3.2.5 Prediction

One of the more complex tasks, prediction often involves deriving patterns from a training set, thus building a model of the population's behavior that incorporates predictor variables and dependable variables. The model is then fed with new predictor data to produce estimations of the dependable variables.
In the strict sense of the word predicting would involve a time difference between the predictor variables (now or in the past) and the dependable variables (in future). However, in statistics or data mining 'predicting' is also often used for the estimation of unknown dependable variables at present or in the past. Our first example is based on the known responses of a test set of customers and tries to predict the set of prospects who are most likely to buy a caravan insurance policy. The next example is of a more complex nature, the authors investigate the existence of predictable sequences of events in the financial markets.

Analytical customer relationship management for insurance policy prospects

Peter van der Putten[1]

In marketing there are two opposed approaches to communication: mass marketing and direct marketing. In mass marketing, a single communication message is broadcast to all potential customers through media such as print, radio or television. Such an approach implies high wastage: only a small proportion of the prospects will actually buy the product. As competition increases and markets get more fragmented the problem of waste worsens. Moreover, in spite of huge investments in market research and media planning, it is still hard to quantify the benefits of mass marketing.

These developments have led to an increased popularity of direct marketing, especially in the sectors of finance, insurance and telecommunication. The ultimate goal of direct marketing is cost-effective, two-way, one-to-one communication with individual prospects. For this it is essential to learn present and predict future customer preferences. In today's chaotic business environment, customer preferences change dynamically and are too complex to be derived straightforwardly.

Continuous mining of customer behavior patterns may offer a flexible solution to this problem [Putten, 1999]. The classical application of data mining for direct marketing is response modeling. Usually, the relative number of customers that

1 Drs P. van der Putten, pvdputten@hotmail.com. pvdputten@liacs.nl, Leiden Institute of Advanced Computer Science, Leiden University, Leiden, The Netherlands

respond to direct mail is very low (5% or less). Predictive models can be built to identify the prospects most likely to respond. Historical data on previous mailings or product ownership are used to construct the model. The resulting model can be applied to filter prospects from the existing customer base or from address lists acquired from list brokers. We present a case from insurance marketing to illustrate this approach.

Business problem

The business objective in the insurance case was to expand the market for an existing consumer product, a caravan insurance, with only moderate cost investment. Data mining analysis should answer the following question: can we predict who would be interested in buying a caravan insurance policy and explain why?

Actually, this real world business case was re-used for the CoIL Challenge 2000, a data mining competition organized by CoIL, the European Network of Excellence for Computational Intelligence and Learning [Putten, 2000] [2]. The problem description and data were posted on the web and participants had a little more than one month to send in results. In the remainder of this article we will focus mainly on the challenge results for the predictive data mining task.

Data mining solutions

The data used in the case were very similar to data in other direct marketing projects for other financial clients. Each customer was characterized by a selection of 95 attributes. The attributes could be divided in two groups. The product usage attributes defined the product portfolio of an individual customer, consequently these attributes can be considered internal (company owned), behavioral attributes. We also purchased sociodemographic survey data that had been collected on zip code level. All customers belonging to the same zip code area have the same value for these attributes. This data included information on education, religion, marital status, profession, social class, house ownership and income. Most participants performed extensive data preparation to get the best input for the models, including using derived attributes, recoding attributes and selecting attributes.

To select prospects a model had to be constructed to predict the attribute 'owns a caravan policy', given all other attributes. For the challenge a random sample, the training set, was drawn from the customer base. This set was used to construct the models. The training set contains over 5,000 descriptions of customers, including the information whether or not they have a caravan insurance policy. Furthermore, there was a test set containing 4,000 customers from whom only the organizers knew whether they had a caravan insurance policy or

2 On the 'The Insurance Company Benchmark Homepage' the problem statement, data and results are made available for research purposes (http://www.liacs.nl/~putten/library/cc2000/).

not. The test set was used to determine how well the model performed for cases that were not used in training.

For a prediction task like this, the underlying problem is to find the subset of customers with a probability of having a caravan insurance policy higher than some boundary probability. The known policyholders can then be removed and the rest receive a mailing. The boundary depends on the costs and benefits such as the costs of mailing and benefit of selling insurance policies. In the case of the challenge, we simplified the problem: we wanted the participants to find the set of 800 customers in the test set of 4,000 customers that contained the most caravan policy owners. For each solution submitted, the number of actual policyholders was counted and this gave the score of a solution. So in summary, the prediction model had to be able to calculate a reasonable estimate of the probability that a customer that was not in the training set owned a caravan policy.

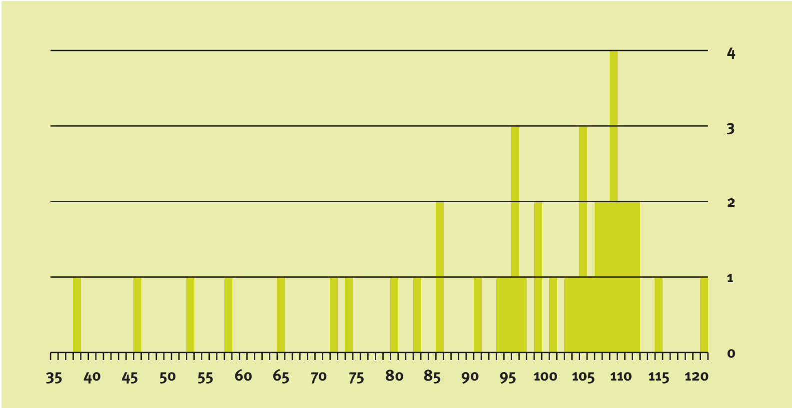

Figure 1
Prediction results. Frequency distribution of the number of real caravan policy owners in the submitted selections. Random selection would result in 42 policy owners, so in most cases there was a substantial improvement in response rate by using the prediction model.

In the end 43 solutions were sent in. In the majority of the cases approaches from more than one area in computational intelligence and statistics were used. The frequency distribution of scores for the prediction task are displayed in Figure 1. The maximum number of policy owners that could be found was 238, the winning model selected 121 policy owners. Random selection results in 42 policy owners. Our standard benchmark tests result in 94 (k-nearest neighbor), 102 (naïve Bayes), 105 (neural networks) and 118 (linear!) policy owners. A wide variety of methodological approaches were used by the participants including boosting, bootstrapping and cost-sensitive classification. Algorithms used included standard statistics, neural networks, evolutionary algorithms, genetic programming, fuzzy classifiers, decision and regression trees, support vector machines, inductive logic programming and others (see Part 6, Data mining methods and technology). A general result reported by most participants was that the product usage variables were better predictors than the sociodemographic variables. This was to be expected. Previous customer behavior is the best predictor for future behavior. Furthermore, the basic assumption behind the sociodemographic variables, that every one in the same zip code area is similar, is often violated in practice. However, these zip code-based variables can still be valuable, for instance for descriptive data mining or when product usage is not available.

The challenge confirmed our findings in the prior commercial project: that by using data mining prediction a clever selection of prospects could be made. Such models can be used in situations where manual selections are not feasible. For example, marketeers might be able to make a coarse grained manual selection by using their marketing knowledge. Models can then be used to make a fine grain selection from this set. Furthermore, marketeers sometimes lack the time to experiment with different queries to determine an optimal mail shot. Automated modeling can help in this case as well. In both cases it is important that the model can be explained and that the prediction performance on new cases is demonstrated.

A number of findings in this direct marketing case might be relevant for other applications.
First, the spread in the prediction scores is rather large. Apparently, expertise with the method is required. To let data mining grow into a tool for end users, this expertise must be made explicit, formalized and automated where possible. This should include steps that go beyond the core algorithm, such as data preparation (e.g. feature selection) and evaluation (application specific evaluation measures, boosting, combining models).
Second, an approach which was suggested to be the most prudent in the after challenge discussions was to 'try the simplest first and be self-confident'. On real world prediction problems like the one in the challenge, one should try a wide variety of approaches, starting with the simplest ones, because they seem to work best. This indicates that simple computational learning algorithms can play an important role, when used as alternatives to standard statistical techniques and thus improving the choice range of algorithms. On the other hand, simple statistics should always be included.

Vision for the future

With respect to current marketing applications, there are a lot of similar useful predictions that can be made, such as predicting customer retention or estimating turnover potential.

Looking into the future always boils down to some degree of speculation, but it is reasonable to assume that the trend towards more personalized and ultimately one-to-one marketing will pull through. Most modern marketeers claim

that this is the holy grail of marketing, however, it will most certainly make life a lot harder for them. As public awareness on privacy issues increases, companies that perform marketing on a non-selective, wasteful manner will simply be neglected by customers. A so called one-to-one relationship must be meaningful from the perspective of the customer too. Privacy protectors and data mining might form a paradoxical alliance with the same objective: less undesirable sales contacts.
Because of this trend towards personalization, the average mailing campaign size will be orders of magnitude smaller, so the selection quality requirements will increase. On the other hand, the amount of campaigns might grow dramatically. So manual selection or even 'manual' construction of data mining models will become merely impossible. In modern analytical customer relationship management and marketing campaign management software this trend towards automation of the selection process is clearly visible.

The next step would be that instead of off-line, centrally organized campaigns marketing actions will be real-time, distributed and customer driven. For instance, any change in customer profile data might result in a changed product propensity level and fire off a marketing action. Even changes in customer profiles of people that resemble a customer might result in such an action.
Finally, it is reasonable to expect that customers will start to use mature, automated mining software to turn the tables. Intelligent agents will scour electronic markets to search for necessary, interesting and useful products. Customers will keep profile information private, and only release anonymous profile information, if they directly profit from it.
If this black-and-white scenario becomes reality, data mining will be a blessing for consumers, but for marketeers it will seem to be more like a devil in disguise. But in the end a more efficient exploitation of marketing budgets will benefit both.

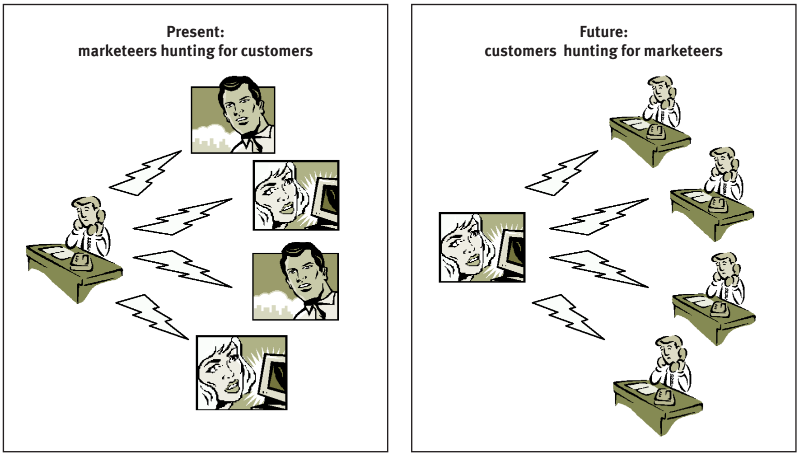

Figure 2
In future customers will use data mining to find the best product.

References

– Putten, P. van der. (1999). Data Mining in Direct Marketing Databases. In: Walter Baets. (ed.). Complexity and Management: A Collection of Essays. World Scientific Publishers, Singapore
– Putten, P. van der, M. van Someren. (eds.). (2000). CoIL Challenge 2000: The Insurance Company Case. Sentient Machine Research, Amsterdam. Also a Leiden Institute of Advanced Computer Science Technical Report 2000-9. June 22

Pockets of predictability in financial markets

Willem-Max van den Bergh[3], *Jan van den Berg*[4]

Predictability of financial market returns

In this chapter we address the potential of advanced data mining techniques in financial time series analysis. But first of all we must deal with the question of how return series[5] can be more than accidental strings of random, unrelated market-level changes. Actually, we bring the efficient market hypothesis (EMH), which states that current market prices reflect all relevant information, under discussion. By definition, new information arrives randomly and thus, as many finance scholars will agree, a securities market that does not exhibit a random walk cannot be efficient. On the other hand there is no doubt among practitioners that financial market returns are predictable. However, most admit that this predictability varies over time. Traders make money by exploiting market opportunities. By doing so they have become very good at recognizing momentary (local) features of the market. This might explain why technical analysis is so popular among traders. Technical analysis gives a description of the market that is very different from the usual statistical description. Technical patterns are truly local and only once in a while, a distinct buy or sell signal emerges. The focus of traders on tools that give a local description of the market cannot be ignored. These situations, also called local patterns or pockets of predictability, are hard to find with conventional techniques that try to model the overall market structure. In the last decade, the notion that there are periods of higher than normal predictability in the market has also entered financial literature, see for an example [Pesaran, 1995] and [Dunis, 1996]. Since these pockets of predictability are very specific and undoubtedly different from the normal everyday market behavior, we will refer to them as exceptions.

Financial markets are characterized by a large number of participants, each having a different appetite for risk, a different time horizon and different motivations and reactions to unexpected news. In the circumstances it would come as

3 Drs W.-M. van den Bergh, vandenbergh@few.eur.nl, Erasmus University Rotterdam, Department of Financial Economics, Rotterdam, The Netherlands

4 Dr Ir J. van den Berg, jvandenberg@few.eur.nl, Erasmus University Rotterdam, Department of Informatics and Economics, Rotterdam, The Netherlands, http://www.eur.nl/few/people/vandenberg

5 Return series relate here to the relative value change (in %) of stocks between time t-1 and t.

return =

$$\frac{\text{stock value}_t - \text{stock value}_{t-1}}{\text{stock value}_t} * 100\%$$

a surprise, if all these complex interactions were to average out in a linear fashion. Consequently, new mathematical and statistical tools, which rely heavily on the analysis of non-linearities, are being developed. Many of these new tools can be characterized as data mining techniques, i.e. methods of exploratory analysis looking for meaningful patterns and rules. Some of them use a bottom-up approach, called knowledge discovery, where no prior assumptions are made; the data is allowed to speak for itself. Undirected knowledge discovery [Berry, 1997], a main issue in this article, attempts to find patterns or similarities among groups of records without the use of a particular data field or predefined classes.

A quick search in the existing literature of undirected knowledge discovery did teach us that in general, only modest attention is put on the discovery or learning of exceptions. Instead, most algorithms put the emphasis upon the discovery of the common rules. Exceptions are often an afterthought. In this chapter, we explicitly look for exceptional patterns having certain predicting power. The remainder of this chapter presents two main issues. The first is more economically oriented and pursues theoretical backgrounds for the existence of exceptions in financial markets. We will go into some very specific exceptional situations in financial markets like stock market crashes. Some novel attempts to model market behavior during these situations are presented in order to get more insight into the nature of such exceptions. Our main focus will be on the system dynamics of financial markets and we will look at different fuzzy states (market regimes) and the transition between these. Analogies with physical systems and the existence of exceptions like avalanches and traffic jams will be discussed in an intuitive manner. In the second section a Competitive Exception Learning Algorithm (CELA, see below) is introduced where exceptional patterns are inferred from a given set of data pairs. Though our inspiration for developing CELA stems from the wish to analyze financial time series and finding pockets of predictability, we think the algorithm is applicable in a much broader field.

Exceptions in financial markets

According to many advocators of the efficient market hypothesis, phenomena like bulls[6], bears[7] and market bubbles are viewed as accidental strings of random, unrelated market-level changes. But how irrational are bubbles really? Recent academic literature [Treynor, 1998] indicates that it is not obvious that, if the market level is temporarily perturbed, equilibrium forces will return it to its original level. If, for example, the perturbation drives the market level up slightly, the bulls will gain at the expense of the bears. Because of the wealth shift, the market will accord greater weight to the bulls than formerly and less weight to the bears. So, now the equilibrium market level has risen. Although the actual price change may have been a complete surprise, the consecutive shift to a

6 Bull: buying stocks to cause price rise, or in expectation of a price rise.

7 Bear: selling stocks to cause price drop,planning to buy back larger quantities at a lower price.

new equilibrium may be quite deterministic. Obviously a potential regime shift is determined by the opportunities bears have to continue their game: which options do they have under these extreme conditions? If the perturbation is quite large, and many investors reach the end of their financial capacity, the new equilibrium may be some distance from the old. Furthermore, the transition will not be smooth and shorter or longer periods of overshooting may exist, which may be related to the way the market is organized and is able to process the order flow during hectic periods. The information available to investors about what is actually happening will certainly influence market behavior. We will call all such deviations from random walk exceptional. It is not necessary to argue that early detection of these exceptions is of great value to investors and risk managers.

The nature of market crashes

One of the most interesting examples of exceptional situations in financial markets is a market crash. Such crises show some notable similarities with traffic jams. The transaction flow in financial markets knows the same sudden transition between orderly normal behavior and sudden explosive and seemingly irrational bursts, eventually followed by one or more collisions, when market parties default. For both, financial crises and traffic jams, it is questionable whether they are predictable. Undoubtedly this is a hot topic: J.P. Morgan has an 'Event Risk Indicator'[8], Credit Suisse First Boston an 'Emerging Markets Risk Indicator', and Lehman Brothers a 'Currency Jump Probability Measure'. However, if we may believe The Economist ('The Perils of Prediction', August 1999): "Don't expect them to work ... sophisticated stuff, to be sure. Yet the real question is whether this improves on what investors use today...!" On the other hand, for complex phenomena like the weather there is a clear breakthrough to better insights and new technological tools.

Different market regimes as fuzzy system states

The understanding of traffic jams increases little by little, at the same time providing insights for financial markets. Most models used originated more or less in theoretical physics. Consequently, a first source for these models is professional literature like 'Nature, Science and Physica'. E.g. recent research at Daimler Benz [Kerner, 1977] emphasizes the relation between the development of traffic congestion and crystallization. A distinguished feature of this process is very abrupt state transition. A single dust particle can act as a catalyst in a supersaturated solution and lead to sudden crystal formation. The new crystals act themselves as catalyst and crystallization spreads like a wave. Thus, the process leading to traffic congestion shows several distinct states. When flow increases, we observe 'clusters' of high traffic density moving with increasing intensity backwards. Above a distinct density all car speeds become, so to

8 http://www.jpmorgan.com/CorpInfo/PressReleases/1998/01301998_ERI.html

speak, ‘synchronized’. One single speed is imposed on all cars and overtaking becomes difficult. When density increases even further, collisions may occur.

In financial analogy, one single sales order might well trigger many others and lead to a high volatility. And when certain boundaries are encountered, a crash might even follow. When volatility increases, price rises and price drops of increasing magnitude follow each other even faster and suddenly the price breaks through in a distinct direction. Generally this takes the shape of a persistent price drop, but sometimes we observe a sharply increased price level (usually referred to as euphoria, but holders of a short position will definitely disagree). It seems as if returns freeze for some time at a more or less constant value. This resemblance to state-change in pure physics suggests the investigation of the properties of physical models for predicting market behavior. Identification of different states and the exact context for state transitions is crucial in these models. However, the financial equivalent of such models is truly complex, mainly on account of the many possible ways open to individuals to react to sudden market movements. Furthermore, state transitions will undoubtedly be fuzzy and less distinct than in many physical situations (‘ice below zero and steam above 100^{o} Celsius’). Advanced data mining techniques used on databases containing the full contextual market setting of large amounts of transactions have a high potential in this respect, especially if they are able to deal appropriately with fuzzy events (see also Section 6.2.16, Fuzzy logic techniques).

State transition and options theory

Financial theory and physics are less dissimilar than one might think. The kernel of the options pricing model as formulated by [Black and Scholes, 1973] is a differential equation that is not different in any respect from what is known in physics as ‘heat transfer equation’. If a source of heat is put into contact with some fluid, the temperature of the fluid will rise gradually until the boiling-point is reached. The analogy in options pricing: if the price of the underlying value of an option changes, the pay-off for the owner changes gradually until the strike price is reached. Beyond that point the pay-off is fixed at the exercise price. Clearly the relation between changes in the underlying value and the pay-off of an option is non-linear. This situation is quite customary in financial contracting. If a fixed interest bond is used by a firm to finance an asset position with a variable value, the pay-off to the bondholder is constant (principal plus interest payments) if, at maturity, the asset is worth more. If the asset is worth less, the bondholder only gets this lower value, i.e. what remains in case of default.

The higher the threat of default, the more it will influence the behavior of the contractual parties, which in return may influence market transactions. In other words, in the proximity of position limits, the return/volatility pattern of asset prices may differ from the usual, normal pattern. Things will become really difficult, if there is no counter-party for a transaction or if the maximal order processing capacity of the market has been reached. The only way open may be a stop-loss order resulting in a reinforcement of the current market direction. Others are triggered to react and so on. Clustering increases and in extreme conditions crystallization starts and the market freezes and the returns maintain to fall (or rise) for some time. Such events should be identifiable early in market return time series by shorter or longer periods of serial correlation. Advanced data mining techniques appear to have great potential for finding patterns in the market context, when distinct behavior has occurred. The hunt should be for a method of separating these specific (and sparse) contexts from situations that are not followed by abnormal market behavior.

From a physically orientated explanation for market crashes we can formulate the following hypotheses:

- Crashes of all magnitude are possible, only limited by the amount of all positions taken.
- The probability for a crash may be higher than predicted by standard risk management models [9] based on standard deviation or semi-variance.
- There is no necessary relation between large returns and special news events [10].
- There is no typical time interval between two crashes.

Institutional and behavioral bias

The probability of crashes is, apart from the order flow itself, determined by 'institutional' properties like the way the market is organized, order processing, availability of information about recent prices and order positions. As mentioned earlier, the way positions are limited and other regulations like 'circuit breaking', i.e. stopping the order flow for a given period when the market gets over-heated, will certainly play an important role. Market organization is important for the speed of order processing in active moments [Martens, 1998]. Capacity and reliability of the existing automated system are obviously important.

A large and rapidly growing body of literature attributes various stock market anomalies to behavioral biases. An example is the finding of [De BondtThaler, 1985] that people tend to overreact to dramatic events: e.g. investors seem to attach unrealistic probabilities to stock market crashes. Some researchers mention what is called 'representativeness bias', which, simply put, means that people tend to think: "if it walks like a duck and quacks like a duck, it must be a duck." [Scott, 1999]. In other words, if a given market context somehow resembles a situation of the past that was followed by a specific market behavior like

9 The Black and Scholes [Black and Scholes, 1973] model discussed earlier is one of these.

10 Note that this is incompatible with the (among econometricians) popular 'jump-diffusion model', which relates large external shocks to large market movements.

a crash, they get anxious and react accordingly. The same may occur, when people get euphoric when they recognize a former prosperous setting for a given stock or the whole market. In this sense the market reaction becomes a 'self-fulfilling prophecy'. In our view, such findings provide a strong argument for the use of advanced data mining techniques that map historic abnormal (exceptional) market behavior to the contextual environment immediately before and during such events.

Mining with a Competitive Exception Learning Algorithm (CELA)

In the previous section we discussed the typical behavior of financial time series in exceptional situations. We will now consider some properties of a truly intelligent agent (either a person or a computer program) being able to early detect such events. In mathematical terms, his (or its) task is to correctly predict a future market state $y(t+1)$ given m historical states $x(t)$, $x(t-1)$, $x(t-2)$, ..., $x(t-m+1)$. Here, any historical state $x(t-i)$ is supposed to potentially affect the future market state. E.g., by purely analyzing the time series $y(t)$, $x(t-i)$ may simply describe the historical market state $y(t-i)$ itself. Otherwise, $x(t-i)$ may be composed of 'environmental' state values at time $t-i$ (like economic news facts about rental, unemployment, growth rates, etc.) or institutional setting, which, when changing, have their impact on future market states. The Competitive Exception Learning Algorithm (CELA) introduced here is especially designed to play the above mentioned role of 'intelligent agent'.

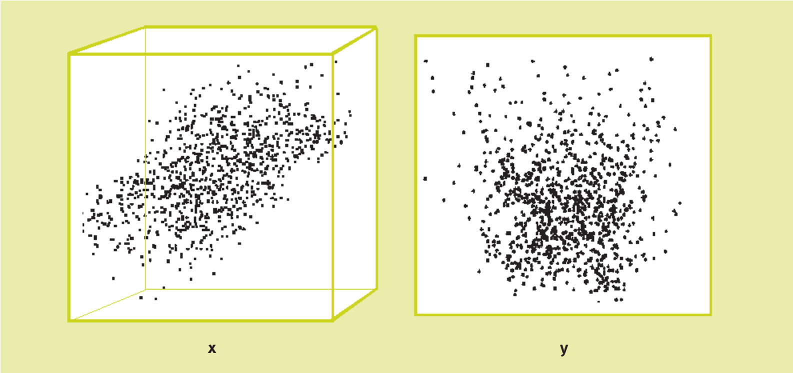

Figure 1
The 3-dimensional input X and the 2-dimensionial output data Y of the working example.

As a working example to illustrate the operation of CELA, we will use a simulated data set containing conditional serial dependencies, in line with market concepts set out previously. We assume that prices are bid up in a bullish market or bid down in a bearish market, resulting in shorter or longer periods of serial correlation in returns. So, a set of data pairs has been constructed having the above given characteristics of noise trading: output space Y (see Figure 1) consists of data points (cases) representing the future 1-period return and the future 5-period volatility. The input space X consists of data points representing

the standard deviation over the last 5 periods, the average return over 5 periods and the 1-period past return. Each new return is generated using a transformation from a uniformly distributed random variable to a normal distribution under 3 regimes (2 of which are exceptional in the sense that they generate serial dependency). The goal of CELA is to infer the regimes together with the corresponding exceptions.
Generally speaking, the task of the algorithm is to infer a conditional relationship between an M-dimensional input space (*X*) and an N-dimensional output space (*Y*). A more formal description of the algorithm is given in [Bergh, 2000]. Five distinct steps can be distinguished:

1. Determine clusters in *Y*, using a competitive learning approach, in such manner that overall fuzziness is minimized.
2. Determine the unconditional output fuzzy membership distribution (UOD) for these output clusters.
3. Determine clusters in *X* where the conditional output fuzzy membership distribution maximally deviates from the UOD as found in the previous step. In this step we also use a competitive learning approach.
4. Identify a fuzzy rule base from input and output clusters as found in the previous steps.
5. Perform a function approximation in order to estimate an output estimate for each individual input case.

We show the result of the first 4 steps on our working example in Figure 2. The number of clusters in our method is user defined. We chose 4 output clusters, 3 of which have their centroïds[11] at a fixed location in a corner of the output space (see the right-hand side in Figure 2). The location of the 4-th cluster centroïd is determined using the above mentioned competitive learning approach. The centroïds are located so that all cases are as near as possible to one of these centroïds. All cases compete, in a way, to attract a centroïd. The clustering is characterized by minimal fuzziness, since the average membership of each case for the cluster with the nearest centroïd (the winner cluster) is maximal. Each case is member of all (in the example 4) clusters, while the memberships value for the winner cluster is highest. The membership values sum up to 1. In our approach we consider cluster memberships as inversely related to the squared distance from the centroïds.

In the next step these individual memberships are averaged over all cases, arriving at the unconditional output distribution (UOD). The percentages in the square on the right-hand side represent this UOD, which equals $\bar{\mu}^{Y}$ = (0,0805, 0,7427, 0,0852, 0,0916). Thus, having no further information about a specific case, it's expected relative membership is 74,27% to the largest class, 9,16 % to the next, etc.

11 Center of gravity of a shape or point cloud.

The third step entails partitioning of the input space. This step is crucial but complicated. The input clusters have been situated so that the local behavior of the associated output is as different from the UOD as possible. In other words, the competitive algorithm searches for exceptional events that jointly occur with a specific input context. This time, the input cases x_p compete for cluster centroïds $\overline{x}_b$ (b indicating the cluster number). Cases that, with respect to their output, deviate strongly from the UOD (measured by the Euclidean distance between the local membership vector and the UOD) have a stronger attraction to input centroïds than 'normal' cases that are distributed more or less similarly as the UOD.

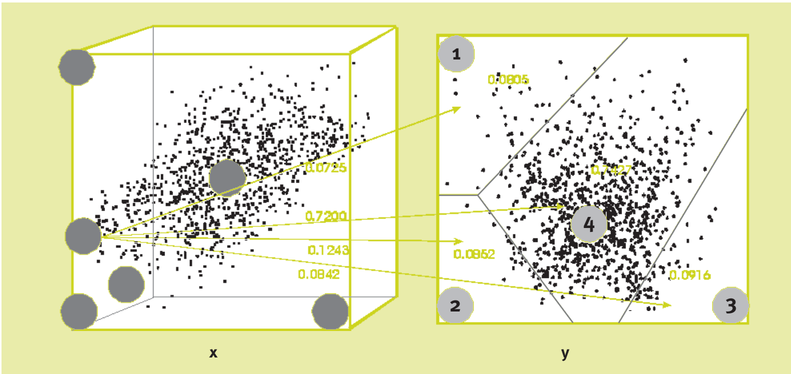

Figure 2
Partitioning of both the input and the output space after learning. The output space Y is divided into 4 fuzzy classes, each predicting a specific return development. For example 1= low at t, high at t+1, 2= low at t, low at t+1, 3= high at t, low at t+1, 4 = around average. Exceptional events at t reflect in other than average memberships for clusters in the input space. leading to a different membership probability for the clusters in the output space of expected returns.

Output values y_p deviating from the UOD are considered 'potentially exceptional' and thought to emit an output exception signal (OES), based on the Euclidean norm as mentioned earlier. Only if they systematically correspond to specific locations in the input space, however, are they actually considered exception or exceptional pattern. This step of the CELA algorithm uses the joint product of OES and winner input cluster membership and seeks to minimize its average value over all cases. The resulting input clustering is characterized by minimal fuzziness as was the output clustering, but now with respect to the characterization of exceptional patterns. Once more, each case is member of all (in the example 6) input clusters.

In step 4 of the algorithm we calculate the weighted average output membership distribution per input cluster. The input memberships are used as weights. The intuition behind this step is clear: the more a specific input is member of a given input cluster, the more we expect its output behavior to behave like the average output behavior of that cluster. This average cluster behavior (a 'local' output distribution) is thus conditional on the input cluster and can be conceived as being located in the centroïd. In this sense such local output distribution can be seen as a rule: knowing we are in a given input cluster centroïd, we have a specific output distribution which differs from its unconditional equivalent. All rules, i.e. the local behavior of all cluster centroïds apply to a certain extent to each input case, depending on the membership for the input clusters. Hence, we may typify the rules as fuzzy. Consequently, input locations lying far away from any cluster centroïd will hardly follow any rule and cannot be classified as exceptional.

Once the fuzzy rule base has been identified, we are also able to estimate (step 5) an output location for any input data point. First we calculate the relative membership for each of the input clusters and next we derive the conditional relative output cluster membership based on the fuzzy rule base and using the input cluster memberships as weights. From this estimated output membership distribution we calculate the expected value as a weighted average from the cluster centroïds.

The 3-dimensional space at the left of Figure 2 shows the result of the input clustering. For one of the input clusters the conditional rule as determined on the working example is shown as an example. We notice that the conditional distribution $\bar{\mu}^Y | \bar{\mu}_c^X$ = (0,0725, 0,7200, 0,1243, 0,0842) which clearly deviates from the UOD of (0,0805, 0,7427, 0,0852, 0,0916). Notably, the conditional membership of 0,1243 for the third cluster is higher. As can be seen from the figure, this cluster has a relatively low return and a relatively low volatility. In economic terminology, we may characterize it as 'a price drop'. The input cluster has a 5-day volatility that is low to normal, a low 5-day return and a relatively low 1-day return. An economic description of this situation might be something like: 'market is falling'. Now we can express the earlier rule as: 'If the market is falling, there is a higher than normal probability of another price drop'. This rule, although it may look trivial, exactly mimics an element of the simulated behavior of the time series during an exceptional regime. In this sense it is a good illustration of 'opening the black box' by studying the fuzzy rule base resulting from CELA.

Conclusions and prospects for the future

Intelligent data mining has the potential to learn very specific characteristics of financial market behavior. Theories that incorporate the institutional setting of the market as well as the behavioral aspects of all market participants and take a true local view of the conditional setting imply non-linear market behavior. In this contribution we have shown some extra capabilities of specialized data mining techniques which are able to deal with such underlying non-linear relationships. The Competitive Exception Learning Algorithm (CELA) is an example of such a technique with potential value for professional investors as well as regulators. It may help not only to anticipate irrational market movements, but also to understand them better. Of course, extensive future research is necessary.

We are convinced that new highly dedicated algorithms (like CELA) will be detected and eventually become available as standard, albeit specialized, tools. The possibility of such tools to get insight in the 'rules behind the system' makes them appropriate for embedding in highly domain-oriented, but user-friendly decision support systems. Without doubt, they will make financial markets more efficient in their important task of risk-diffusion between individuals.

References

- Bergh, W.M. van den, J. van den Berg. (2000). Competitive Exception Learning using Fuzzy Frequency Distributions. ERIM-report ERS-2000-06-LIS. http://www.erim.eur.nl/. Publications in Management, Section Business Processes, Logistics and Information Systems
- Berry, M.J.A., G. Linoff. (1997). Data Mining Techniques for Marketing, Sales, and Customer Support. John Wiley & Sons, New York
- Black, F., M. Scholes. (1973). The Pricing of Options and Corporate Liabilities. Journal of Political Economy **81**:637-654
- Bondt, W. de, R. Thaler. (1985). Does the Stock Market Overreact? Journal of Finance **60** (3):793-805
- Dunis, Ch. (ed.). (1996). Forecasting Financial Markets, Exchange Rates, Interest Rates and Asset Management. John Wiley & Sons
- Kerner, B.S., H. Rehborn. (1977). The Physics of Traffic Jams. Physical Review Letters. November 3
- Martens, M. (1998). Price Discovery in High and Low Volatiliy Periods: Open Outcry Versus Electronic Trading. Journal of International Financial Markets, Institutions and Money **8**:243-260
- Pesaran, M.H., A. Timmermann. (1995). Predictability of Stock Returns: Robustness and Economic Significance. Journal of Finance **4**:1201-1228
- Scott, J., M. Stupp, P. Xu. (1999). Behavioral Bias, Valuation and Active Management. Financial Analysts Journal **55** (4), July/August
- Treynor, J., Bulls, Bears and Market Bubbles. (1998). Financial Analysts Journal **54** (2):69-89