



VICTORIA UNIVERSITY
MELBOURNE AUSTRALIA

Dynamic optimisation based fuzzy association rule mining method

This is the Accepted version of the following publication

Zheng, Hui, He, Jing, Huang, G, Zhang, Yanchun and Wang, Hua (2019)
Dynamic optimisation based fuzzy association rule mining method.
International Journal of Machine Learning and Cybernetics, 10 (8). pp. 2187-2198. ISSN 1868-8071

The publisher's official version can be found at
<https://link.springer.com/article/10.1007%2Fs13042-018-0806-9>
Note that access to this version may require subscription.

Downloaded from VU Research Repository <https://vuir.vu.edu.au/39353/>

Dynamic optimisation Based Fuzzy Association Rule Mining Method

Hui ZHENG · Jing HE · Guangyan HUANG ·
Yanchun ZHANG · Hua WANG

Received: date / Accepted: date

1 **Abstract** Techniques of performance analysis, comprising of various metrics such as
2 accuracy, efficiency and consuming time, have been conducted to evaluate the mea-
3 sures of properties and interestingness for the association rule mining method. There-
4 fore, these metrics combined with different parameters (partitioning points, fuzzy
5 sets) should be analysed thoroughly and balanced simultaneously to enhance the en-
6 tire performance (effectiveness, accuracy and efficiency) for an algorithm. As a result,
7 Most of the current algorithms face the pressure from the tradeoff of these metrics and
8 parameters, which becomes even rougher when we employ it in different resources
9 of data (discrete data, categorical data and continuous data). Specifically, serial data
10 (i.e., sequences or transactions of floating point numbers), such as analysis of sen-
11 sor streaming data, financial streaming data, medical streaming data and sentimental
12 streaming data, are different from discrete variables, such as boolean data (e.g., sen-
13 timent: negative and positive represented as ‘0’ and ‘1’ separately) and categorical

Hui ZHENG
CAS Research Center on Fictitious Economy and Data Science,
University of Chinese Academy of Sciences, Beijing, 100190, China
Victoria University, Melbourne, Australia
and Fudan University, Shanghai, China
E-mail: hui.zheng2@live.vu.edu.au

Jing HE (corresponding author)
Institute of Information Technology, Nanjing University of Finance and Economics, College of Engineer-
ing and Science, Victoria University, Melbourne, Australia

Guangyan GUANG
School of Information Technology, Deakin University, Melbourne, Australia

Yanchun ZHANG
Centre for Applied Informatics, Victoria University, Melbourne, Australia
and Fudan University, Shanghai, China

Hua WANG
Centre for Applied Informatics, Victoria University, Melbourne, Australia

14 data (e.g., ‘young age’, ‘middle age’, ‘old age’). The main difference is that serial
15 data face sharp boundary’s problem. That is, it is hard to decide the boundary values
16 (i.e., the single points to partition data into different value groups), which is few to
17 be solved in association rule mining methods. This paper aims to resolve the problem
18 of sharp boundaries and balance multiple performances of our algorithm simultane-
19 ously by developing a novel dynamic optimisation (parameters and metrics) based
20 fuzzy association rule mining (DOFARM) method. The proposed method can be ap-
21 plied in a wide range of classifying problems, such as the classification of sentiment
22 strength (negative and positive). In our DOFARM method, instead of single partition-
23 ing points, we use a range of values to smoothly separate two consecutive partitions
24 and develop a corresponding membership function to generate fuzzy sets for original
25 data sets of physical and emotional diseases. Mainly, we design a dual compromise
26 scheme: the first tradeoff balances better performance of out-putting association rules
27 and more widely applicable fuzzy membership function while the second tradeoff re-
28 duces the time parameter as well as enhances the entire performance of our DOFARM
29 method. The feasibility and accuracy of DOFARM we proposed have been certified
30 theoretically and experimentally. Besides, we demonstrate the accuracy, effectiveness
31 and efficiency for our DOFARM method by experiments according to both synthesis
32 and real datasets.

33 **Keywords** Association Rule · Optimised Parameters · Multiple Objective Function ·
34 Data Mining

35 1 Introduction

36 Efficient analysis of serial data (i.e., sequences or transactions of floating point num-
37 bers) has become a crucial issue to be successfully resolved with the advancement
38 of computing technology, such as data streams in financial, medical applications and
39 physiological factors acquisition. Traditional classifiers can manage serial data and
40 classify them into different groups conveniently. However, the hidden relationships
41 in original data are also required to mine to provide further information, e.g., the
42 possible product in a shopping process or the potential reason of type 2 diabetes. As-
43 sociation rule mining is therefore generally chosen for mining hidden relations and
44 associations. The problem of the association rule mining method is that it concerns
45 only non-continuous factors such as categorical sequence objects and customer trans-
46 action records and cannot handle continuous data quickly.

47 Suppose we have a constant feature: ‘Age’, a direct method is to divide this fea-
48 ture into intervals . When the number of intervals is fixed as three, we can use labels:
49 ‘young age’, ‘middle age’ and ‘old age’ as the feature classes (crisp sets) after choos-
50 ing the partitioning points. While, by using fuzzy theory [1] for the feature of ‘Age’,
51 we can combine the three segments with membership functions by extending the
52 boolean values 0 and 1 (respectively indicating absence and presence) to the contin-
53 uous values from 0 to 1 ($[0, 1]$). Thus, the crisp transactions have been changed into
54 fuzzy ones as shown in Figure 1. Specifically, crisp sets can only define whether a
55 tuple contains an item, while in the fuzzy sets, we can define the degree of a tuple
56 belonging to each interval. Still taking feature ‘Age’ as an example, we can generate

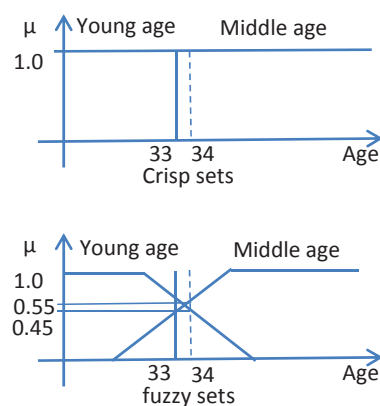


Fig. 1. An example of crisp sets and fuzzy sets.

57 three intervals $(0, 33]$, $(33, 67]$ and $(67, +\infty)$ with three classes ‘young age’, ‘middle
 58 age’ and ‘old age’ respectively. But it is non-reasonable to classify a person with 33
 59 years old into the ‘young age’ class, an individual with 34 years old into the ‘middle
 60 age’ class. This problem is called sharp boundaries. In comparison, the fuzzy sets in
 61 Figure 1 can regard the ‘Age’: 34 as (young age, 0.45) (middle age, 0.55) instead of
 62 (young age, 0) (middle age, 1) in the crisp sets. Also, the feature of ‘Age’ can be ap-
 63 plied for various purposes. Sometimes, we use ‘Age’ to judge the personal incomes;
 64 then we would like to change the partitioning points according to the modern work-
 65 ing age and the modern retirement age. Sometimes, we distinguish ‘Age’ to measure
 66 the risk of heart disease or sentiment strength, in this situation, the feature ‘Age’ is
 67 better to be partitioned by using the changes of ‘Age’ rather than the absolute value.
 68 The characteristics of one person vary from gender, district, and country and all fea-
 69 tures may evolve. So all of these parameters, such as the partitioning points and fuzzy
 70 sets need to be improved and balanced simultaneously. When the continuous data are
 71 involved, it is not an easy task to extend the approaches introduced above.

72 As mentioned in paper [2], fuzzy logic was applied first to extend the association
 73 rule mining method with fuzzy sets of range, which keeps the advantage of numeric
 74 data with a membership value and diminishes the problem of the sharp or abnormal
 75 boundary in dividing the interval. Besides, a general model to discover association
 76 rules is proposed in work [3], which consists of the user-defined filter of certainty fac-
 77 tors and the definition of very strong rules to generate interesting association rules.
 78 Apart from the paper [3], researchers have already presented some approaches to
 79 improve the method of fuzzy association rule mining. An assessment method to par-
 80 tition the data into different groups according to the features of data that are related
 81 to a given rule, that against the rule (the counterexamples) and that are irrelevant with
 82 the rule is developed in [4]. Another work in [5] introduces the novel measurements
 83 of quality by distinguishing the correlations of positive from the correlations of neg-
 84 ative association rules; while extra measures (clustering, classifying, weighting and

85 extracting membership function) are used to modify fuzzy association rule models
86 [6], [7], [8] and [9].

87 Paper [10] proposes a classifying model called TME to distinguish social emo-
88 tions of readers. Also, the generated topic indicators are utilised for the alleviation
89 of overfitting problems. Additionally, the framework SenticRank of paper [11] aims
90 to rank content-based sentiment and collaborative sentiment. Compared to it, this pa-
91 per applies rule-based sentiment to further reveal to relationships between sentiments
92 and individual features. As mentioned in papers [10], [12] and [13], most emotional
93 or sentimental classification problems are solved by text mining. This paper will ap-
94 ply fuzzy association rule to partition sentiments into positive and negative, which
95 means emotions benefit for well-being or harm to well-being.

96 As mentioned above, the method of fuzzy association rule discovering is not per-
97 formed without its downside. The problems contain lacking the tradeoff scheme to
98 select the most suitable partitioning points for association rules generating (while
99 the continuous original data sets are transformed into fuzzy sets and corresponding
100 membership values, the partitioning points are chosen as the points between any two
101 adjacent fuzzy sets). The procedures of selecting partitioning points and calculating
102 their membership values [3] and [4] are two essential processes of constructing fuzzy
103 sets in building Fuzzy Association Rule Mining (abbreviation of FARM) model steps.
104 Suppose $(0, 33]$, $(33, 67]$ and $(67, +\infty)$ of the feature 'Age' are three fuzzy sets with
105 two partitioning points 33 and 67, which is not an accurate definition of middle age
106 and it is inconsistent with general knowledge 44 and 59 in [14] or 40 and 60 in [15].
107 As the definition of middle age varies from domain, application and time - the par-
108 ticular algorithm is required to adjusting partitioning points regarding the accuracy
109 improvement of distinguishing diabetes. However, there is not the task, on which re-
110 searchers focus. What's more, any individual with 'Age' of 80 pay more attention
111 with their emotional and physical well-being than a person with only 40 years old,
112 but more and more kinds of diseases such as heart disease, diabetes and emotional
113 disease are hitting on middle-aged people. So the relations with illness for the middle
114 age (non-high value of feature) are ordinarily more useful than that related to old age
115 (high value of feature). Also, the more related features (e.g., Age, systolic pressure,
116 diastolic pressure, blood glucose) we consider, the more accurate result we can get.
117 To be more specific, a person of middle age combined with other feature, i.e., systolic
118 pressure of 130 mm HD, which is diagnosed as one of the two criteria of per hyper-
119 tension (non-high value of element). In that case, a slight high in blood glucose will
120 sharply increase the possibility of having diabetes problem than an old aged person
121 with only a high value of 'Age'. Beyond this, the metrics of association-rules filtering
122 and the parameters for the membership function smoothing still need to be improved
123 and balanced simultaneously.

124 As these restraints of current FARM method, a generic method: Dynamic optimi-
125 sation (parameters) based Fuzzy Association Rule Mining (DOFARM) is proposed,
126 working with both continuous data and discrete data. It firstly offers a dual compro-
127 mise scheme to balance the accuracy, effectiveness and efficiency of our algorithm
128 simultaneously; Besides, the DOFARM method we proposed smoothes membership
129 function of fuzzy sets and consequently reduces sharp boundary problems to a great
130 extent. Moreover, our novel method which is based on the parameter selecting en-

hances the entire performances of fuzzy association rule mining by optimising parameters (partitioning points, fuzzy sets, the number of association rules) and metrics (support, confidence, certainty factor [3]). Fourthly, the efficiency - the most critical part of a method is improved almost two times by our DOFARM method when it skips some unnecessary steps with direction parameter selecting and reduces time-consuming of our DOFARM method. Therefore, combining these contributions of our DOFARM, we can say that it can finally achieve two interacting tradeoffs. To be more specific, it balances the effectiveness and accuracy (parameters: partitioning points and fuzzy sets) with multiple objective function scheme of the first tradeoff. Also, it adjusts the smoothly cognitive membership function and better performance of association rule mining (parameters and metrics of association rules) together, which is called dual compromise in this paper.

The rest of this paper is organised as follows. Section 2 describes the first tradeoff of our DOFARM method, which optimises three user-defined metrics to balance the accuracy and effectiveness of our algorithm simultaneously. In section 3, detailed algorithms of our second tradeoff are proposed through the interval partitioning, membership function constructing, parameters based metrics balancing and dual compromise mechanism proving. To further illustrate the DOFARM method and the dual compromise scheme we proposed, the computing processes and procedures are represented in section 4. The theorem, which demonstrates the universal applied dual compromise scheme is also illustrated in this chapter. The experimental performance evaluation of accuracy, effectiveness and efficiency of the proposed DOFARM method is studied in Section 5. Finally, conclusions are summarised in Section 6.

2 The First tradeoff: Balancing Different Metrics of Fuzzy Association Rules Simultaneously for Better Performance

Distinct from the classic fuzzy association rule mining method, our DOFARM method optimises frequent itemsets and association rules according to two additional tradeoff processings. It optimises the association rules of mining-output based on the previous frequent itemsets and the parameters based on selected metrics, which are used as metrics for optimisation functions. As a consequence, we can balance the effectiveness and accuracy using the proposed method (the second tradeoff). However, before that process, we should attempt to optimise the performance of our dynamic optimisation based fuzzy association rule mining, that is, balancing all of the metrics of fuzzy association rules: better-performed results and more interesting association rules (the first tradeoff).

2.1 A Multi-objective optimisation Scheme

In this subsection, we will introduce our scheme to optimise metrics of fuzzy association rules simultaneously which is based on Richardson Extrapolation and Gradient-based optimisation methods [16], [17], [18], [19], [20], [21]. A theorem is illustrated to indicate the correctness of our multi-objective optimisation process. Among all of the processes, the effective metrics should be listed and applied in our scheme firstly.

172 With the definition of multiple objectives, the problem that we are facing becomes
 173 optimising our metrics based objectives $\varphi_1, \varphi_3, \varphi_5, \varphi_{10}, \varphi_{n/2}$ simultaneously, by the
 174 procedure of selecting partitioning points according to the result of the direction from
 Algorithm 1.

Algorithm 1 Direction-computation Algorithm

Input: three thresholds: min_Supp, min_Conf and (min_CF; the user-defined number of objective function: N_objective (default value as 5); the total outputting number of association rules (n) a initialising set of partitioning points: \mathcal{X}_0 ; and the initialising gradients for every objective functions at the point of \mathcal{X}_0 :

$$\mathbf{g}_1 = \nabla\varphi_1, \mathbf{g}_2 = \nabla\varphi_3, \mathbf{g}_3 = \nabla\varphi_5,$$

$$\mathbf{g}_4 = \nabla\varphi_{10}, \mathbf{g}_5 = \nabla\varphi_{n/2},$$

Output: the chosen direction $\boldsymbol{\eta}$ (according to that direction all of the five objective functions can keep the condition of increasing in the limited area of the neighbourhood of the current point of \mathbf{x}).

```

1: At first:  $\boldsymbol{\eta} \leftarrow \{0, \dots, 0\}$ .
2: for  $i = 1, \dots, N\_objective$  do
3:    $\boldsymbol{\alpha} \leftarrow \mathbf{g}_i$ ;
4:   for  $j = 1, \dots, N\_objective$  do
5:     if  $j \neq i$  and  $\langle \boldsymbol{\alpha}, \mathbf{g}_j \rangle < 0$  then
6:        $\boldsymbol{\alpha} \leftarrow \boldsymbol{\alpha} - \frac{\langle \boldsymbol{\alpha}, \mathbf{g}_j \rangle}{\langle \mathbf{g}_j, \mathbf{g}_j \rangle} \mathbf{g}_j$ ;
7:     end if
8:   end for
9:    $\boldsymbol{\eta} \leftarrow \boldsymbol{\eta} + \boldsymbol{\alpha}$ ;
10: end for
11: Return  $\boldsymbol{\eta}$ ;
```

175

176 **Theorem 1** In every optimisation step, \exists the direction $\boldsymbol{\eta}$ makes all of objective func-
 177 tions be optimised simultaneously.

Proof Using the idea of Algorithm 1, the $\boldsymbol{\eta}$ can be calculated as

$$\boldsymbol{\eta} = \sum_{i=1}^5 \sum_{j=1}^5 \left(\mathbf{g}_i - \frac{\langle \mathbf{g}_i, \mathbf{g}_j \rangle}{\langle \mathbf{g}_j, \mathbf{g}_j \rangle} \mathbf{g}_j \right),$$

178 where $j \neq i$ and $\langle \mathbf{g}_i, \mathbf{g}_j \rangle < 0$.

179 With the above algorithm, we can sum up the Theorem 1. The parameter sets of
 180 partitioning points for the selected association rules are gradually optimised by the
 181 optimisation objective Algorithm 1. In the next section, we will consider about the
 182 optimisation procedure of the membership function and the corresponding fuzzy sets.
 183 Due to the necessity of keeping the priority of these metrics for multiple objective
 184 functions in our theorem, we should consider remaining this good performance of
 185 association rules in the second tradeoff.

186 2.2 The First tradeoff

187 The first tradeoff aims at balancing different metrics for the multiple objective func-
188 tions mentioned above. For this purpose, Richardson Extrapolation formula and the
189 steepest descent method are utilised and extended to multiple objective functions
190 which can balance different metrics of fuzzy association rule simultaneously. The
191 pseudo-code is shown in Algorithm 2. *Line 2* compute the parameters based ob-
192 jective functions, while *Line 3* calculate corresponding derivatives by Richardson
193 Extrapolation method. Then, *Line 4* represents the processing of Algorithm 1 that
194 computes the direction for increasing objective functions together. After that, we up-
date the value of metrics φ through the selected path η and step size λ .

Algorithm 2 The First tradeoff: Balancing Multiple Metrics of Performance for As-
sociation Rule Mining

Input: initial (or previous) metrics φ , the maximum number of tradeoff rounds I ;

Output: optimised parameters φ ;

- 1: **for** $i = 0$ to I **do**
- 2: The processing of objective functions' computing;
- 3: The processing of derivatives' calculating [16] (Richardson Extrapolation approachis applied);
- 4: The processing of direction η selecting by Algorithm 1;
- 5: With the selected direction η , we can update the objective functions to a larger one

$$\varphi \leftarrow \varphi + \lambda\eta,$$

where λ is a user-defined step size;

6: **end for**

7: Return φ ;

195

196 **3 The Second tradeoff: Balancing the Effectiveness and Accuracy of Our**
197 **DOFARM Method**

198 The first level of our dual compromise scheme aims at optimising all of the metrics
199 for fuzzy association rules. While it has already optimised the preselecting metrics
200 of association rule mining, maintaining this optimised performance of these prese-
201 lecting metrics in the first level is becoming one of the basic tasks for our second
202 tradeoff spontaneously. Also, the first tradeoff has not updated the partitioning points
203 of the fuzzy-set membership functions with the parameter based metrics. Therefore
204 this updating procedure should be considered in the second tradeoff. In the meantime,
205 our dual compromise is still required to update the sets of frequent items and rules of
206 generated from them according to the optimised partitioning points. The parameters
207 related to the frequent itemsets are used to balance the number of elements in every
208 fuzzy set of our method. Therefore, our DOFARM method will dynamically discover
209 the optimised rules by the partitioning intervals and their frequent items of fuzzy
210 sets, which can be used to analyse new coming data and supplied to decision-making
211 efficiently [23], [24], [25], [26].

Algorithm 3 The Second tradeoff: Balancing Fuzzy sets and Partitioning Parameters

Input: the previous set of partitioning parameters \mathcal{X}_0 and the user-defined maximum number of rounds N_round_I (default as 5);

Output: optimised partitioning parameters \mathcal{X} ; the optimised set of frequent item-sets F and the optimised set of Association Rules R ;

- 1: Initialising \mathcal{X} with the previous set of parameters \mathcal{X}_0 ;
- 2: Generate frequent item-sets F and association rules R , make sure R contains only strong rules;
- 3: **for** $i = 0$ to N_round_I **do**
- 4: **for** $\forall f \in F$ **do**
- 5: Compute the value of $\text{Supp}(f, \mathcal{X})$;
- 6: **end for**
- 7: **for** $\forall r \in R$ **do**
- 8: Compute the value of $\text{Supp}(r, \mathcal{X})$; Compute the value of $\text{Conf}(r, \mathcal{X})$; Compute the value of $\text{CF}(r, \mathcal{X})$;
- 9: **end for**
- 10: The processing of the objective functions' computing;
- 11: The processing of the corresponding derivatives' calculating;
- 12: The processing of the suitable directions' searching η according to Algorithm 1;
- 13: Update the set of parameters with the searched direction η for larger value of our objective functions, update

$$\mathcal{X} \leftarrow \mathcal{X} + \lambda\eta,$$
 where λ denotes step size;
- 14: **end for**
- 15: Return the current value of parameters \mathcal{X} , the present set of frequent fuzzy items F and the current set of association rules R ;

212 The aim of our second tradeoff is to update the previous set of partitioning points
 213 generated from the first tradeoff. Also, the last set of association rules is applied
 214 when the current iteration is not the first one (Under-Optimised set of association
 215 rules is used in the first iteration). In our Algorithm 3, we can see that our the whole
 216 processing of the first tradeoff is shown as the *Lines* 2 to 14 and the three thresholds
 217 are updated from *Line* 4 to 9. After it, our multiple objectives optimisation procedure
 218 of the weighted parameter w is illustrated by the *Lines* 10 – 13, while we compute
 219 the proper direction of the Algorithm 1 through *Line* 12.

220 **4 The Dynamic optimisation based Fuzzy Association Rule Mining Method**

221 In this section, we describe the fuzzy association rule mining method based on dy-
 222 namic optimal parameters and metrics. The first subsection utilises an algorithm to
 223 further demonstrate the features of our dual compromise scheme and our DOFARM
 224 method. In the second subsection, the concrete steps of our DOFARM method are
 225 listed to interpret our process from the view of data processing further. As these
 226 processes and procedures displayed, we witness the operations of balancing the cor-
 227 responding metrics (support, confidence and certainty factor) with the first tradeoff,
 228 and the methods of adjusting the current parameters (partitioning points, fuzzy sets)
 229 in the second tradeoff. Eventually, we conduct a global dual tradeoff between the
 230 predefined metrics and optimised parameters.

231 Apart from all of these details, another theorem based on the previous theorem
 232 proposed in the last section is also certified rigorously in this section. It not only

233 demonstrates that the dual tradeoff enhances the performance of association rule min-
 234 ing theoretically but also illustrates a widely applied scheme to balance metrics of
 235 multiple functions and parameters related with high result-performing and low time-
 236 consuming simultaneously.

237 4.1 The dual compromise scheme

238 Our dual compromise scheme searches for the appropriate sets of association rules
 239 and frequent items through multi-aspect parameters, such as fuzzy sets and partition-
 240 ing points improved by the second tradeoff. The pseudo-code is shown in Algorithm
 241 4. The whole steps of our dual compromise scheme are introduced in *Lines* 1 – 8,
 242 while *Lines* from 1 to 5 illustrate the processing of initialisation and the *Lines* 6 – 8
 243 show the processing of how to optimise the set of partitioning points by the Algorithm
 244 3.

Algorithm 4 The dual compromise scheme

Input: the original data set D ; the threshold of support (min_Supp); the threshold of confidence (min_Conf); the user-defined maximum number of rounds in the second tradeoff N_round_J and the user defined maximum number of round in the first tradeoff N_round_I ;
Output: the balanced set of frequent items F and the balanced set of association rules R ;
 1: DP_1, \dots, DP_4 that is applied to distinguish different intervals is computed for a given continuous feature as described in paper [16];
 2: **for** $\forall \mathcal{X}$ components x_0, x_1 in every continuous feature **do**
 3: $x_0 \leftarrow 0.5 * (DP_0 + DP_1)$;
 4: $x_1 \leftarrow 0.5 * (DP_2 + DP_3)$;
 5: **end for**
 6: **for** $j = 0$ to N_round_J **do**
 7: The balance processing of optimising the set of fuzzy frequent items and partitioning parameters according to the Algorithm 3;
 8: **end for**
 9: Return the set of fuzzy frequent items F and the set of association rules of R by the *Line* 7;

245 As the first tradeoff optimises the metrics by using multiple objective functions
 246 and the second tradeoff aims to balance the performance of fuzzy association rules
 247 and the partitioning points, the strong rules of the second tradeoff process will be
 248 different from that of the first tradeoff. To fulfil the dual tradeoff and it's optimising
 249 operations, the value of our multiple objective functions should be kept nondecreas-
 250 ing. Taking φ_{10} , which is one of the most popular metrics in association rule mining,
 251 as an instance, we have the Theorem 2, the other objective functions are just as the
 252 same situation as φ_{10} .

253 **Theorem 2** *The value of objective function φ_{10} is non-decreasing during the dual*
 254 *tradeoff optimisation we proposed.*

255 *Proof* The optimisation we proposed consists of two levels of tradeoff. The second
 256 level of tradeoff reselect the association rules by redoing the frequent itemset discover-
 257 ing algorithm. The re-selection will either replace the original top 10 rules with ten

258 better rules whenever it is possible or keep the original ten rules otherwise. So the
259 dual tradeoff will either improve the value of φ_{10} , or maintain the value as it is. The
260 first level of tradeoff perform a gradient-based Multi-Objective optimisation (we call
261 it the first optimisation for convenience). The first tradeoff won't replace the top 10
262 rules; instead, it improves the quality of the top-ten rules since this quality is one of
263 its objective function according to Theorem 1. So both the second level and the first
264 level of our tradeoff ensure that the value of φ_{10} is nondecreasing.

265 In this way, our DOFARM method is proved as a generic measurement which
266 can be widely used to balance multiple objective functions. This theorem means our
267 objective function will be improved continuously both in the first and the second
268 tradeoff. Precisely, in the first tradeoff, we optimise the preselecting metrics by us-
269 ing our objective functions, then we change the partitioning points to enhance the
270 quality of the whole strong rule set in the second tradeoff. Afterwards the entire pro-
271 cedures of our dual compromise, we replace our entire strong rule set with better
272 one. If improving the partitioning points of the fuzzy-sets will increase the number of
273 rules above the given thresholds, then the dual compromise scheme we proposed will
274 hopefully increase this number as well because the optimisation is performed with
275 a set of objective functions that are related to the quality of the partitioning points.
276 The experimental study will be illustrated to show the further achievements of our
277 DOFARM method.

278 4.2 The Concrete Steps of the DOFARM Method We Proposed

279 Our DOFARM method is differing from the method of classic FARM concerning an
280 additional dual tradeoff. It can optimise the set of frequent items and calculate the
281 parameter based metrics, which are used as parameters for optimisation functions.
282 The first tradeoff shown in Figure 2 can generate the optimised set of parameters
283 for multiple objective functions. It optimises the performance of association rule en-
284 tirely. Then, the second level of our tradeoff balances the partitioning points of the
285 fuzzy-set membership functions based on optimal dynamic parameters. Eventually,
286 our dual compromise scheme computes the set of frequent items and the set of asso-
287 ciation rules concerning the fuzzy sets optimised in the algorithm of the first tradeoff.
288 The parameters are related to the frequent itemsets and used to balance the number
289 of elements in every fuzzy set in our method. Therefore, our DOFARM method will
290 dynamically discover the optimised set of association rules according to the continu-
291 ously improving our multiple objective functions, which can be used to analyse new
292 coming data and supplied to decision-making. All of the details will be presented in
293 this section.

294 5 Experimental Study

295 There are four subsections in this section. The first subsection explains the cor-
296 responding methods, parameters and datasets. The second subsection lists the an-
297 tecedents of strong rules and the results of partitioning points which indicate our

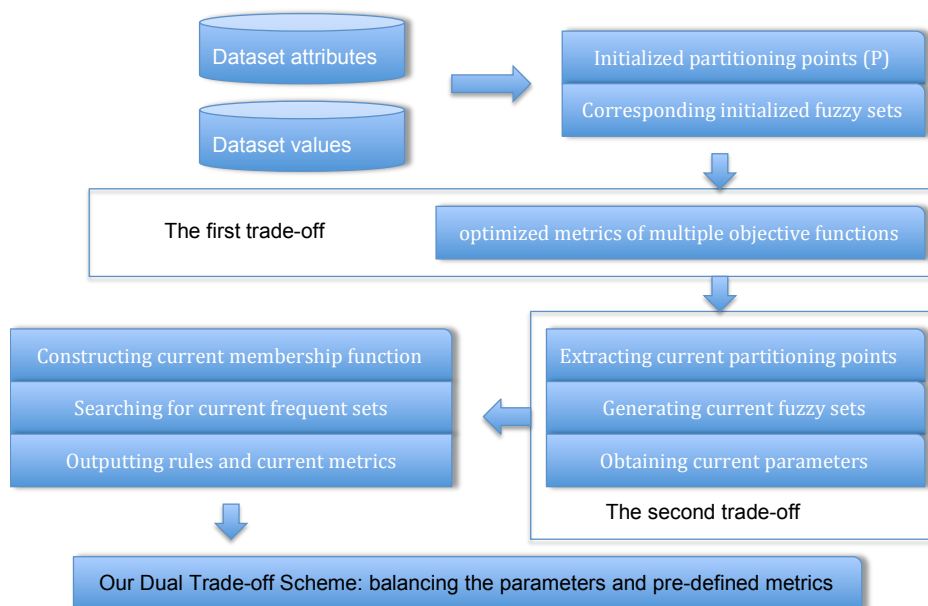


Fig. 2. Flowchart of Dynamic Optimal Parameter Based Fuzzy Association Rule Mining (DOFARM) method

298 rules have high accuracy and it follows the actual application as well. The third sub-
 299 section illustrates the statistics of the three methods which further account for the ef-
 300 fectiveness of our proposed DOFARM method. Finally, we compare our DOFARM
 301 with our previous work OFARM method [16] with data analysis of efficiency, as the
 302 method GFARM lose the general comparison conditions (details will be explained in
 303 this section). With all of the experimental studies, we can simply further represent
 304 the benefits of our DOFARM method, including the good performance of efficiency,
 305 effectiveness and accuracy expecting theoretical demonstrations in section 3.

306 5.1 Corresponding Methods and Experimental Datasets

307 In our experiment, the proposed DOFARM method is evaluated by comparing with
 308 GFARM method [3] and OFARM method of our previous work [16]. From the exper-
 309 imental descriptions among this section, we see our novel DOFARM method extends
 310 GFRAM and OFARM method to arbitrary parameters and metrics and improves it
 311 on accuracy, effectiveness and efficiency. The function of membership values for the
 312 GFARM, OFARM and our DOFARM method is already shown in the paper [16]
 313 and the strong rules are defined in [17]. A data set of “Pima Indians Diabetes” from
 314 UCI repository, is applied to display the outputting rules and compare the differences
 315 among the partitioning points of the three methods. The other data set coming from
 316 the Massachusetts General Hospital/Marquette Foundation (MGH/MF) Waveform

317 Database is applied to demonstrate the effectiveness and efficiency of our DOFARM
 318 method. The metric φ_{10} in section 2.1 is collected as one metric of effectiveness. The
 319 other metric of effectiveness is the number of the strong outputting rules. The user-
 320 defined maximum number of the second tradeoff algorithm and the number of the set
 321 of frequent fuzzy items are $N_round_J = 5$ and $q = 3$. The details of DOFARM
 322 we proposed are shown in Figure 2. The pruning method [22] of our experiments
 323 is applied to filter the set of association rules and prevent the huge amount of the
 324 number of rules. Following results in the three methods: GFARM, OFARM and our
 325 DOFARM will be shown as the average of five procedures, which is used to cut the
 326 randomness during our experiments.

327 The higher the value of thresholds are chosen, the better rules are generated, and
 328 then there will be a limited number of strong rules. So if the value of thresholds is
 329 set to be too high, the generated rules will normally be too narrow, while the value of
 330 thresholds is set to be too low, the quality of the generated rules will be too poor to
 331 be interesting. Thus, to manifest the exquisite adaptability of our DOFARM method,
 332 different thresholds of min_Supp and min_Conf are outputted and compared in our
 333 experiments. Therefore, we can prove the proposed DOFARM method according to
 334 a vast range of thresholds and then compare the differences.

335 5.2 Outputting of Strong Rules and Accuracy Comparisons

336 The set of association rules that are related to diabetes we discovered from “Pima
 337 Indians Diabetes” data set is represented in this subsection. All of the features and
 338 their IDs are described in the following items.

- 339 – 0: Number of times pregnant;
- 340 – 1: Plasma glucose concentration a 2 hours in an oral glucose tolerance test;
- 341 – 2: Diastolic blood pressure (mm Hg);
- 342 – 3: Triceps skin fold thickness (mm);
- 343 – 4: 2-Hour serum insulin (μ U/ml);
- 344 – 5: Body mass index (weight in kg/(height in m)²);
- 345 – 6: Diabetes pedigree function;
- 346 – 7: Age (years);
- 347 – 8: Class variable (0 or 1).

348 The interesting and strong rules are defined and generated in this section. We
 349 firstly group continuous features from ‘0’ to ‘7’ into three sets of frequent fuzzy
 350 items. The leaving feature ‘8’ is a label of having diabetes or not (the value of ‘0’
 351 is recognised as healthy people, and the value of ‘1’ represents the people who are
 352 suffering from diabetes). We only print the strong rules with their consequent (8, 1)
 353 in our experiments, which denotes the 8 – th feature and its value is 1. So one of the
 354 interesting and strong rules can be shown as the form as $(4, 2)(7, 2) \rightarrow (8, 1)$. This
 355 outputted rule means when the second fuzzy set of the 4 – th feature and the second
 356 fuzzy set of the 7-th feature coincide; then the current individual can be indicated
 357 as diabetes. Our three thresholds are defined by $min_Supp = 0.1, min_Conf =$
 358 0.7 and $min_CF = 0.1$. All of the antecedents are shown separately without the

Table 1. Comparison of Interesting and Strong Rules in three methods.

Comparing Item	GFARM	OFARM	DOFARM
Antecedent (4, 2)	Containing	Containing	Containing
Antecedent (7, 2)	Containing	Containing	Containing
Antecedent (1, 2)	Containing	Containing	Containing
Antecedent (3, 2)	None	Containing	Containing
Antecedent (5, *)	None	(5, 2)	(5, 1)
Antecedent (2, *)	None	None	(2, 1)
Antecedent (6, *)	None	None	(6, 1)
Total Antecedent	3	5	7
Total Mid-Antecedent	0	0	3
Total rule	2	7	9

359 common consequent (8,1) in Table 1, where ‘*’ means any possible value. Take
 360 (4,2)(7,2) → (8,1) as instance, it will be divided into two antecedent (4,2) and
 361 (7,2).

Table 2. Partitioning points comparisons in three methods.

Model	Partitioning Points (Fuzzy Sets)
GFARM	$M_{L,0} = 1.5, M_{R,0} = 5.5;$
	$M_{L,1} = 102, M_{R,1} = 136;$
	$M_{L,2} = 66, M_{R,2} = 78;$
	$M_{L,3} = 25.5, M_{R,3} = 32.5283;$
	$M_{L,4} = 121.372, M_{R,4} = 168.519;$
	$M_{L,5} = 28.3, M_{R,5} = 35.75;$
	$M_{L,6} = 0.2615, M_{R,6} = 0.572;$
OFARM	$M_{L,7} = 25, M_{R,7} = 38;$
	$M_{L,0} = 1.8494, M_{R,0} = 6.9014;$
	$M_{L,1} = 95.0285, M_{R,1} = 125.031;$
	$M_{L,2} = 69.8885, M_{R,2} = 74.0415;$
	$M_{L,3} = 27.9882, M_{R,3} = 30.0862;$
	$M_{L,4} = 137.702, M_{R,4} = 158.103;$
	$M_{L,5} = 30.3525, M_{R,5} = 33.7087;$
DOFARM	$M_{L,6} = 0.2198, M_{R,6} = 0.6890;$
	$M_{L,7} = 26.9063, M_{R,7} = 33.0771;$
	$M_{L,0} = 1.9866, M_{R,0} = 6.9007;$
	$M_{L,1} = 108.966, M_{R,1} = 125.057;$
	$M_{L,2} = 69.902, M_{R,2} = 81.9074;$
	$M_{L,3} = 27.9887, M_{R,3} = 30.0819;$
	$M_{L,4} = 137.717, M_{R,4} = 163.632;$
$M_{L,5} = 30.3597, M_{R,5} = 37.7868;$	
$M_{L,6} = 0.2197, M_{R,6} = 0.6890;$	
$M_{L,7} = 26.9798, M_{R,7} = 33.0393;$	

362 According to the Table 1, we can observe that our proposed DOFARM method
 363 discovers seven different antecedents in all, while the OFARM gets five and GFARM
 364 has only three antecedents. Different from these common five antecedents with GFARM
 365 and OFARM, our DOFARM has two new antecedents (2,1) and (6,1), which means
 366 Diastolic blood pressure and Diabetes pedigree function have some relations with

367 diabetes. The proposed DOFARM finds strong rules with more disease-related an-
 368 tecedents and more non-high antecedents. In real-world applications, the more amount
 369 of the features of the disease-related antecedents are, the more useful of the rule is.
 370 The DOFARM method we proposed, therefore, shows its first merit with two more
 371 disease-related antecedents. The second merit of our DOFARM method is that two
 372 rules associated with non-high value antecedents are discovered by our DOFARM
 373 method, while the methods of GFARM and OFARM find nothing. With general
 374 knowledge about association rule [22] and the interesting rule we defined (which
 375 is related to disease), the more interesting rules are filtered, better is the method.
 376 The GFARM and OFARM perform not well since they find less disease-related an-
 377 tecedents, less non-high value antecedents and less interesting rules. By contrast, our
 378 DOFARM generates a higher amount of interesting and strong rules, and it outputted
 379 rules seem to be more useful and productive in this light.

380 If the new continuous data is coming, we can use the same membership func-
 381 tion and fuzzy sets defined by previous data to handle new data. Suppose there is
 382 an individual like Diastolic blood pressure of 80 mm Hg and 2-Hour serum insulin
 383 164 μ U/ml. Firstly, we can look up and find there are expressed as '2' and '4'
 384 respectively and transform them into fuzzy sets: Diastolic blood pressure (0, 0.8894,
 385 0.1106) and 2-Hour serum insulin (0, 0.028, 0.972); Secondly, in Table 2, there are
 386 two related antecedents (2, 1) and (4, 2) and the rule (2, 1)(4, 2) \rightarrow (8, 1) is found in
 387 our proposed DOFARM method; Thirdly, we can see the partitioning points in Table
 388 1 as $M_{L,2} = 69.902$, $M_{R,2} = 81.9074$ and $M_{L,4} = 137.717$, $M_{R,4} = 163.632$; At
 389 last, we can see the individual have a high possibility of diabetes disease since the
 390 membership grades of (2, 1) and (4, 2) are high.

391 5.3 An example of segmental computing

392 Our algorithm of DOFARM can be widely applied in different applications, such
 393 as medicine, finance and affective and segmental computing. This subsection will
 394 illustrate an example of how our DOFARM applied in emotional and sentimental
 395 computing.

396 Emotions and sentiments have profound influences on medical treatments. In this
 397 paper, two sentiment strengths will be considered: positive (sentiment benefits for
 398 well-being) and negative (sentiment harms to well-being). For instance, people whose
 399 sentiment strengths are extremely positive would be active in treatments of control-
 400 ling their unhealthy conditions. Patients with positive sentiment can enjoy their lives
 401 even if they are diagnosed with type 2 diabetes, coronary heart disease or cancers.

402 Subsection 5.2 shows the rules of diagnosing diabetes that can classify people into
 403 two groups: diabetes and nondiabetes. Our primary concern of classifying sentiment
 404 is the group of people who are diabetes, so we assume that the sentiment strength of
 405 nondiabetes will be extremely positive and then we entirely ignore this group in this
 406 subsection.

407 To be more simple and without loss of generality, we suppose only two attributes
 408 (body mass index and age) of the diabetes group that are related to sentiments. Then,
 409 applying our DOFARM on data of Table 3, the proposed method may generate rules

410 related with sentiments as $BMI \leq 25$ and $Age \leq 33 \rightarrow sentiments : positive$ and
 411 $BMI \geq 30$ and $Age \geq 67 \rightarrow sentiments : negative$. Therefore, we can predict
 412 sentiments by the generated rules of our DOFARM.

Table 3. Sentimental data.

BMI	Age	Sentiments
60	23	positive
20	80	negative
...

413 5.4 Effectiveness Comparisons and Analysis

414 In this subsection, we use two metrics to evaluate our DOFARM's effectiveness ac-
 415 curately:

- 416 1. the number of rules;
- 417 2. the average value of top ten strong rules: φ_{10} in section 2.1, which combines
 418 values of min_Supp, min_Conf and min_CF, is used as metrics for the quality of
 419 rules.

420 The data set coming from the Massachusetts General Hospital / Marquette Found-
 421 ation (MGH/MF) Waveform Database is applied to compare our proposed DO-
 422 FARM method with GFARM and OFARM. For simplicity, we only take the record
 423 of mgh10 to assess our DOFARM method. The recording includes eight features:
 424 three ECG leads, arterial pressure, pulmonary arterial pressure, central venous pres-
 425 sure, respiratory impedance, and airway CO2 waveforms. We computed the average
 of five-time procedures to reduce the randomness of our experimental results.

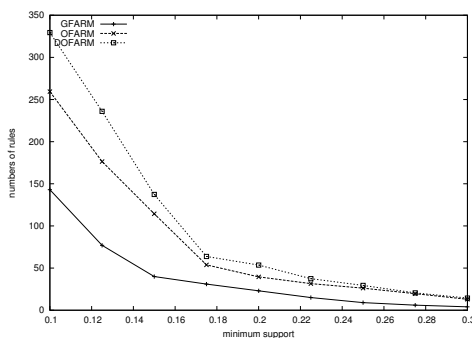


Fig. 3. Number of Rules comparison with different Minimum support

427 Among all these three methods: GFARM, OFARM and DOFARM, all of their
 428 number of rules show a downward trend with the growing min_Supp, which follows
 429 the property of min_Supp: the larger the min_Supp, the fewer association rules are
 430 filtered. However, there are still some differences in the changing process: the gap
 431 between OFARM and DOFARM is smaller than the difference between GFARM and
 432 OFARM at almost every point. That is to say, our DOFARM which is extended from
 433 OFARM not only inherits the benefits of OFARM but also exceeds the OFARM.
 434 Moreover, our DOFARM performs much better than other methods whether the orig-
 435 inal results are good or not (the min_Supp is small or large).

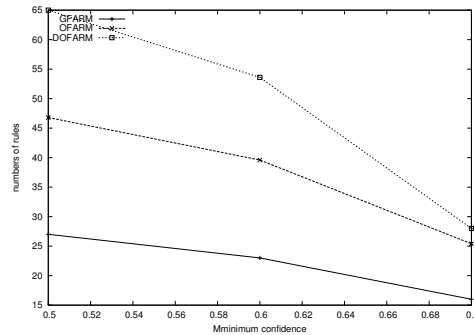


Fig. 4. Number of Rules comparison with different Minimum confidence

436 Combining the results in Figures 3 and 4, the number of rules for our method
 437 DOFARM is greater. The improvements of proposed DOFARM are more satisfactory
 438 when the original result is poor (with small min_Supp or small min_Conf) than the
 439 improvements with large min_Supp or large min_Conf. Then, the DOFARM method
 440 we proposed will get a much higher number of strong rules when the OFARM and
 441 GFARM are not good enough. Moreover, our DOFARM method can retain the ben-
 442 efits of OFARM method and can get a better result even if the results of OFARM are
 443 already pretty good.

444 The quality of fuzzy association rules is also used to verify the effectiveness of
 445 our DOFARM method. Taking the top ten rules as an example, the Figure 5 witnesses
 446 the quality of average of top ten rules decreasing according to the gradually increas-
 447 ing value of min_Supp with fixed min_Conf (0.6) and fixed min_CF (0.1). As we can
 448 manifest from the Figure 5, the φ_{10} , which means the average of quality (the mini-
 449 mum of support, confidence and certainty factor) of the top ten rules, drops when the
 450 min_Supp rises from 0.1 to 0.3. Among all these three methods, the DOFARM we
 451 proposed is always staying the highest column; the OFARM lies the column which is
 452 a little lower than the DOFARM column, while the GFARM is illustrated as the low-
 453 est. The difference of OFARM and our DOFARM in the histogram is still noticeable,
 454 and the column of our DOFARM shows its improvement at every different min_Supp
 455 in our experiments.

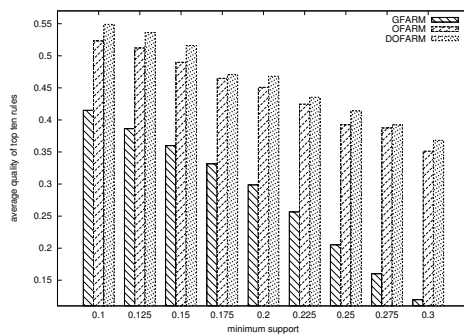


Fig. 5. Top Ten rules quantity comparison with different Minimum support

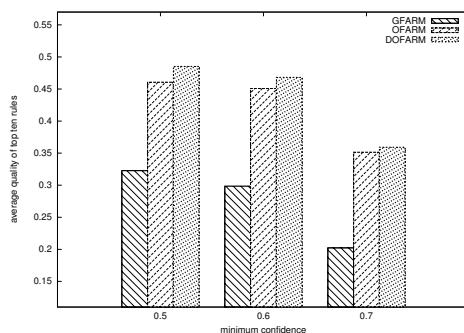


Fig. 6. Top Ten rules quantity comparison with different Minimum confidence

456 The situation of quality φ_{10} with the increasing value of min_Supp is just similar
 457 to the case in increased min_Conf with fixed min_Supp (0.2) and min_CF (0.1), which
 458 is indicated in Figure 6. The column of our DOFARM is still higher than the other
 459 two columns of GFARM and OFARM. So we can say our DOFARM can generate
 460 more suitable rule sets than other compared methods.

461 With Figure 5 and 6, our DOFARM method demonstrates greater effectiveness
 462 comparing with the GFARM and OFARM method. Then, the proposed DOFARM
 463 method outperforms the other two methods concerning both the quantity of outputted
 464 rules and the quality of interesting rules. To be more specific, optimising the set of
 465 partitioning parameters enhances the amount of our outputted rules; while the selected
 466 parameters of the functions of objectives increase the quality of interesting
 467 rules with the thresholds of min_Supp and min_Conf.

468 5.5 Efficiency Comparisons and Analysis

469 In this section, we consider estimating the efficiency of our DOFARM method. To
 470 certify the performance of our DOFARM method, we compare our method with the
 471 previous work OFARM method. As GFARM method runs once only and OFARM,

472 DOFARM method runs several times to balance the different metrics of fuzzy asso-
 473 ciation rule mining process, GFARM method lose the necessity of comparison while
 474 parameter time is related.

Following the general rule, we use the program running time t and the two effective-
 ness metrics: the number of strong rules N and the average of the quantity of best
 ten rules φ_{10} together, then generate the formula for efficiency as follows:

$$Efficiency = \frac{N \cdot \varphi_{10}}{t}.$$

475 For example, when $\min_Supp = 0.3$ and $\min_Conf = 0.7$, we can compute $Efficiency$
 476 of OFARM and DOFARM by $Efficiency_{(0.3,0.7)}^{OFARM}$ and $Efficiency_{(0.3,0.7)}^{DOFARM}$ as
 477 follows:

$$Efficiency_{(0.3,0.7)}^{OFARM} = \frac{25.4 \times 0.3766452}{21.18} = 0.45168971 \approx 0.4517.$$

$$Efficiency_{(0.3,0.7)}^{DOFARM} = \frac{28 \times 0.3866068}{11.0388} = 0.980631083 \approx 0.9807.$$

478 As this formula illustrated, we sort out the $Efficiency$ with different \min_Supp
 479 from 0.1 to 0.3 and fixed \min_Conf (0.6) in Table 4.

Table 4. Efficiency comparison with different Min-support.

Min-support	0.1	0.125	0.15	0.175	0.2
OFARM	2.7118	1.6863	1.8614	0.7225	0.9366
DOFARM	4.6485	3.2722	3.8347	1.8567	3.9215
Min-support	0.225	0.25	0.275	0.3	Average
OFARM	0.5047	0.1918	0.1548	0.1277	0.9886
DOFARM	0.7902	0.3067	0.2697	0.1384	2.1154

480 As it is described in Table 4, we can witness the good performance of the DO-
 481 FARM we presented, as the $Efficiency$ of our DOFARM is always larger than the
 482 $Efficiency$ of OFARM method. Also, the differences between our DOFARM and
 483 the OFARM method show decreasing trends with more and more strict threshold
 484 (\min_Supp) from 0.1 to 0.2. During the period of changing the value of \min_Supp ,
 485 the gap shrinks marginally from 0.225 to 0.3. More particularly, in Table 4 there
 486 are nearly four times $Efficiency$ of our DOFARM at $\min_Supp = 0.2$ than the
 487 $Efficiency$ of OFARM, which demonstrate the DOFARM we proposed in this pa-
 488 per is almost two times of $Efficiency$ as its counterpart.

489 The $Efficiency$ with different \min_Conf s and fixed $\min_Supp = 0.2$ are il-
 490 lustrated in Table 5, which also describe better performance of efficiency of our
 491 DOFARM method than that of the OFARM method. From Table 5, we can further
 492 demonstrate the $Efficiency$ of the DOFARM method we proposed is much higher
 493 than the $Efficiency$ of OFARM method. In summary, we can conclude that our
 494 DOFARM performs much better than other two methods: GFARM and OFARM in

Table 5. Efficiency comparison with different Min-confidence.

Min-confidence	0.5	0.6	0.7
OFARM	0.7124	0.9366	0.4517
DOFARM	1.9896	1.4543	0.9806

495 both effectiveness and accuracy metrics. So the DOFARM method is demonstrated as
 496 the better one to generate larger amount of quantity of strong rules and better quality
 497 efficient rules.

498 6 Conclusion

499 A dynamic optimisation fuzzy-association-rule mining method has been proposed ac-
 500 cording to the definition of the dual compromise measurement. We have shown that
 501 the balancing procedures of the parameter-based-metrics make the proposed method
 502 easy to formulate and valid to balance parameters and metrics simultaneously for con-
 503 tinuous data. In the algorithm of our dual compromise, the set of fuzzy association
 504 rules and the set of frequent items are optimised by the selected set of partitioning
 505 parameters. After outputting association rule of the three methods GFARM, OFARM
 506 and our DOFARM, the accuracy of the rule sets has been certified. The experiment
 507 also demonstrates that our DOFARM method is capable of balancing the parameters
 508 of the quality of interesting rules and the quantity of outputted rules; that is, the ef-
 509 fectiveness of our DOFARM method exceeds the other two approaches. Also, after
 510 comparing with OFARM method, the efficiency of the DOFARM we proposed is al-
 511 most two times of its counterpart, as the consuming time of our method has been
 512 reduced to half as OFARM method averagely. In conclusion, our DOFARM method
 513 outperforms its peers - GFARM and OFARM in accuracy, effectiveness and effi-
 514 ciency. Furthermore, the results of our experiments for gradual changes of min.Supp
 515 and min.Conf show the stability and robustness of the DOFARM method we pro-
 516 posed. In this paper, problems of both diabetes and sentiment strength employ our
 517 DOFARM method to accomplish their solutions of classification.

518 **Acknowledgements** This work is supported in part by The ARC Discovery Early Career Research Award
 519 (DE130100911), The ARC Discovery Project (DP130101327), The ARC Linkage Project (LP100200682),
 520 The NSFC funding(61332013), The International Science and Technology Cooperation Projects(No.2016D10008,
 521 2013DFG12810, 2013C24027), The Municipal Natural Science Foundation of Ningbo(No.2015A610119),
 522 The Natural Science Foundation of of Zhejiang Province (No. Y16F020002).

523 References

- 524 1. Q. Zhou, P. Shi, S. Xu, H. Li, "Adaptive Output Feedback Control for Nonlinear Time-Delay Systems
 525 by Fuzzy Approximation Approach," *IEEE Transactions on Fuzzy Systems*, vol. 21, no. 2, pp. 301-313,
 526 2013.
- 527 2. J.H. Lee and H.L. Kwang, "An extension of association rules using fuzzy sets," *In Proceedings of the*
 528 *Seventh IFSA World Congress*, pp. 399-402, 1997.

- 529 3. M. Delgado, N. Marin and D. Snchez, "Fuzzy association rules: general model and applications," *IEEE*
530 *Transactions on Fuzzy Systems*, vol. 11, no. 2, pp. 214-225, 2003.
- 531 4. X.Z. Wang, R.A.R. Ashfaq, A.M. Fu, "Fuzziness based sample categorization for classifier performance
532 improvement", *Journal of Intelligent and Fuzzy Systems*, vol. 29, no. 3, pp. 1185-1196, 2015.
- 533 5. M. De Cock, C. Cornelis, E. E. Kerre, " Fuzzy Association Rules: a Two-Sided Approach," *In Proceed-*
534 *ings of Internal Conference on Fuzzy Information Processing-Theories and Applications*, pp. 385-390,
535 2003.
- 536 6. C.L. Chen, F.S.C. Tseng, T. Liang, "An integration of WordNet and fuzzy association rule mining for
537 multi-label document clustering," *Data and Knowledge Engineering*, vol. 69, pp. 1208-1226, 2010.
- 538 7. J. Alcalá-Fdez, R. Alcalá, and F. Herrera, "A fuzzy association rule-based classification model for high-
539 dimensional problems with genetic rule selection and lateral tuning," *IEEE Transactions on Fuzzy Sys-*
540 *tems*, vol. 19, no. 5, pp. 857-872, 2011.
- 541 8. V.V. Rao, E. Rambabu, and G. Sriramganes, "Effective Association rule mining using Fuzzy Apriori
542 and Weighted Fuzzy Apriori," *IJECCE*, vol. 3, no. 3, pp. 381-386, 2012.
- 543 9. J. Alcalá-Fdez , R. Alcalá, M.J. Gacto, and F. Herrera, "Learning the membership function contexts for
544 mining fuzzy association rules by using genetic algorithms," *Fuzzy Sets and Systems*, vol. 160, No. 7,
545 pp. 905-921, 2009.
- 546 10. Y. Rao, H. Xie , J. Li and F. Jin , F. L. Wang, and Q. Li , "Social emotion classification of short text
547 via topic-level maximum entropy model," *Information and Management*, vol. 53, No. 8, pp. 978-986,
548 2016.
- 549 11. H. Xie , X. Li , T. Wang , R. Lau , T. L. Wong, L. Chen, and Q. Li , " Incorporating sentiment into tag-
550 based user profiles and resource profiles for personalized search in folksonomy," *Information Processing*
551 *and Management*, vol. 52, No. 1, pp. 61-72, 2016.
- 552 12. A. Dridi, and D. R. Recupero , "Leveraging semantics for sentiment polarity detection in social me-
553 dia," *International Journal of Machine Learning and Cybernetics*, vol. 1, No. 11, 2017.
- 554 13. W. Wang , G. Tan, and H. Wang , "Cross-domain comparison of algorithm performance in extracting
555 aspect-based opinions from Chinese online reviews," *International Journal of Machine Learning and*
556 *Cybernetics*, vol. 8, No. 3, pp. 1053-1070, 2017.
- 557 14. O. B. Omar ,C. Boschi-Pinto, and A. D. Lopez, "AGE STANDARDIZATION OF RATES: A NEW
558 WHO STANDARD," *World Health Organization*, vol. 13, no. 2, pp. 167-192, 2001.
- 559 15. X. Li, H. Xie, L. Chen, J. Wang, and X. Deng, "News impact on stock price return via sentiment
560 analysis", *Knowledge-Based Systems*, vol. 69, pp. 14-23, 2014.
- 561 16. H. Zheng, J. He, G.Y. Huang, and Y.C. Zhang, "optimised Fuzzy Association Rule Mining for Quan-
562 titative Data," *In Proceedings of 2014 IEEE International Conference on Fuzzy Systems*, pp. 396-403,
563 2014.
- 564 17. M. Delgado, D. Sánchez, M. J. Martín-Bautista, and M.A. Vila, "Mining association rules with im-
565 proved semantics in medical databases." *Artificial Intelligence in Medicine*, vol. 21, no. 1, pp. 241-245,
566 2001.
- 567 18. J. Huang, M. Peng, H. Wang, J. Cao, W. Gao, and X. Zhang , "A probabilistic method for emerging
568 topic tracking in microblog stream." *World Wide Web*, vol. 20, no. 2, pp. 325-350, 2017.
- 569 19. J. Ma, L. Sun, H. Wang, Y. Zhang, and U. Aickelin, "Supervised anomaly detection in uncertain
570 pseudoperiodic data streams." *ACM Transactions on Internet Technology (TOIT)*, vol. 16, no. 1, pp. 1-4,
571 2016.
- 572 20. H. Wang, Y. Zhang, and J. Cao, "Effective collaboration with information sharing in virtual universi-
573 ties." *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 6, pp. 840-853, 2009.
- 574 21. H. Wang, J. Cao, and Y. Zhang, "A flexible payment scheme and its role-based access control." *IEEE*
575 *Transactions on Knowledge and Data Engineering*, vol. 17, no. 3, pp. 425-436, 2005.
- 576 22. J. Han, M. Kamber, and J. Pei," *Data mining: concepts and techniques*, Morgan kaufmann, pp. 229-
577 242, 2006.
- 578 23. J. Zhang, X. Tao, and H. Wang," *Outlier detection from large distributed databases*, World Wide Web,
579 vol. 17, no. 4, pp. 539-568, 2014.
- 580 24. M.E. Kabir, H. Wang, and E. Bertino," *Efficient systematic clustering method for k-anonymization*,
581 *Acta Informatica*, vol. 48, no. 1, pp. 51-66, 2011.
- 582 25. F. Khalil, J. Li, and H. Wang," *An integrated model for next page access prediction*, International
583 *Journal of Knowledge and Web Intelligence*, vol. 1, no. 1, pp. 48-80, 2009.
- 584 26. X. Sun, H. Wang, J. Li, and J. Pei," *Publishing anonymous survey rating data*, Data Mining and
585 *Knowledge Discovery*, vol. 23, no. 3, pp. 379-406, 2011.