Fuzzy Genetic Algorithms for Pairs Mining*

Longbing Cao, Dan Luo, and Chengqi Zhang

Faculty of Information Technology, University of Technology Sydney, Australia
{lbcao, dluo, chengqi}@it.uts.edu.au

Abstract. Pairs mining targets to mine pairs relationship between entities such as between stocks and markets in financial data mining. It has emerged as a kind of promising data mining applications. Due to practical complexities in the real-world pairs mining such as mining high dimensional data and considering user preference, it is challenging to mine pairs of interest to traders in business situations. This paper presents fuzzy genetic algorithms to deal with these issues. We introduce a fuzzy genetic algorithm framework to mine pairs relationship, and propose strategies for the fuzzy aggregation and ranking of identified pairs to generate final optimum pairs for decision making. The proposed approaches are illustrated through mining stock pairs and stock-trading rule pairs in stock market. The performance shows that the proposed approach is promising for mining pairs helpful for real trading decision making.

1 Introduction

Pairs mining aims to mine pair relationships between a couple of instances. A *pair relationship* represents that two instances are correlated or associated in terms of some form of statistic or probabilistic measures. Pairs identified, either intra class or inter class, may indicate dynamics of interest to users. For instance, in stock data mining [1,6,7], pairs mining can discover stock pairs linked by correlation relationship from a series of stocks in the market. Pairs mining may also find pair relationships between stocks and derivatives (namely stock-derivative pairs) lodged in an exchange or of the same listed companies. These pairs identified are helpful for traders to make smart decisions. For example, pairs trading strategy can be designed to trade a basket of stocks to distribute potential trading and investment risk rather than putting all money on one instrument.

Real-world pairs mining is quite complicated because the pairs are hidden in high dimensional data and must satisfy uncertain user requests. For instance, stock pairs may hide in all combinations of stocks validly listed in an exchange, while the number of listed stocks can be over 1,000. This may lead to big computational cost. More challenging issue is to mine pairs showing the correlation between stocks and trading rules in a market. On the other hand, user preference and business needs are the drivers to enhance pairs of interest to real user needs [3]. Therefore, it is important to tackle the above factors in identifying and evaluating pairs.

* This work is sponsored by Australian Research Council Discovery Grant (DP0667060, DP0449535), China Overseas Outstanding Talent Research Program of Chinese Academy of Sciences (06S3011S01), and UTS ECRG and Chancellor grants.

Q. Yang and G. Webb (Eds.): PRICAI 2006, LNAI 4099, pp. 711 – 720, 2006.

Correlation [5] and association mining [5] can be used to mine pair relationships. However, the problem for them to mine pairs from the above real-world situations is the ineffectivity in mining pairs from high dimensional data and considering business preference into the mining process. To efficiently address high dimensional data, genetic algorithms [4] are widely used. However, genetic algorithms are not good at dealing with domain-oriented business requests and user preference.

This paper investigates pairs mining in high dimensional data and by involving user preference. The main contribution of the paper is that it proposes fuzzy genetic algorithms by integrating fuzzy set [8] and genetic algorithms, and fuzzy aggregation and ranking mechanisms for generating optimum pair set for final decision support. Correlation analysis techniques are used as a pair selection method and merged into fuzzy genetic algorithms.

The remaining sections are organized as follows. Section 2 introduces pairs mining. Section 3 presents a framework for fuzzy genetic algorithms and fuzzy aggregation and ranking of mined pairs. We illustrate fuzzy genetic algorithm-based rule-stock pairs mining in stock market in Section 4. Performance evaluation is presented in Section 5. Section 6 concludes this work.

2 Pairs Mining

Pairs mining targets mining *pair relationships* existing in a pair of instances of an attribute, or between a pair or a number of attributes. Pair relationship indicates that there are two instances that are highly correlated or associated in terms of some form of probabilistic or statistic metrics. For instance, in stock data mining, pair relationships between stocks can be analyzed according to the coefficient of correlation of stock prices. We call these pairs of stocks as *stock pairs*. We can also mine pair relationships between stocks and markets, namely stock-market pairs, based on the correlation between stock price and market index.

Pairs coming from the same class are called *kindred* or *homogeneous* pairs, for instance, stock pairs and market pairs in stock data mining. On the other hand, those come from different families are *alien* or *heterogeneous* pairs such as stock-market pairs and stock-trading rule pairs. Both kindred and alien pairs are noteworthy, they are exchangeable in terms of some conditions. For instance, normally the pair relationship among stocks is kindred. While they may present as alien pairs if the two parties come from different sectors and we want to highlight the significance of varying sectors in assessing the pair relationship.

Furthermore, pairs may be linked together in either positive or negative pair relationships. Positive pair relationships mean that the two parties are linked through certain obverse relation such that both of them follow similar patterns. In correlation-based pairs mining, pairs represented by positive correlation coefficients follow obverse change trends. For instance, in stock pair mining, some stocks are found to be positively linked with some others when both prices go up following similar patterns.

Negative pair relationships indicate that two parties are coupled in opposite trends or patterns. In correlation-based pairs mining, if the correlation coefficient is negative, then one goes up while another goes towards the reverse direction. For instance, in stock pairs identified, the trend of price dynamics of some stocks are found to be

opposite to its paired partners. Another interesting type of negative pairs exist in terms of negative association rules. In this case, one item presented indicates that another definitely will not come up, which is against commonsense.

Correlation mining and association mining can be used for the above pairs mining. However, they are not good at handling high dimensional data and user preference. To tackle these issues, we present an effective framework of fuzzy genetic algorithms.

3 Fuzzy Genetic Algorithms

This section introduces a framework of fuzzy genetic algorithms. We also introduce the mechanisms for aggregating and ranking pairs to generate final workable pairs.

3.1 A Fuzzy Genetic Algorithm Framework

A fuzzy genetic algorithm [2] is a fuzzy set-coded genetic algorithm where each individual (chromosome) is composed of a set of membership functions. The principle of fuzzy optimization is as follows. Suppose we have a membership function F so that $\bar{Y} = F(\bar{X})$, where $\bar{X}$ is fuzzy set-based input, fuzzy set $\bar{Y}$ is the output from F given $\bar{X}$. Fuzzy optimization is to find a proper $\bar{X}$ in its valid value range to "maximize" the fuzzy set $\bar{Y}$. This is achieved through maximizing a suitable mapped measure of $\bar{Y}$, for instance the center of gravity centroid($\bar{Y}$). Here the measure can map the fuzzy sets into real numbers.

In designing fuzzy genetic algorithms, issues in conventional genetic algorithms are put into fuzzy context and converted into fuzzy versions, for instance, fuzzy representation, fuzzy genetic operators, etc. In particular, fuzzy genetic algorithms need to consider the validation and ranking of the created fuzzy sets. In mining pairs in financial markets, we develop the following fuzzy genetic algorithm framework.

ALGORITHM 1. pseudo code for fuzzy genetic algorithm
Input: real number set X
Output: optimal fuzzy set Y for decision support
Procedure: FGA($\mu, X(t), \bar{X}(t), \bar{X}'(t), Y$)

```
//start with an initial time
t := 0;
//initialize a fuzzy random population of individuals X̄(t) by fuzzifying the real number
sets X(t) with proper membership functions μ_X̄,
initialize X̄(t) = {(x, μ_X̄(x)) | x ∈ X(t), μ_X̄ : X(t) → [0,1]};
//evaluate the fitness of all initial individuals of population based on fuzzy evaluation
evaluate X̄(t);
//test for termination criterion
While (not done) do
    //increase the time counter
    t := t + 1;
    //select a fuzzy sub-population set X̄'(t) for offspring production
    X̄'(t) := select X̄(t);
    //crossover the "genes" of the selected parents X̄'(t)
```

```
        crossover X̄'(t);
        //perturb the mated population stochastically
        mutate X̄'(t);
        //fuzzily evaluate its new fitness
        evaluate X̄'(t);
        //select the survivors Ȳ from actual fitness
        Ȳ := survive X̄(t), X̄'(t);
    End
  //fuzzily rank the survivors
  rank Ȳ;
  //defuzzify and export the final survivors
  export Y;
```

The above fuzzy genetic algorithm framework deals with all basic issues such as initialization, selection, crossover, mutation and evaluation of stock pairs mining in fuzzy context. For initialization, all individuals are sampled randomly within the valid domain. In stock pair mining, we identify the business interestingness measure *sharpe ratio* as the fitness function for all individuals in the pair population. Real coded sharpe ratio measures the performance of a pair in terms of both return and risk. If sharpe ratio is high, the rule leads to high return with low risk. The following defines real number sharpe ratio *SR* used in stock market.

$SR = (R_p - R_f)/\sigma_p$ Where R_p is expected portfolio return, R_f is risk free rate, σ_p is portfolio standard deviation.

We fuzzify the real coded *SR* into the interval [0,1] to get its fuzzy sets $\overline{SR}$. We use triangle piecewise linear membership function to fuzzify the universal sets. For instance, we specify ten levels of linguistic values, namely 1st, 2nd, …, 10th from the lowest to the highest, for the fuzzy linguistic variable *sharpe ratio* $\overline{SR}$. Hereby we generate top $\overline{N}$ target objects, for instance the corresponding trading rules, $\overline{N}$ refers to those rules which correspond to the first *N* highest linguistic values.

In our case, the genetic algorithms for pairs mining are real coded, therefore its crossover can be in an arithmetic and/or multiple-point manner. We provide multi-point arbitrary crossover in a shuffling probability $p(0 \le p \le 1)$ of alleles on top of the top $\overline{N}$ selected sub-populations. Suppose γ is a random number in [0,1], the following illustrates the shuffling situation if $\gamma \le p$.

$$\overline{X}_1' = (x_{10}, ..., x_{2i}, x_{1(i+1)}, ..., x_{2k}, ...)$$
$$\overline{X}_2' = (x_{20}, ..., x_{1i}, x_{2(i+1)}, ..., x_{1k}, ...)$$

On the other hand, the mutation is based on changing the original value stochastically by the mutation rate q $(0 \le q \le 1)$ either positively or negatively. Our strategy for the mutation is to conduct the mutation operation on top of the shuffled sets $\overline{X}'(t)$ with the rate q around 0.03.

Based on the above operations, a collection of optimized individuals emerge from the candidate population. They are possible optimal candidates for the final recomm - endation. In order to generate the final optimal list, special attention should be paid to the aggregation, evaluation and ranking of fuzzy functions and fuzzy sets.

3.2 Fuzzy Evaluation, Aggregation and Ranking

Fuzzy evaluation refers to evaluate the validity of fitness functions, fuzzy sets and individuals. The validity is directly induced and embodied through membership functions. Therefore, it is essential to check the validity of membership functions and fitness functions. Taking the fuzzification of sharpe ratio as an instance, its real values can be negative or positive in the real number field. However, in fuzzy genetic algorithms, the universe of the discourse of the fuzzified sharpe ratio $\overline{SR}$ must be in [0,1]. Otherwise, they are abnormal and invalid, and we need to normalize membership grades into the interval of [0,1]. We also check the monotone of $\overline{SR}$ membership grades to guarantee that only one grade exists for each $\overline{SR}$ element, which is mapped to single real value of sharpe ratio. If it is not monotonic, corresponding strategy should be taken to monotonize it. Based on these policies, we can check the validity of individuals and the fuzzy sets.

The output of fuzzy genetic algorithms is to recommend a set of optimal individuals *Y*. To this end, fuzzy aggregation and fuzzy ranking play an important role in generating the final survivors. In the following paragraphs, we illustrate the ranking strategy for discovering trading rule-stock pairs. There are three steps for us to find out the actionable trading rules highly correlated to given stocks. One of the steps is to mine and rank the in-depth trading rules [7] for a specific stock. Another step is to detect and order the very appropriate stocks for a given trading rule. Then we aggregate these two lists through fuzzy aggregation rules to obtain a set of composite optimal stock-trading rule pairs. Finally, we fuzzily rank the trading rule-stock pairs, and further defuzzify them to generate final outputs. See Section 4 for more details about discovering rule-stock pairs.

Let *SR* be fitness function for the above first two steps. Suppose we build ten ascending linguistic values from 1^{st} to 10^{th}. To distinguish the two cases, as illustrated in Figure 1 and 2, we use fuzzy linguistic terms *a* to *j* and fuzzy values *A* to *J* to label the fuzzy sets for the optimal trading rules given a specific stock (we called rule sets), and for the appropriate stocks given a trading rule (called stock sets), respectively.

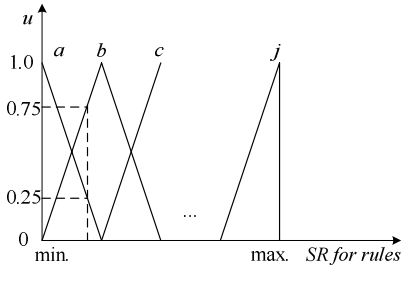

Fig. 1. Fuzzy set for trading rules given stock

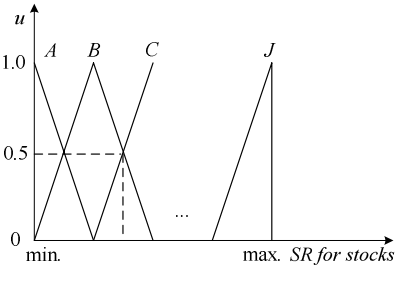

Fig. 2. Fuzzy set for stocks given trading rule

In practice, even though sharpe ratio is used as the fitness and similar linguistic measures are used for both rule set and stock set situations, the meaning of a specific corresponding linguistic term, say *b* and *B* in this case, may be highly varying. This

means that we cannot aggregate the stock-rule pairs based on the equal matching of two linguistic values from different sets. Instead, we develop the following solution to aggregate the two fuzzy groups.

To aggregate the two groups, we set up another fuzzy variable called *rule-stock pair* (in short *pair*) to correlate close partners between the rule set and the stock set. The *pair* has 19 linguistic values ascending from 1^{st}, 2^{nd} to 19^{th}. Figure 3 defines its triangle fuzzy sets. Further, the following fuzzy aggregation rules are defined to merge the fuzzy sets from different groups.

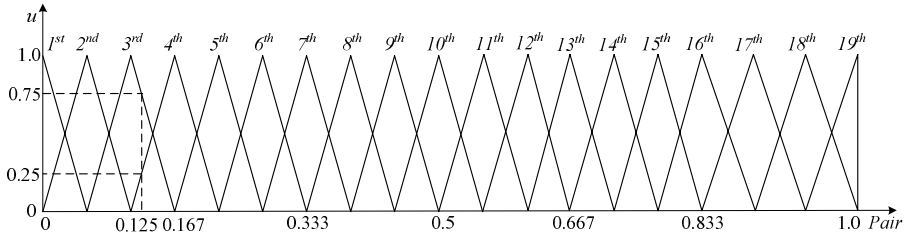

Fig. 3. Fuzzy set for trading rule-stock pairs

DEFINITION 1. *(Fuzzy aggregation rule) if the fuzzy rule set is m-th, and the fuzzy stock set is n-th, then the rule-stock pair is (m+n-1)-th.*

For instance, if the rule set is *c* (i.e., 3^{rd}), and the stock rule is *d* (4^{th}), then the rule-stock pair is ranked as 6^{th} (3+4-1). Table 1 illustrates the fuzzy aggregation and ranking of all linguistic values from the rule set and the stock set, respectively. The fuzzy rule makes it possible to integrate the rule set and the stock set, and output the higher ranked rule-stock pairs as final survivors for optimal decision-making.

Table 1. Fuzzy aggregation and ranking of rule sets and stock sets

	a	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*	*j*
A	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}	8^{th}	9^{th}	10^{th}
B	2^{nd}	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}	8^{th}	9^{th}	10^{th}	11^{th}
C	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}	8^{th}	9^{th}	10^{th}	11^{th}	12^{th}
D	4^{th}	5^{th}	6^{th}	7^{th}	8^{th}	9^{th}	10^{th}	11^{th}	12^{th}	13^{th}
E	5^{th}	6^{th}	7^{th}	8^{th}	9^{th}	10^{th}	11^{th}	12^{th}	13^{th}	14^{th}
F	6^{th}	7^{th}	8^{th}	9^{th}	10^{th}	11^{th}	12^{th}	13^{th}	14^{th}	15^{th}
G	7^{th}	8^{th}	9^{th}	10^{th}	11^{th}	12^{th}	13^{th}	14^{th}	15^{th}	16^{th}
H	8^{th}	9^{th}	10^{th}	11^{th}	12^{th}	13^{th}	14^{th}	15^{th}	16^{th}	17^{th}
I	9^{th}	10^{th}	11^{th}	12^{th}	13^{th}	14^{th}	15^{th}	16^{th}	17^{th}	18^{th}
J	10^{th}	11^{th}	12^{th}	13^{th}	14^{th}	15^{th}	16^{th}	17^{th}	18^{th}	19^{th}

However, the above aggregation and ranking strategy is based on the fuzzification of fitness and membership functions. Therefore, a rule in fuzzy set *c* or a stock in fuzzy set *D* is basically from fuzzy rather than crisp perspective. For instance, as shown in Figure 1 and 2, a trading rule could be classified into fuzzy set *b* with a

membership grade μ=0.75 or set *a* with the grade μ=0.25. Similar thing exists for stock set, a stock could be segmented into fuzzy set *B* or *C* with the same grade μ=0.5. In this case, the outcome of the fuzzy aggregation and ranking could have four options, namely 2^{nd}, 3^{rd}, 4^{th} or 5^{th}.

To manage the above uncertain situations in the fuzzy aggregation and ranking, a ranking coefficient ρ based on moment defuzzification is introduced. It defuzzifies a fuzzy set returning a floating point that represents the fuzzy set. It actually measures how optimal a pair is.

$$\rho = \sum_{l=1}^{m} \eta_l \mu_l^R \mu_l^S \Big/ \sum_{l=1}^{m} \mu_l^R \mu_l^S$$

Where, m refers to the number of triggered linguistic values, l =1, 2, …, m corresponds to each triggered linguistic value. μ_l^R is the membership grade of No. l linguistic term relevant to the sharpe ratio of a rule. μ_l^S is the membership grade of No. l linguistic term corresponding to the sharpe ratio of a stock. η_l is the centroid of the No. l triggered linguistic value, it is calculated in terms of the moment and the area of each subdivision.

The ranking coefficient ρ provides a solution to deal with possible uncertainty when a rule-stock pair is aggregated. A real number can be obtained to measure a fuzzy rule-stock pair in a relatively crisp manner. For instance, we can calculate and get ρ =0.125 in the above example. As shown in Figure 3, this clearly indicates that this rule-stock pair is ranked as 3^{rd} fuzzy set since its membership grade is 0.75 much larger than fuzzy set 4^{th} with grade 0.25.

4 Mining Rule-Stock Pairs

We instantiate the above fuzzy genetic algorithm framework to mine financial pairs such as stock pairs and rule-stock pairs [7]. Due to space limitation, here we only introduce the rule-stock pairs mining through analyzing the correlation between trading rules and tradable stocks. In market trading, some trading rules are tested more profitable to trade a class of stocks, while others are more suitable for other stocks. Using pairs mining we may evidence whether there exists pair relationship between trading rules and stocks or not. The above fuzzy genetic algorithm framework is used to develop algorithms for discovering the target rule-stock pairs.

In identifying rule-stock pairs which can be used for trading support, traders are invited to give suggestions on designing features, interestingness measures and parameter optimization strategy. They also helped us construct interestingness metrics and evaluate and refine rule-stock pairs. Taking the Australian Stock eXchange (ASX) as an instance, six types of trading rules such as Channel Breakout and 27 ASX stocks such as ANZ are chosen for the experiments in the orderbook data. Five different investment plans are used to trade the above identified rule-stock pairs by enhancing trading rules via considering and trading the correlated stocks. The following introduces the method for rule-stock pairs mining algorithms.

ALGORITHM 2. Method for improving trading rules by analyzing correlation between rules and stocks

Input: a set of historical intraday orderbook transactions T, a set of trading rules R, a set of stocks S, a coefficient threshold $coeff_0$, a sharpe ratio threshold sr_0, a return threshold r_0 ,

Output: Fuzzily ranked trading rule-stock pairs

Method:

1. Given a stock S_i, and a type of trading rule R_j, mining actionable rules r_{ijm} (m = 0,1, …) for the stock in T using fuzzy genetic algorithms described in ALGORITHM 1 and 4;
2. Mining all actionable rules r_{ijm} (i =0,1,…, j =0,1,…; m = 0,1, …) for all stocks S_i (i =0,1,…) and all types of trading rules R_j (j =0,1,…);
3. Fuzzily aggregating the rule set r_{ijm} to generate a fuzzily optimal rule for a given stock;
4. Generating fuzzy optimal rule-stock pairs s_i -r_j (i =0,1,…, j =0,1,…);
5. Evaluating the rule-stock pairs s_i -r_j by involving traders' concerns;
6. Fuzzily ranking the rule-stock pairs s_i -r_j ;
7. Exporting classified rule-stock pairs in terms of user acceptable linguistics.

Figure 4 illustrates an excerpt of fuzzily ranked rule-stock pairs discovered in ASX intraday orderbook data in June 2001. Three types of rules (coded as 1 to 3) and 27 ASX stocks (coded from 1 to 27) are selected for this experiment. The total 81 rule-stock pairs are classified into five fuzzy groups in terms of sharpe ratio: VP-very positive, P-positive, W-watchable, N-negative and VN-very negative. We can see that Rule 2-Stock 24, Rule 2-Stock 26, and Rule 1-Stock 24 are three *very positive* pairs in June 2001. These high-ranking pairs may be useful to support trading.

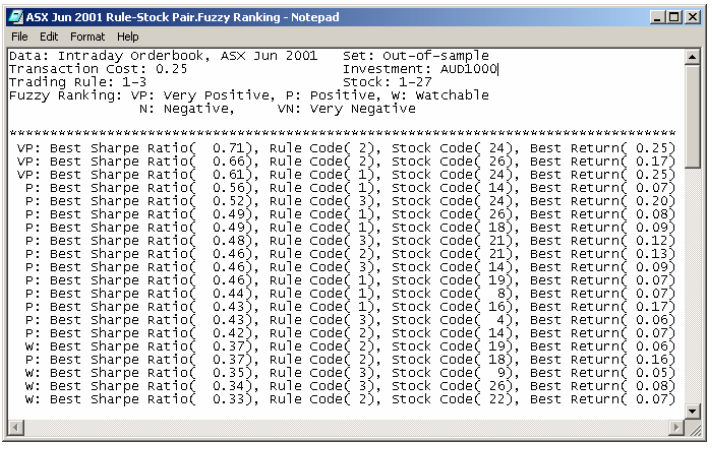

Fig. 4. Fuzzily ranked trading rule-stock pairs

5 Performance Evaluation

In real-world mining, performance evaluation not only demonstrates the advantage of a specific data mining algorithm but also justifies whether the developed approach can satisfy real user needs or not, besides. Therefore, the algorithm should satisfy both technical interestingness and business interestingness [3].

Technically, we focus on predictability and actionability which can be instantiated into different metrics in terms of particular domain problem. Let D be the number of total pairs (e.g., rule-stock pairs) found in in-sample and out-of-sample sets satisfying certain business interestingness (say return must be larger than a threshold R_0). A (B) be the number of total pairs in in-sample (out-of-sample) data set. AB be the number of pairs existing in both in-sample and out-of-sample sets, which satisfy business interestingness request. Then, the following statistics measures the actionability of trading rule-stock pairs in terms of in-sample and out-of-sample comparison.

DEFINITION 2. *(Probability of rule-stock pairs) The following probability functions are defined for rule-stock pairs in or across in-sample and out-of-sample data sets: P(A)=A/D, P(B)=B/D, P(AB)=AB/D, P(A|B)=P(AB)/P(B), P(B|A)=P(AB)/P(A).*

Example 1. In the rule-stock pair mining, 10 pairs are ranked as class 19th in in-sample data, which are the most promising pairs. While 7 of them also satisfy the same condition in out-of-sample set, another 2 new pairs jump into the class 19th in testing. Then *P(A)=83%, P(B)=75%, P(AB)=58%, P(A|B)=78%, P(B|A)=70%.*

DEFINITION 3. *(Actionability metrics of rule-stock pairs) The following metrics*: *Confidence, All_Confidence, Cosine and Coherence are defined for measuring the actionability of the trained rule-stock pairs in in-sample set when they are tested in out-of-sample data.*

$$Confidence = \max(P(A \mid B), P(B \mid A))$$

$$All\_Confidence = P(AB) / \max(P(A), P(B))$$

$$Cosine = P(AB) / \sqrt{P(A)P(B)}$$

$$Coherence = P(AB) / (P(A) + P(B) - P(AB))$$

In general, the value range of all the above metrics is in the interval of [0, 1]. Larger value donates bigger actionability of pairs when they are deployed into the real trading. We count the summarized statistics of actionability for the discovered rule-stock pairs from April to October 2001 using ASX intraday orderbook data using sliding window strategy for training and testing.

Table 2 lists the coherence and confidence of top 10% pairs (in this case, *Confidence* = *All_Confidence* = *Consine*). In this top 10% pairs, we find 11 pairs actionable in out-of-sample set. Among these pairs, two of them are relatively frequent pairs: Rule 1 – Stock 14, Rule 2 – Stock 24 (in this particular pair set, their association supports are larger than 20%).

Table 2. Actionability of top 10% rule-stock pairs

	Apr	May	Jun	Jul	Aug	Sept	Oct
Coherence	6.7%	14.3%	23%	14.3%	6.7%	14.3%	23%
Confidence	12.5%	25%	37.5%	25%	12.5%	25%	37.5%

In addition, we calculate the statistics in terms of fuzzy percentile ranking. For all pairs either in in-sample set or in out-of-sample set, we rank them in terms of five linguistic levels: *very positive*, *positive*, *watchable*, *negative* and *very negative* using asymmetry triangle membership function. We prune pair items at the point where more than 20% in-sample pairs (choose 5% in-sample pairs in this example) have presented in out-of-sample set. Table 3 provides the summarized statistics for the

same results. The above performance analysis shows that the confidence of fuzzy percentile ranked rule-stock pairs seems better than non-fuzzy top x% pairs and crisp threshold based pairs. In this case, 9 actionable pairs are found, while the following pairs are found frequent: Rule 1-Stock 14, Rule 1-Stock 24, Rule 2-Stock 14, Rule 2-Stock 24 (association support ≥ 20%).

Table 3. Actionability of fuzzy percentile ranked rule-stock pairs

	Apr	May	Jun	Jul	Aug	Sept	Oct
Coherence	20%	13.3%	37.5%	22.2%	13.3%	10%	9.1%
Confidence	50%	50%	75%	50%	50%	25%	25%
All_Confidence	40%	15.4%	42.9%	28.6%	15.4%	14.3%	12.5%
Consine	35.4%	27.7%	56.7%	37.8%	27.7%	18.9%	17.7%

However, our field studies in the market show that in real-world stock mining it is very hard to get predictability rate as high (say 80%) as reported in KDD literature. And even difficult thing is that it is very time-consuming and costly to develop a trading rule-stock pair which is workable when deployed into real market.

6 Conclusions

Pairs mining is emerging as a kind of very useful data mining applications, which can discover interesting pair relationship between a pair of instances in one or many classes. Real-world pairs mining must identify pairs in high dimensional data and consider user preference. This brings challenge to the existing association rule mining, correlation mining and genetic algorithms. This paper has proposed fuzzy genetic algorithms to tackle real-world pairs mining. A fuzzy genetic algorithm framework has been designed, which integrates fuzzy set and genetic algorithms and embeds user preference by developing fuzzy aggregation and ranking rules. They are used for mining financial pairs in stock markets. The financial pairs mining in ASX orderbook data has shown that the proposed approach is promising for detecting interesting pairs in high dimensional data and considering user preference.

References

[1] Allen, F. and Karjalainen, R. Using genetic algorithms to find technical trading rules, *Journal of Financial Economics*, 51, 245-271, 1999.

[2] Buckley, J. Hayashi, Y. Fuzzy genetic algorithm and applications, *Fuzzy Sets and Systems* 61: 129-136, 1994.

[3] Cao, L., Zhang, C. Domain-driven actionable knowledge discovery in the real world, *PAKDD2006, LNAI 3918*, 821 – 830, 2006.

[4] Davis, L. (Ed.). *Handbook of genetic algorithms*. New York: Van Nostrand Reinhold, 1991

[5] Han, J W., Kamber, M.: Data Mining: Technologies, Techniques, Tools, and Trends. Morgan Kaufmann Publishers.

[6] Kovalerchuk, B., & Vityaev, E. *Data Mining in Finance: Advances in Relational and Hybrid Methods*, Kluwer Academic, 2000.

[7] Lin, L., Cao, L. Mining In-Depth Patterns in Stock Market*, Int. J. Intelligent System Technologies and Applications,* 2006.

[8] Zadeh, L.A. Fuzzy sets. *Information and Control*, 83: 338–353, 1965.