Early version, also known as pre-print

Link to published version (if available):
10.1109/T2FZZ.2013.6613293

Link to publication record in Explore Bristol Research
PDF-document

University of Bristol - Explore Bristol Research
General rights

Using Nonstationary Fuzzy Sets to Improve the Tractability of Fuzzy Association Rules

Simon Coupland, *Member, IEEE*
and Stephen G. Matthews, *Member, IEEE*

Abstract— **Modern organisations now collect very large volumes of data about customers, suppliers and other factors which may impact upon their business. There is a clear need to be able to mine this data and present it to decision makers in a clear and coherent manner. Fuzzy association rules are a popular method to identifying important and meaningful relationships within large data sets. Recently a fuzzy association rule has been proposed that uses the 2-tuple linguistic representation. This paper presents a methodology which makes use of nonstationary fuzzy sets to post process 2-tuple fuzzy association rules reducing the size of the mined rule set by around 20% whilst retaining the semantic meaning of the rule set.**

I. INTRODUCTION

There has a been surge in the collection, storage and processing of data in recent decades [1]. The problem of obtaining useful information from large databases is both timely and challenging. Association rule mining is a method for finding correlations between in items in a database [2], such as "customers who purchased beer also purchased pizza". A recent method for mining temporal fuzzy association rules has overcome a previously unrecognised problem where not all interesting rules are discovered [3]. The temporal 2-tuple algorithm by Matthews *et al* [3] is a complementary method to existing methods, such as FuzzyApriori [4], that adds more rules to the final set of rules. The solution by Matthews *et al* uses the 2-tuple linguistic representation [5] for fuzzy association rules, which has previously been used for learning the context of fuzzy association rules [6]. However, the additional knowledge produced from the temporal 2-tuple algorithm in the form of rules requires more work by a decision maker or domain expert when determining which rules are actionable. In this paper we present a method using nonstationary fuzzy association rules to reduce the final rule set of 2-tuple fuzzy association rules. We obtain a reduction of around 20% in the size of the rule set whilst maintaining the rule semantics.

The remainder of the paper is structured as follows. Section II describes the motivation and the algorithm that uses the 2-tuple linguistic representation. Section III describes nonstationary fuzzy sets. Section IV details our method for creating a nonstationary fuzzy association rule. In Section V a case study with Web log data is presented. In Section VI

there is a discussion of the findings, and in Section VII the conclusion of our work is made.

II. TEMPORAL 2-TUPLE FUZZY ASSOCIATION RULES

The motivation that led to a new approach of fuzzy association rule mining is described before an overview of the new approach is given. Fuzzy association rules describe quantities of items with linguistic terms, e.g., "customers who purchased *lots* of beer also purchased *lots* of pizza". Temporal association rules represent the frequency change of a rule, e.g., a rule may have a lifespan [7] when a one-off event occurs such as hurricane Katrina [8]. Matthews *et al* [3] identified a new problem in the field of association rule mining. A problem arises when combining fuzzy association rules with temporal association rules. Some temporal fuzzy association rules can be lost.

Traditional fuzzy association rule mining algorithms define the linguistic terms and membership functions first and then use these in the mining algorithm. This means the linguistic terms and membership functions remain the same from the start of the dataset to the end of the dataset. However, contextual changes in the meaning of linguistic terms can vary over time. For example, the contextual meaning of linguistic terms may change with events such as seasonal weather or sports events [9]. When the contextual meaning changes, the membership functions do not change to match the contextual change of the linguistic terms. Matthews *et al* [3] demonstrated how traditional approaches can lose some temporal fuzzy association rules as a result of this problem.

The temporal 2-tuple algorithm was demonstrated in [3] to be able to discover temporal fuzzy association rules that a traditional approach could not. The novel approach used the 2-tuple linguistic representation [5] and a genetic algorithm (GA) [10]. The 2-tuple linguistic representation is a symbolic translation of a fuzzy set. A symbolic translation is the lateral displacement of the fuzzy set within the interval $[-0.5, 0.5)$ that expresses the domain of a term when it is displaced between two linguistic terms. The 2-tuple linguistic representation maintains the semantic interpretability because the the linguistic term is not changed. As with the traditional approach, the linguistic terms and membership functions are defined first. However, the difference is the GA simultaneously searches for lateral displacements and rules. The benefit is that each lateral displacement is specific to a rule in a temporal interval, i.e., the context of the linguistic term is specific to a rule in a temporal interval. One run of the GA produces one rule using the Michigan representation. So

Simon Coupland is with the Centre for Computational Intelligence, De Montfort University, Leicester, United Kingdom, LE1 9BH and Stephen G. Matthews is with the Department of Engineering Mathematics, University of Bristol, Queen's Building, Bristol, BS8 1TR (email: simonc@dmu.ac.uk,stephen.matthews@bristol.ac.uk).

Iterative Rule Learning (IRL) [11] repeatedly runs the GA for a fixed number of iterations to produce a fixed number of rules.

The flexibility of the 2-tuple linguistic representation and the GA allows the discovery of rules that traditional a method cannot. The temporal 2-tuple algorithm is not an exhaustive search method, so it is a complementary method that provides additional knowledge to traditional methods. However, more rules are a burden on the decision maker or domain expert who uses the rules, so a method of reduction can enhance rule tractability.

III. NONSTATIONARY FUZZY SETS

Nonstationary fuzzy sets were first proposed by Garibaldi *et al* [12], [13], [14] in order to capture the variation facet of uncertainty within a fuzzy set. Many fuzzy concepts are subject to slight variations with time or other latent variables. If we consider the temporal 2-tuple algorithm, one of the main purposes of using the 2-tuple linguistic representation of a fuzzy set is to is allow rules to be mined which do not quite fit the existing fuzzy sets in a particular temporal period. With small variations to the membership functions, additional rules containing important information can be mined.

Nonstationary fuzzy sets allow these variations to be modelled within a single set which is a function of both the domain and time at which the fuzzy set is used. More formally Garibaldi *et al* define a nonstationary fuzzy set $\dot{A}$ over universe X as characterised by a membership function $\mu_{\dot{A}} : T \times X \to [0,1]$ which maps an ordered pair of values (t,x) to a membership grade in $[0,1]$. Note also that:

$$\mu_{\dot{A}}(t,x) = \mu_A(x, \bar{\tau}(t)) \qquad (1)$$

where $\bar{\tau}(t) = [\tau_1(t), \ldots, \tau_m(t)]$ and $\tau_i(t) = \tau_i + k_i f_i(t)$ for $i = 1, \ldots, m$. So, the variation of each parameter depends on time and a weighted combination of perturbation functions. In this particular application we are only interested in one type of variation referred to by Garibaldi *et al* as variation in location. Consider a simple discrete perturbation function $f(t) : T \times X$ where $f = \{1 \mapsto -0.5, 2 \mapsto 0.75, 3 \mapsto 1.5\}$ and the fuzzy set A with a triangular membership function defined by the parameters $(1.5, 3, 5)$. Let $\dot{A}$ be a nonstationary fuzzy set with the membership function of A and the perturbation function f. Figure 1 depicts the fuzzy A and instantiations of $\dot{A}$ at times $t = 1$, 2 and 3.
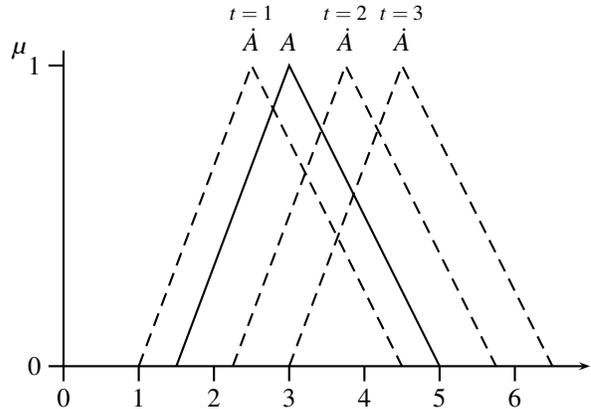

Fig. 1. The Fuzzy Set A and Nonstationary Fuzzy Set $\dot{A}$

Nonstationary fuzzy sets have been employed in a range applications mostly in the medical decision making domain [15]. They have also been used in control problems [16], [17] and have been compared to systems employing type-2 fuzzy sets [18]. There is a clear relationship between nonstationary fuzzy sets and type-2 fuzzy sets. Both approaches are attempting to capture uncertainty about a concept: type-2 are attempting to model the vagueness about a membership function, and nonstationary fuzzy sets are trying to model the variation in a membership function.

There is still a question about what causes observed variation in the membership function of a fuzzy set. Our opinion based on working with observed variation is that it is natural and is due to a number of contextual variables such as time and environmental conditions plus some additional noise.

IV. NONSTATIONARY FUZZY ASSOCIATION RULES

We define a nonstationary fuzzy association rule as a fuzzy association rule where at least one of the fuzzy sets is a nonstationary fuzzy set. We now consider how such a rule can be constructed and used to reduce the cardinality of a fuzzy association rule set with the goal of improving tractability.

A. Rule Construction

In this paper we construct the nonstationary fuzzy association rule by taking the rules mined using the temporal 2-tuple algorithm as a starting point. The temporal 2-tuple algorithm produces a set of n fuzzy association rules where the fuzzy sets make use of the 2-tuple linguistic representation. So, each rule contains associative relationships between fuzzy variables at a particular time interval. Since the 2-tuple linguistic representation is used the exact construction of these membership functions varies according to time. In order to represent these sets using nonstationary fuzzy sets the relationship between the lateral displacement value α and time must be understood and modelled with a function. There are a wide range of approaches to do this and we do not prescribe any particular method, but instead recommend that

a careful analysis of the data be used to inform which method is appropriate. Some approaches that we believe would work well include:

- Appropriate line fitting: linear, polynomial or other.
- Some form of Fourier analysis.
- Discrete function based on simple statistic such as mean or median.

Once the relationship between α values of each fuzzy set and time have been identified it is trivial to replace the 2-tuple linguistic representation with nonstationary fuzzy sets. The relationship between α and time becomes the perturbation function and the time interval that the rule was mined for becomes the time parameter for the perturbation function. We can now explore reducing the number of rules.

B. Rule Reduction

When the temporal 2-tuple algorithm produces a set of rules it identifies similar rules for different temporal intervals. For example the rule "customers who purchased *lots* of beer also purchased *lots* of pizza" may exist for Friday afternoon and may also exist for Friday evening. However, fuzzy set *lots* for the domain *beer* for the Friday evening rule may well have a higher α value. Using nonstationary fuzzy sets rather than the 2-tuple linguistic representation renders these co-occurring rules redundant: only one occurrence of the rule is required to capture both the linguistic relationship and its variation over time.

We can now give a complete algorithm for mining non-stationary fuzzy association rules as presented in Algorithm 1.

Algorithm 1 Nonstationary Fuzzy Association Rule Mining

Input: A set of data to be mined including timestamps.
Output: A collection of nonstationary fuzzy association rules.
Step 1: Apply temporal 2-tuple algorithm to existing data to produce set of fuzzy association rules R_1.
Step 2: Identify the relationship between time and α for all fuzzy sets contained in R_1 denote as $f : T \times X$.
Step 3: Replace each rule in R_1 with its nonstationary equivalent using f as the perturbation function producing a new set of rules R_2.
Step 4: Remove any duplicate rules from R_2 creating a final rule set R_3.
Step 5: Output R_3.
End

In the next Section we explore a case study of this approach using Web log data as an example.

V. CASE STUDY: WEB USAGE MINING

A Web log dataset has both temporal and quantitative features. The temporal feature is the timestamp of a request made to the server, and the quantitative feature is the Web page view time. The United States Environmental Protection Agency (EPA) dataset[1] is a collection of Hypertext Transfer Protocol (HTTP) requests to a Web server recorded during a 24-hour interval. The geographical location of the Web server is Research Triangle Park, NC, USA. The EPA dataset was recorded from 23:53:25 EDT 29th August 1995 to 23:53:07 30th August 1995. The EPA dataset has 47748 requests: 46014 GET requests, 1622 POST requests, 107 HEAD requests and 6 invalid requests.

The data was preprocessed as follows. Requests with the following suffixes were removed: the following suffixes: gif, xbm, zip, pdf, exe, gz, wpd, wp, dct, jpg, and imf. Maximal forward reference transaction identification [19] was set to 10 minutes to reconstruct a sequence of uniform resource locators (URLs) requested by a user. Sequences of URLs that contain multiple occurrences of the same URL next to each other in the sequence are removed, because these are assumed to be a Web page refresh. Two URLs requests are required to determine the Web page view time of a URL by calculating the difference in timestamps. Association rules have a minimum of two clauses. To ensure sequences of URLs can produce rules of length 2, transactions with 2 or less URLs were removed.

A. Experimental Method

The temporal 2-tuple algorithm produces a set of *n* fuzzy association rules with a variety of of antecedent lengths and 2-tuple linguistic representation of each fuzzy set in the rule. A number of sets of *n* rules may be produced by repeated running of the temporal 2-tuple algorithm with different random seed generator numbers. In this experiment we ran the temporal 2-tuple algorithm with 30 different seeds with each run producing 200 rules with a range of antecedent lengths, thereby completing step 1 of Algorithm 1. We then listed all the 2-tuple α values for each fuzzy set that were mined across the 30 difference seeds along with the timestamp of the data which this rule was generated from. This gives a range of lateral displacement values for each fuzzy set against time. We do this in order to capture the nature of the relationship between time and lateral displacement for each fuzzy set and use this to form nonstationary fuzzy sets. With some fuzzy sets there is a definite relationship between lateral displacement and time, with others the two variables appear to be unlinked. Figures 2 to 6 depicts this data as box plots for the five most frequently occurring fuzzy sets *low* in the domains 6, 90, 73, 131 and 130. This number are simply numeric labels assigned to longer web page URLs for the sake of brevity.

---

[1]Available from The Internet Traffic Archive (http://ita.ee.lbl.gov/)

Fig. 2.   Box Plot of Lateral Displacement against Time for the Domain 6 and the Fuzzy Set *Low*.

Fig. 3.   Box Plot of Lateral Displacement against Time for the Domain 90 and the Fuzzy Set *Low*.
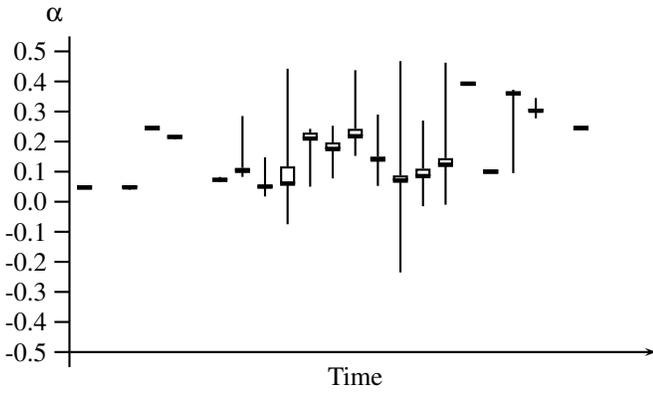

Fig. 4.   Box Plot of Lateral Displacement against Time for the Domain 73 and the Fuzzy Set *Low*.
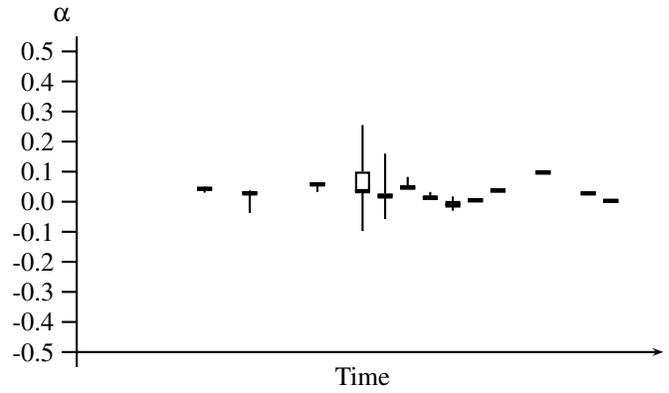

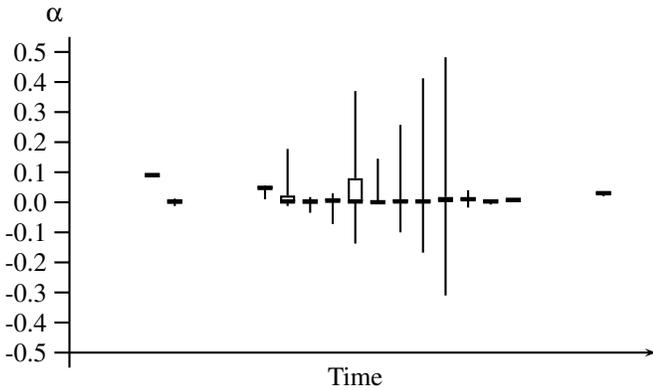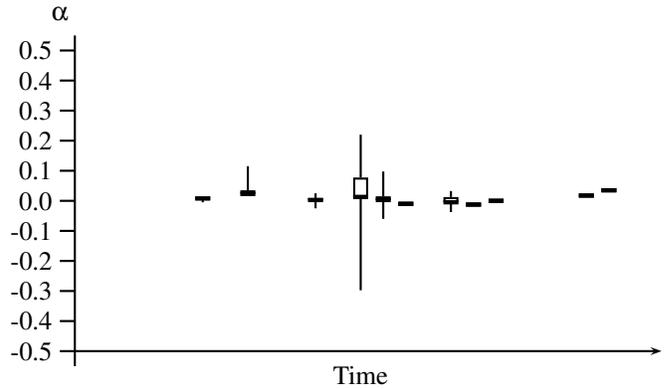Fig. 5.   Box Plot of Lateral Displacement against Time for the Domain 131 and the Fuzzy Set *Low*.
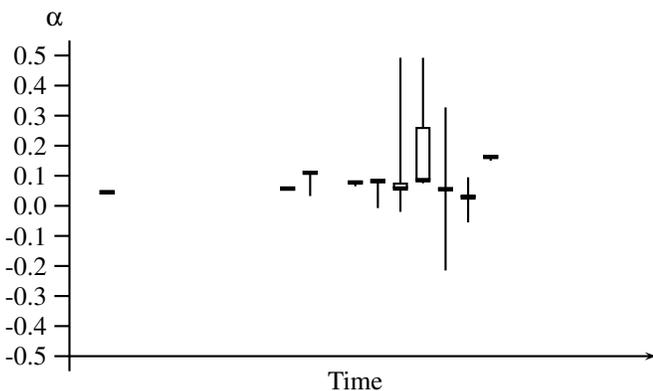

Fig. 6.   Box Plot of Lateral Displacement against Time for the Domain 130 and the Fuzzy Set *Low*.

These plots give a pictorial representation of the relationship between α and time for the five most frequently occurring fuzzy sets. We can see that in some of the fuzzy sets (6, 73, 131) there is some kind of non-linear relationship. Whilst with others (90, 130) there does not appear to be any meaningful relationship between α and time other than a small amount of noise. The source of this noise is unknown, it could be related to the random seeds, a more general feature of the output from the GA, or it could simply be that it is a natural phenomenon perhaps from a range of users using the website for different purposes.

From the data and the box plots produced we decided to use a discrete function to model α over time. We map each point in time to the median value of the α values identified for that set at that point in time by the temporal 2-tuple algorithm. Where there are no α values for a particular time interval the discrete function returns a null value to indicate that the rule does not hold for that point in time. This completes step 2 of Algorithm 1. We complete step 3 by replacing all 2-tuple linguistic representations with nonstationary fuzzy sets. Consider the following four rules:

1) IF 73 is LOW($\alpha = 0.0553$) THEN
   6 is LOW($\alpha = 0.1082$)
2) IF 73 is LOW($\alpha = 0.0558$) THEN
   6 is LOW($\alpha = 0.0415$)

TABLE I

RATES OF RULE REDUCTION ACROSS 30 SEEDS

Seed	No. of Rules Retained	No. of Rules Removed	Rate of Rule Reduction
0	168	32	19.05
1	167	33	19.76
2	171	29	16.96
3	166	34	20.48
4	171	29	16.96
5	166	34	20.48
6	165	35	21.21
7	168	32	19.05
8	161	39	24.22
9	171	29	16.96
10	168	32	19.05
11	164	36	21.95
12	171	29	16.96
13	167	33	19.76
14	169	31	18.34
15	164	36	21.95
16	166	34	20.48
17	167	33	19.76
18	166	34	20.48
19	167	33	19.76
20	167	33	19.76
21	170	30	17.65
22	168	32	19.05
23	166	34	20.48
24	165	35	21.21
25	168	32	19.05
26	167	33	19.76
27	167	33	19.76
28	165	35	21.21
29	163	37	22.70

TABLE II

MEAN JACCARD SIMILARITY ACROSS 30 SEEDS

Seed	Mean Similarity
0	0.9896
1	0.9912
2	0.9894
3	0.9906
4	0.9906
5	0.9892
6	0.9896
7	0.9891
8	0.9898
9	0.9905
10	0.9900
11	0.9895
12	0.9898
13	0.9896
14	0.9903
15	0.9903
16	0.9899
17	0.9904
18	0.9897
19	0.9897
20	0.9911
21	0.9894
22	0.9900
23	0.9902
24	0.9890
25	0.9896
26	0.9903
27	0.9897
28	0.9891
29	0.9900

3) IF 73 is LOW($\alpha = 0.0832$) THEN
    6 is LOW($\alpha = 0.1290$)
4) IF 73 is LOW($\alpha = 0.1113$) THEN
    6 is LOW($\alpha = 0.1670$)

which were all identified by the temporal 2-tuple algorithm in a single run. By replacing the 2-tuple linguistic representation with a nonstationary one we can replace all four of these rules with the following single rule:

1) IF 73 is LOW($f(t)$) THEN
    6 is LOW($g(t)$)

where $f(t)$ and $g(t)$ are functions of the median values of α over time, in this case the centre line of the box plots depicted in Figures 2 and 4 respectively. We now consider the effects of this process on the meaning of the rules.

B. Comparison with Original Rules

There are two questions to explore when comparing the rules across the 2-tuple and nonstationary approaches. Firstly how many rules can be eliminated, and secondly, what difference does this make to knowledge contained in the association rules? We begin by answering the first question. Table I gives the rate of rule reduction across the 30 seeds used in the original temporal 2-tuple experiment. The rate of rule reduction is given by dividing the number of nonstationary rules by the number of 2-tuple rules and multiplying by 100. There is a clear result with this particular example dataset that around 20% of the rules have been eliminated

from the 2-tuple rule set. It is worth noting that these are the rules where the same relationship keeps reappearing in a slightly different way across a range of time intervals. It could well be the case that the rules which are replaced are actually the most important rules as they reoccur across a range of time intervals. If this is the case then this approach may lead to a method for prioritising the rule set.

We now consider how the meaning of these rules has changed. To do this we calculate the similarity between the sets which make up the two rule sets using the Jaccard similarity measure given by:

$$s(A,B) = \frac{\sum_{i=1}^{n} min(\mu_A(x_i), \mu_B(x_i))}{\sum_{i=1}^{n} max(\mu_A(x_i), \mu_B(x_i))} \qquad (2)$$

We only consider the five most frequently occurring fuzzy sets in the rule set as we expect that these are the sets where the greatest dissimilarity will be. Table II gives the mean similarity for the five most frequently occurring fuzzy sets across the 30 seed values. Observe that the 2-tuple and nonstationary fuzzy sets have high degrees of similarity indicating little difference between the two rule sets.

VI. DISCUSSION

The results given in Tables I and II demonstrate that the proposed approach performs the desired function of reducing the size of the rule set. In the example given, around 1 in 5 of all rules were removed. Although this is a useful level of rule reduction it should be noted that the level of rule reduction is

entirely dependent of the rules mined by the temporal 2-tuple algorithm and with only on data explored so far we can not say whether this level of reduction is typical or a-typical.

The high level of similarity between the temporal 2-tuple rule set and the nonstationary rule set gives confidence that, despite 1 in 5 of the rule being eliminate, the knowledge contained in the rule set is maintained.

VII. CONCLUSION

In this paper we have presented a method for using nonstationary fuzzy sets to improve the tractability of a mined rule set. In a case study we have shown that 20% of the original rules could be removed and that the removal of these rules has little affect on the knowledge contained in the rule set.

We have shown that nonstationary fuzzy sets are a powerful tool for modelling fuzzy associative relationships that have a temporal component. We intend to continue working with nonstationary fuzzy sets and in particular look at a data mining algorithm which only makes use of nonstationary fuzzy sets without the need for the intermediate stage of the 2-tuple linguistic representation.

REFERENCES

[1] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers, "Big data: The next frontier for innovation, competition, and productivity," McKinsey Global Institute, Tech. Rep., May 2011.

[2] R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," in *Proceedings of ACM SIGMOD International Conference on Management of Data, Washington, DC, USA*, 1993, pp. 207–216.

[3] S. G. Matthews, M. A. Gongora, A. A. Hopgood, and S. Ahmadi, "Temporal fuzzy association rule mining with 2-tuple linguistic representation," in *Proceedings of the 2012 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2012), Brisbane, Australia*, 2012, pp. 1–8.

[4] T.-P. Hong, C.-S. Kuo, and S.-C. Chi, "Trade-off between computation time and number of rules for fuzzy mining from quantitative data," *International Journal of Uncertainty, Fuzziness & Knowledge-Based Systems*, vol. 9, no. 5, pp. 587–604, 2001.

[5] F. Herrera and L. Martínez, "A 2-tuple fuzzy linguistic representation model for computing with words," *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 6, pp. 746–752, 2000.

[6] J. Alcalá-Fdez, R. Alcalá, M. J. Gacto, and F. Herrera, "Learning the membership function contexts for mining fuzzy association rules by using genetic algorithms," *Fuzzy Sets and Systems*, vol. 160, no. 7, pp. 905–921, 2009.

[7] J. M. Ale and G. H. Rossi, "An approach to discovering temporal association rules," in *Proceedings of the 2000 ACM Symposium on Applied computing, Como, Italy*, 2000, pp. 294–300.

[8] D. Leonard, "After katrina: Crisis management, the only lifeline was the wal-mart," *FORTUNE Magazine*, no. October 3, 2005.

[9] B. Saleh and F. Masseglia, "Discovering frequent behaviors: time is an essential element of the context," *Knowledge and Information Systems*, vol. 28, pp. 1–21, 2010.

[10] J. H. Holland, *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, MI, 1975.

[11] A. González and F. Herrera, "Multi-stage genetic fuzzy systems based on the iterative rule learning approach," *Mathware & Soft Computing*, vol. 4, no. 3, pp. 233–249, 1997.

[12] J. M. Garibaldi, M. Jaroszewski, and S. Musikasuwan, "New concepts related to non-stationary fuzzy sets," in *Proceedings of the 2007 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2007), London, UK*, 2007, pp. 1–6.

[13] J. Garibaldi and M. Jaroszewski, "Generalisations of the concept of a non-stationary fuzzy set - a starting point to a formal discussion," in *Proceedings of the 2008 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2008), Hong Kong, China*, 2008, pp. 2313–2319.

[14] J. Garibaldi, M. Jaroszewski, and S. Musikasuwan, "Nonstationary Fuzzy Sets," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 4, pp. 1072–1086, August 2008.

[15] J. Xiao-Ying Wang andGaribaldi, S.-M. Zhou, and R. John, "Methods of interpretation of a non-stationary fuzzy system for the treatment of breast cancer," in *Proceedings of the 2009 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2009), Jeju Island, Korea*, 2009.

[16] Benatar, N. and Aickelin, U. and Garibaldi, J.M., "A comparison of non-stationary, type-2 and dual surface fuzzy control," in *IEEE International Fuzzy Systems Conference, 2011. FUZZ-IEEE 2011.*, 2011, pp. 1193–1200.

[17] Duc Thang Ho and Garibaldi, J.M., "A novel fuzzy inferencing methodology for simulated car racing," in *IEEE International Fuzzy Systems Conference, 2008. FUZZ-IEEE 2008.*, 2008, pp. 1907–1914.

[18] Musikasuwan, S. and Garibaldi, J.M., "On Relationships Between Primary Membership Functions and Output Uncertainties in Interval Type-2 and Non-Stationary Fuzzy Sets," in *Proceedings of the 2006 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2006), St. Louis, Missouri, USA*, 2006, pp. 1433–1440.

[19] R. Cooley, B. Mobasher, and J. Srivastava, "Grouping web page references into transactions for mining world wide web browsing patterns," in *Proceedings of the 1997 IEEE Knowledge and Data Engineering Exchange Workshop, Newport Beach, CA, USA*, 1997, pp. 2–9.