Nonlinear Time Series Analyses in Industrial Environments and Limitations for Highly Sparse Data

Emili Balaguer-Ballester

School of Design, Engineering and Computing
Address: Talbot Campus, Poole, Dorset, BH12 5BB
Bournemouth University, UK

Abstract. This work presents case studies of effective knowledge transfer in projects that focused on using nonlinear time series analyses in varied industrial settings. Applications, characterized by intricate dynamical processes, ranged from e-commerce to predicting services request in support centres. A common property of these time series is that they were originated by nonlinear and potentially high-dimensional systems in weakly stationary environments. Therefore, large amount of data was typically required for providing useful forecasts and thus a successful transfer of knowledge. However, in certain scenarios, classifications or predictions have to be inferred from time windows containing only few relevant patterns. To address this challenge, we suggest here the combined use of statistical learning and time series reconstruction algorithms in industrial domains where datasets are severely limited. These ideas could entail a successful transfer of knowledge in projects were more traditional data mining approaches may fail.

1 Introduction

With the increasing possibilities of accessing massive amount of data, prediction of customer's behaviour in industrial domains has become a very active research area during the last twenty years. For instance, one of the most important goals within data mining in Web environments consists of predicting interesting characteristics of users using their previous navigation patterns or *clickstreams* (Martin et al., 2004). These kinds of data analyses enables to better profile the customers commercial Web portals, and in turn, to customize the interface by providing the most suitable services for reach individual user in advance (Fu et al., 1999; Carberry, 2001; Martin et al., 2006). Therefore, it is particularly relevant to anticipate users' demands. Recently, it was shown that powerful nonlinear algorithms for predicting of users' preferences can significantly increase the effectiveness of large web portals of organizations or institutions; provided we have access to a significant amount of historical data (e.g. Martin et al., 2004; Martin et al., 2006, 2007).

R.J. Howlett et al. (Eds.): *Innovation through Knowledge Transfer 2012*, SIST 18, pp. 51–60.
DOI: 10.1007/978-3-642-34219-6_6

Besides web mining, time series analyses have been successfully applied in other industrial settings. For instance, a recent study addressed the more efficient management of Support Centres (Balaguer-Ballester et al., 2008). Support Centres (SCs) usually deal with all the requests reported by either external customers or internal users of a company. The formal contract between the service provider and the customer (the service recipient) usually contains clauses that economically penalize or incentive the SC depending on its performance. Therefore the successful anticipation of service request is an essential aspect in the efficient management of both human and technological resources that are used to solve these eventualities; and thus is of highly significant economical relevance.

Another domain in which predictive analyses have been successful is targeted marketing. The latest marketing trends focus on maintaining and optimizing the behaviour of loyal customers rather than getting new ones i.e. to increase the *net* value of the customer in the long term or lifetime value, LTV (Reichheld, 2001). The relationships between a company and its costumers follow a sequence of action-response cycles, and customer's behaviour typically evolves according to the marketing actions (Pfeifer and Carraway, 2000). For this reason, in targeted marketing, is also essential to predict which subsequent marketing action (for instance, offering a discount to a particular customer) will result in an increase of customer‘s LTV (Gomez-Perez et al., 2009).

The case studies outlined above are good examples of successful knowledge transfer between Academia and Industry. The algorithms developed in a scientific environment were then implemented in several commercial software Java© platforms and used for improving the performance of the web sites, incidences management centres or marketing policy of the companies, respectively. The processes underlying such applications were typically nonlinear and therefore very challenging for prediction models (Balaguer-Ballester et al., 2008). Nevertheless, forecasts were still possible because large amount of data was available in all situations, enabling us to robustly optimize the parameters of sophisticated machine learning approaches; and thus knowledge transfer was successfully implemented (Martin et al., 2006).

However, in certain challenging scenarios, predictions have to be inferred from very short time series, for instance in biomedical applications using functional magnetic resonance brain-imaging, where the sampling rate is very low and therefore only tens of data patterns are available. Moreover, sometimes the number of accessible variables does not suffice in accounting for the real complexity of the system. In these challenging research situations, the underlying system dynamics is not accessible from the available data. Therefore knowledge transfer may not be successfully implemented using standard machine learning and computational statistics approaches. In this work, we propose the combined use of statistical learning and time series reconstruction algorithms for analysing intrinsically nonlinear problems in industrial settings, particularly when the number of relevant temporal patterns is severely limited. These approaches could be useful in real-life industrial domains were more traditional inferential approaches are not easily applicable due the complexity of the system (many potentially independent variables) and data limitations.

2 Case Studies and Discussion

2.1 Web Mining and Viability of Recommender Systems

The first case study of this work consists of designing recommendation strategies in commercial or institutional web portals using predictive models.

Personalized recommender systems has been a very active field of research during the last decade (Carberry, 2001), and they become an important component of e-commercial Web sites (for instance, Amazon.com). Such systems have been extensively tested and can reliably process large amount of data (Zukerman & Albrecht, 2001). Collaborative filtering is perhaps the most popular approach. This type of systems computes indicators of interest such as frequency of access or indexes of user similarities; and then suggests the most suitable services or products on-line to the user. Some early and well-known examples are Recommender (Hill et al., 1995) or NetPerceptions (Resnick et al., 1994). Those "coarse-grained" analyses have the advantage that they do not need any complex object representation, but only aggregated statistics. Therefore, collaborative filtering is a good strategy in commercial portals provided that "people-to-people correlations" i.e. variations in tastes, accounts for the majority of the variation in user's preferences (Schaffer et al., 1999). In contrast, content-based approaches are based on users' past preferences in order to forecast his/her future behaviour; thus, those learning algorithms are particularly suitable for situations in which users exhibit a predictable behaviour i.e. a significant "item-to-item correlation". Figure 1 shows a simple schema of these two approaches.

An example of the successful applications of recommender systems to commercial portals is the Java platform iSUM©; developed by the Spanish IT and software company TISSAT© (www.tissat.es). The data mining capabilities of this product were entirely based on the results shown in Martin et al. (2004, 2006, 2007). The two classes of recommender algorithms outlined above were implemented, in order to provide the optimal solution depending on data characteristics. For instance, in Martin et al. (2004), content-based recommenders based on Associative Memories, Time-Delayed Perceptrons (Balaguer-Ballester et al., 2002), Classification and Regression Trees (CART; Duda et al., 2001), and Support Vector Machines (SVM; Vapnik, 1998) were used for predicting the services related to bureaucracy, shopping or entertainment etc. based on the historical information of individual users. The optimal models where able to predict with an 80% of success the preferred service by users in a large institutional portal; thus validating the "item-to-item" assumption in this scenario (Martin et al., 2004).

In other situations however, "people-to-people" collaborative recommendation is more appropriate, since the aim is to find inter-user similarities rather than idiosyncratic behaviours of individual users. In particular, in Martin et al. (2006, 2007) users' behaviour was profiled by using different neurologically inspired clustering algorithms such as Adaptative Resonance Theory-2 Networks (ART2, Carpenter & Grossberg, 1991); which have been designed to adapt clusters to new data patterns, without disrupting the already established clusters. This feature can support on-line tracking of user profiles. Moreover, in this work, we developed an artificial data generator; which permitted to benchmark different recommendation algorithms in arbitrarily complex web scenarios. Our approach was able to accurately find groups of similar users on a range of real and simulated e-commerce websites; and afterwards new users were offered services which were primarily accessed by the users of the same group. As a result, this approach improved by 1.5-2.5 times the prediction accuracy obtained with more standard approaches (Martin et al, 2007). As briefly mentioned earlier, this functionality was transferred to the web platform iSUM©. The recommender systems above described were integrated in the Artificial Intelligence Module, available in versions 4.x and superior of such software product. These novel functionalities permitted to successfully position iSUM© within the international market of intelligent web mining software.

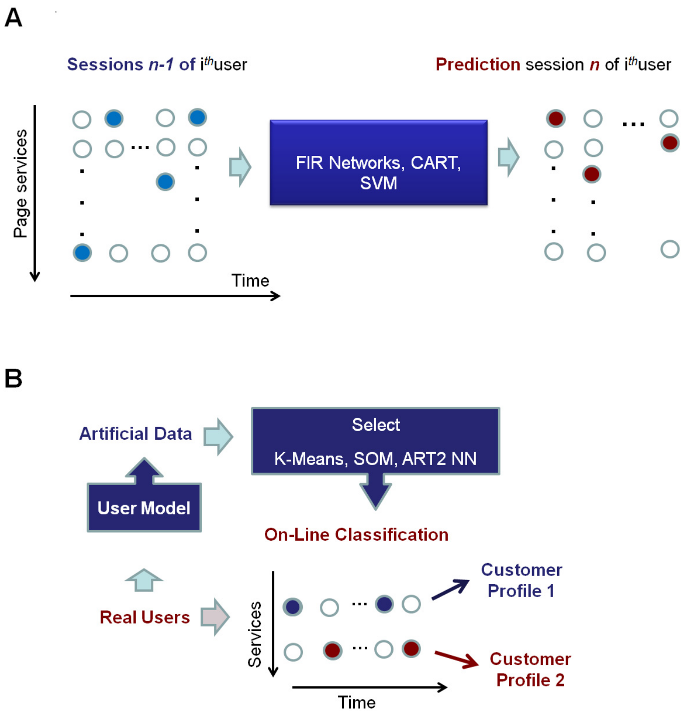

Fig. 1 The most widely used recommender systems. **A**. Content-based recommendation, where “item-to-item” correlation is significant and thus prediction of the future preferences of customers is feasible (FIR NN: Finite Impulse response Networks; see e.g. Balaguer et al., 2002). **B.** Collaborative filtering, where the aim is to find inter-user similarities rather than idiosyncratic behaviours of individual user, thus unsupervised clustering algorithms are typically used (SOM: Self-Organizing Map, Kohonen, 1997). In both approaches (especially in Content-based-approaches), accuracy for short time series is severely compromised.

Unfortunately, these two recommendation algorithms require collecting large amounts of data from each user in order to reliably estimate the parameters of such nonlinear models. This requirement is of course not exclusive of e-commerce websites but is common to many predictive scenarios in other industrial domains; as discussed in the following sections.

2.2 Predicting Service Requests in Support Centres

A second example of successful predictive modelling in industrial domains is the automatization of Support Centres management.

Although telephone has been the traditional way to provide support, nowadays the Internet is widely used (Balaguer-Ballester et al., 2008). The contracts between service provider and customers (service level agreement) define a series of quality measures or service level measurements (SLMs) that are used to evaluate the quality of the support service. Different system performance and business parameters can be considered for SLMs: aggregated statistics of time-dependent parameters, service availability, number of affected users, metrics based on a particular business process, etc. These measurements are automatically collected, maintained and analysed in order to manage the service support process. The “Help Desk” platform consists of several software applications that allow recording all the information involved in a SC; which is considerably wide: general information, purchases, complaints, etc. SCs are especially interested in managing these forthcoming events as good as possible. Moreover, is preferable that, whenever it is possible, problems are solved by operators at the first levels of the system hierarchy. Therefore the prediction of the number of forthcoming service requests as well as the time when they occur will permit to optimize the resources used to solve these eventualities.

For instance, in Balaguer-Ballester et al. (2008) we analysed different eventualities managed by an official SC (CETESI©; a project which aimed for a complete integration of institutional services provided by local authorities in Valencia, Spain). In this project, we used time series reconstruction algorithms in order to evaluate the complexity of the system dynamics, which varied during the four years of recordings. Therefore, we were able to determine when a linear predictor will suffice (for low-complex multivariate time series of incidences) or rather highly-parameterized nonlinear approaches where more suitable. Then, we selectively applied linear/nonlinear models (auto-regressive moving averages with exogenous inputs and neural networks) depending on the previous analyses in order to apply the most robust and reliable method to predict future events in each situation.

Using this “meta-approach” for selecting the most suitable model, we obtained an 86–94% of out-of sample prediction accuracy in certain management projects; contributing to the successful fulfilment of the service level agreement. As a result, local authorities renewed the contract with TISSAT©; yielding to an increase in the company's revenue and to a better positioning in the Support Centres Management sector.

Nevertheless, those nonlinear time series predictions were less accurate in the projects were long time series (two years of recordings) were not accessible (Balaguer-Ballester et al. 2008).

2.3 Optimizing Strategies in Marketing Campaigns

The most common strategy for increasing the loyalty of customers is by offering them the opportunity to obtain some gifts as the result of their commercial transactions. For instance, the company can assign "virtual credits" for each purchase of those products which wants to be promoted. After a certain number of purchases, the customers can exchange their virtual credits for the gifts offered by the company.

Therefore, the policy of this targeted marketing campaign consists of establishing the appropriate number of virtual credits for each promoted item. The goal is to achieve an optimal trade-off between the cost of the marketing campaign (for instance, the value of the gifts) and the increase in the amount of purchases as a result of the campaign. Commercial domains of this kind can be viewed as a Markov decision problem, in which a company decides what action to take given the current customer state (Abe et al., 2004).

This is however a difficult problem, due to the large number of variables that can be potentially used for characterizing the state of the customer. In Gomez-Perez et al. (2009), we profiled the behaviour of the customer using three optimal features in Relational Marketing which are the last transaction date a.k.a. Recency, Frequency and Monetary value (Figure 2). After that, we used reinforcement learning algorithms (RL, Sutton and Barto, 1998) in order to determine the optimal marketing policy i.e. the assignment of the appropriate number of virtual credits to each customer for each purchase.

Moreover, targeted marketing can have a very complex characterization of the transactions that are involved. This high-dimensionality requires the combined use of RL algorithms with some dimensionality reduction approaches. For overcoming this drawback, we combined Q-Learning methods with SOM networks (Smith, 2002), which enabled us to compute the optimal policy in a complex credit assignment problem (Gomez-Perez et al., 2009; see Figure 2).This combined approach indicated that incurring in more marketing costs for certain customers (some of the most loyal ones) results in a significant increase (over 50% in some cases) of the long-term LTV with respect to the previous strategy (established manually by marketing experts).

Therefore our algorithm revealed that the company's policy had a lack of exploration of apparently non-profitable marketing actions in the short term, which results in larger long-term benefits. Nevertheless, for other groups of customers the manual strategy could not be largely improved suggesting that, in spite we have about 10^6 data patters, a much large amount of data was still required for optimizing the marketing policy on such customers.

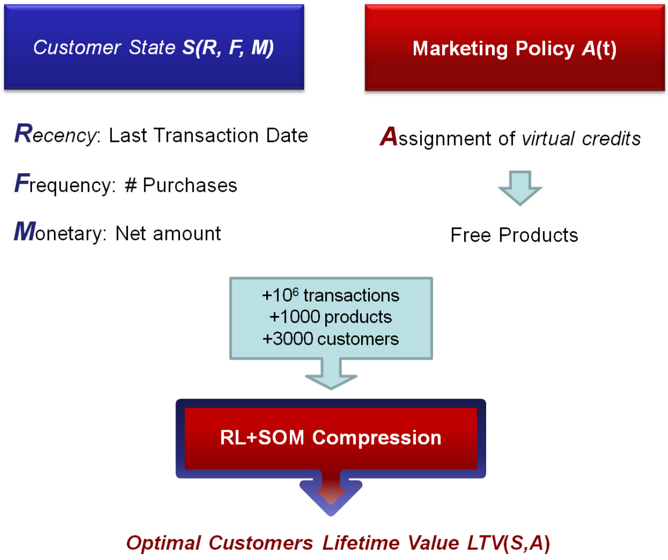

Fig. 2 Methodology for finding the optimum policy $A(t)$ of a Marketing campaign, that is the sequence of marketing actions which maximizes the predicted customers long-term life time value (LTV). A reinforcement learning algorithm (RL, in this case *Q*-learning) combined with a SOM for compressing the variables space (consisting of customer states *S* and marketing actions) enabled us to improve the company benefits (Gomez-Perez et al., 2009).

2.4 Limitations of Standard Approaches in Short Time Series

The element that is common for all three presented above case studies is the availability of large number of "independent" variables and, simultaneously, of large amount of data. Therefore, the underlying dynamical process, which was generating the observable temporal patterns, could be inferred using machine learning approaches. Nevertheless, these algorithms were in general less successful in improving the expert-supervised strategies when long time series where not available.

The reason of this failure can be understood in the light of nonlinear time series analyses theorems (Sauer et al., 1992): When the number of variables in the system is not sufficiently large, we cannot precisely determine the underlying dynamics. For instance, trajectories corresponding to two different customer's behaviours may look tightly entangled in a two dimensional space of recorded variables (*x*-*y* plane in Figure 3), and therefore two truly separated trajectories exhibit multiple crossings. In those ambiguous patterns (crosses in Figure 3), the future direction of the flow cannot be precisely predicted. It is well-known that this ambiguity can be resolved if we add sufficient number of delayed version of our original axes (Takens, 1980); provided we have noise-free and unlimited data. Nevertheless, those assumptions can be particularly inaccurate for short time series.

A possible alternative consists of “supplanting the missing variables” of the system by nonlinear functions of the recorded variables and suitable temporal delays (for instance see Figure 3, z-axis). This simple approach aims to create a sufficiently expanded space such that trajectories are not ambiguous anymore. Therefore, trajectory analyses are feasible by using linear classifiers which have few parameters to estimate and thus may not require large datasets for obtaining robust and generalizable predictions.

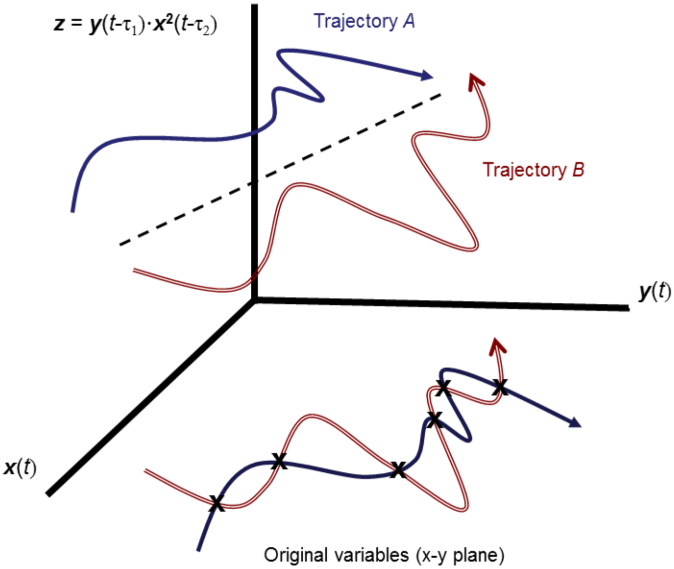

Fig. 3 Reconstructing nonlinear dynamics in noisy time series. When the number of recorded variables is not sufficient (for instance, the *x-y* plane in the figure), trajectories are entangled and thus is particularly difficult to discern whether those trajectories correspond to different classes. The expended space including a nonlinear combination of variables and their delayed version as new axes (for instance the z variable) can disambiguate the system’s dynamics, enabling robust classification using simple (low-parametric) models.

For instance, in Balaguer-Ballester et al. (2011), we used simple discriminants (with the aid of *kernel*-representations) for successfully predicting the rodent’s behaviour in maze based on neurophysiological recordings. These ideas can be applied to other complex scenarios and thus potentially provide solutions to particularity problematic problems in different industrial settings.

3 Conclusions

This study presents three examples of successful knowledge transfer in very varied industrial domains. Towards that goal, a wide range of machine learning algorithms specialized in temporal patterns processing was used. These approaches helped to improve the existing strategies in e-commerce, supported centres management or targeted marketing scenarios, and others. These novel functionalities in data mining and artificial intelligence were implemented in existing or new commercial software products of TISSAT©. As a result, the company strengthened its position in these strategic sectors. Nevertheless, predictions were much more effective in situations where many system variables and large amounts of data were accessible. This work suggests an alternative approach designed for classification problems based in short nonlinear time series in noisy conditions; which has been recently used in neuroscientific applications. These ideas could help in the successful completion of projects in industrial environments where standard approaches are unfeasible.

References

Abe, N., Verma, N., Schroko, R., Apte, C.: Cross channel optimized marketing by reinforcement learning. In: Proceedings of the KDD, pp. 767–772 (2004)

Balaguer-Ballester, E., Camps-Valls, G., Carrasco-Rodriguez, J.L., Soria, E., del Valle-Tascon, S.: Effective one-day ahead prediction of hourly surface ozone concentrations in eastern Spain using linear models and neural networks. Ecological Modeling 156, 27–41 (2002)

Balaguer-Ballester, E., Lapish, C., Seamans, J., Durstewitz, D.: Attracting dynamics of frontal cortex ensembles during memory-guided decision making. PLoS Computational Biology 7(5), e1002057 (2011), doi:10.1371/journal.pcbi.1002057

Balaguer-Ballester, E., Soria, E., Palomares, A., Martín-Guerrero, J.D.: Predicting service request in support centres based on nonlinear dynamics, ARMA modelling and neural Networks. Expert Systems with Applications 34, 665–672 (2008)

Carberry, S.: Techniques for Plan Recognition. User Modeling and User Adapted Interaction 11, 31–48 (2001)

Carpenter, G.A., Grossberg, S.: ART2: Self-Organization of Stable Category Recognition Codes for Analog Input Patterns. In: Pattern Recognition by Self-Organizing Neural Networks. MIT Press (1991)

Duda, R.O., Hart, P.E., Stork, D.G.: Pattern classification. John Wiley and Sons (2001)

Fu, Y., Shandu, K., Shih, M.: Fast clustering of web users based on navigation pattern. In: Proceedings of SCI 1999/ISAS1999, Orlando, USA (1999)

Gómez-Pérez, G., Martín-Guerrero, J.D., Soria-Olivas, E., Balaguer-Ballester, E., Palomares, A., Casariego, N.: Assigning discounts in a marketing campaign by using reinforcement learning. Expert Systems with Applications 36, 8022–8831 (2009)

Hill, W., Stead, L., Rosenstein, M., Furnas, G.: Recommending and Evaluating choices in a virtual community of use. In: CHI 1995: Conference Proceedings on Human Factors in Computing Systems, Denver, USA, pp. 194–201 (1995)

Kohonen, T.: Self-Organizing Maps, 2nd edn. Springer, Berlin (1997)

Martín-Guerrero, J.D., Balaguer-Ballester, E., Camps-Valls, G., Palomares, A., Serrano-López, A.J., Gómez-Sanchís, J., Soria, E.: Machine Learning Methods for One-Session Ahead Prediction of Accesses to Page Categories. In: De Bra, P.M.E., Nejdl, W. (eds.) AH 2004. LNCS, vol. 3137, pp. 420–424. Springer, Heidelberg (2004)

Martín-Guerrero, J.D., Lisboa, P.J.G., Palomares-Chust, A., Soria, E., Balaguer-Ballester, E.: An approach based on Adaptive Resonance Theory for analyzing the viability of recommender systems in a citizen web portal. Expert Systems with Applications 33, 743–753 (2007)

Martín-Guerrero, J.D., Soria, E., Gómez-Sanchis, J., Soriano-Asensi, A., Palomares, A., Balaguer-Ballester, E.: Studying the feasibility of a recommender in a citizen Web Portal based on user modeling and clustering algorithm. Expert Systems with Applications 30, 299–312 (2006)

Pfeifer, P.E., Carraway, R.L.: Modeling customer relationships as markov chains. Journal of Interactive Marketing 14, 43–55 (2000)

Reichheld, F.F.: The loyalty effect: The hidden force behind growth, profits, and lasting value. Harvard Business School Press, Boston (2001)

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J.: An Open Architecture for Collaborative Filtering of Netnews. In: Proceedings of the Conference on Computer Supported Cooperative Work, pp. 175–186. Chapel Hill (1994)

Sauer, T., Yorke, J., Casdagli, M.: Embedology. J. Stat. Phys. 65, 579–616 (1992)

Schafer, J.B., Konstan, J., Riedl, J.: Recommender Systems in E-Commerce. In: Proceedings of the First ACM Conference on Electronic Commerce EC 1999, Denver, USA, pp. 158–166 (1999)

Smith, A.J.: Applications of the self-organising map to reinforcement learning. Neural Networks 15, 1107–1124 (2002)

Sutton, R.S., Barto, A.G.: Reinforcement learning: An introduction. MIT Press, Cambridge (1998)

Takens, F.: Detecting strange attractors in turbulence. Springer lecture notes in mathematics, vol. 898, pp. 366–381 (1981)

Vapnik, V.N.: The nature of statistical learning. Springer, New York (1999)

Zukerman, I., Albrecht, D.W.: Predictive Statistical Models for User Modeling. User Modeling and User Adapted Interaction 11, 5–18 (2001)