). Bank credit risk analysis with k-nearest-neighbor classifier: Case of Tunisian banks. *Accounting and Mangement Information System*, *14*, 79–106.
Bache, K., & Lichman, M. (2013). UCI machine learning repository. Irvine: University of California, School of Information. http://archive.ics.uci.edu/ml/.
Beasley, M., Branson, B., & Pagach, D. (2015). An examination of the assessment of top risks on the horizon: Evidence from executives and risk professionals. *The Journal of Enterprise Risk Management*, *1*, 1–22.
Darwish, N. R., & Abdelghany, A. S. (2016). A Fuzzy logic model for credit risk rating of Egyptian commercial banks. *International Journal of Computer Science and Information Security*, *14*, 11–19.
Dimitrijevic, M., & Bosnjak, Z. (2010). Discovering interesting association rules in the web log usage data. *Interdisciplinary Journal of Information, Knowledge and Management*, *5*, 191.
Fang, J. (2016). Integrating market and credit risk: A simulation and optimisation perspective. *International Journal of Security and Its Applications*, *10*, 21–32. doi:10.14257/ijsia.2016.10.4.03.
Govindan, K., Jafarian, A., Azbari, M. E., & Choi, T. M. (2016). Optimal Bi-objective redundancy allocation for systems reliability and risk management. *IEEE Transactions on Cybernetics*, *46*, 1735–1748. doi:10.1109/TCYB.2014.2382666.
He, X. J. (2014). Business intelligence and big data analytics: An overview. *Communications of the IIMA*, *14*, 1–11.
Iyer, R., Puri, M., & Ryan, N. (2016). A tale of two runs: Depositor responses to bank solvency risk. *The Journal of Finance*, *71*, 2687–2726. doi:10.1111/jofi.12424.
Izvercian, M., & Ivascu, L. (2014). Semantic assessment of the risk: A multidimensional data model for sustainable enterprise. *Procedia Technology*, *16*, 598–605. doi:10.1016/j.protcy.2014.10.008.
Katal, A., Wazid, M., & Goudar, R. H. (2013). Big data: Issues, challenges, tools and good practices. In *2013 6th international conference on contemporary computing, IC3 2013*. pp. 404–409. doi:10.1109/IC3.2013.6612229.
Kliegr, T., Kuchař, J., Sottara, D., & Vojíř, S. (2014). *Learning business rules with association rule classifiers*. Berlin: Springer International Publishing. doi:10.1007/978-3-319-09870-8_18.
Nath, B., Bhattacharyya, D. K., & Ghosh, A. (2013). Incremental association rule mining: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *3*, 157–169. doi:10.1002/widm.1086.
Öner, S. C. (2014). Reformative financial risk management approach: A multistage decision support system with the assistance of Fuzzy goal programming and expertons method. *Journal of Mathematics and System Science*, *4*, 620–636.
Russom, P. (2011). Big data analytics. *TDWI Best Practices Report*, pp. 1–34. doi:10.1109/ICCICT.2012.6398180.
Sarlin, P. (2016). Macroprudential oversight, risk communication and visualization. *Journal of Financial Stability*, *27*, 160–179. doi:10.1016/j.jfs.2015.12.005.
Sousa, M., Gama, J., & Brandão, E. (2016). A new dynamic modeling framework for credit risk assessment. *Expert Systems with Applications*, *45*, 341–351. doi:10.1016/j.eswa.2015.09.055.
Verbeke, W., Martens, D., Mues, C., & Baesens, B. (2011). Building comprehensible customer churn prediction models with advanced rule induction techniques. *Expert Systems with Applications*, *38*, 2354–2364. doi:10.1016/j.eswa.2010.08.023.
Weng, S.-S., Yang, M.-H., Koo, T.-L., & Hsiao, P.-I. (2016). Modeling the prediction of business intelligence system effectiveness. *Springerplus*, *5*, 737. doi:10.1186/s40064-016-2525-6.
Wu, D., & Birge, J. R. (2016). Risk intelligence in big data era: A review and introduction to special issue. *IEEE Transactions on Cybernetics*, *46*, 1718–1720. doi:10.1109/TCYB.2016.2580239.
Wu, D. D., Chen, S. H., & Olson, D. L. (2014). Business intelligence in risk management: Some recent progresses. *Information Sciences (Ny)*, *256*, 1–7. doi:10.1016/j.ins.2013.10.008.