DATA FUSION FOR TREND IDENTIFICATION IN LARGE RETAIL BUSINESSES USING FUZZY TECHNIQUES

Alessandra Orsoni, Sarah Viazzi, Marina Massei
DIP University of Genoa, Via Opera Pia 15, 16145 Genoa, Italy. Tel. +39 010 353 2883 Fax +39 010 31 77 50,
Email aorsoni@itim.unige.it

Chiara Briano, Matteo Brandolini,
Liophant Simulation Club, Via Molinero 1, 17100 Savona, Italy, Tel +39 019 97398, Fax +39 019 97600
Email Liophant@itim.unige.it

KEYWORDS

Data Classification, Dempster-Shafer, Bayesian, Fuzzy Logic, VV&A.

ABSTRACT

The work discussed in this paper is a response to major perceived needs for higher organizational standards in the retail business infrastructure. In this context effective planning of resources strongly relies on the ability to understand customer behavior and market trends based on high volumes of data made available at different levels of detail by different sources. The objective is to integrate sales information from heterogeneous sources and to combine this data using fuzzy techniques, along with traditional ones, in order to identify incipient sales trends and customer behavior patterns, tracking their evolution over time for improved store organization and planning. A set of software tools was developed and validated in order to build a model architecture capable of gathering and processing information from different sources, such as cashier's receipts, individual customers (through store membership/advantage cards), and aggregate weekly sales indicators across stores located in a geographic area of interest, either individually, by source type, or in combinations.

INTRODUCTION

Fuzzy Logic techniques are key components for several Data Fusion applications because of their ability to account for the random behavior of the systems under observation, as well as the uncertainty associated to the output of the designated observation tools. The application of intelligent systems in high-level fusion is considered a strategic requirement for the effective integration of the main data fusion entities. This study is based on a long-term collaboration project between DIP, the University of Genoa, and one of the largest Italian retail chains. This phase of the project develops a software package to simulate and evaluate the effectiveness of Fuzzy techniques in fusing highly heterogeneous data gathered from multiple sources.

FUNCTIONAL ARCHITECTURE FOR DATA FUSION

A large variety of application-specific hardware and software architectures are currently available for the functional implementation of Data Fusion. Based on prior experience and knowledge about the features of various functional models, the authors propose a specific architecture to address the problem of data fusion for sales indicators obtained from multiple and diversified sources. These sources include cashier's receipts, individual customers (through store membership/advantage cards), and aggregate weekly sales indicators. Particular attention was paid to the detection, tracking, and classification of sales indicators by product clusters, identified as canned/packaged food, fresh meats, sandwich meats, fresh fish, and ready-made gastronomy. Aggregate sales indicators include cumulative weekly revenues, number of cashier's receipts, and average customer spending per visit on each product group.

The experience developed in collaboration with other research centers led to the development of a Data Fusion system that would operate according to the following logical steps:

- Access to Company Databases
- Knowledge of Pre-existing Seasonal and Daily Sales Fluctuations
- Acquisition of New Sales Records and Aggregate Indicators
- Data Association
- Sales Tracking
- Patterns Identification

Established methodologies for data association and patterns identification are based upon mathematical statistical techniques, the effectiveness of which has been demonstrated through well-known experiments involving the Bayesian and the Dempster-Shafer approaches. Instead, variables tracking and the estimation of dynamic parameters are best addressed using a Kalman filter, a traditional but reliable method.

The experimental data gathered on the field from different sources, whether they are cashier's receipts, individual customers' spending amounts, or aggregate sales indicators, represent complementary views of the same image: from the individual store's perspective, from the individual customer's perspective and from the whole company's perspective, respectively. As the level of detail is increased, moving from aggregate and averaged data, to punctual sales figures, the noise and the fluctuations of the data over time become too significantly high for trend identification and pattern definition: noise and fluctuations may actually disguise the presence of a trend or pattern.

Current data fusion procedures are particularly well-suited for processing images based on pre-defined characteristics which can be extracted out of series of pictures taken at different times and/or from different perspectives. This approach normally ensures high computational efficiency because it is based on the processing of summarized information. In order for the summary to be effective, modern techniques are often required, such as the fuzzyfication of the image's attributes and their evaluation against fuzzy classes considering both significance and variability.

In the functional implementation of the proposed architecture, only some features of the data series were selected as relevant to the analysis, through a comparison with the experiences of other research centers. In the specific case, the authors considered the following parameters extracted from each data set by source type

- Temporal Horizon of Analysis
- Extension and Location of Geographic Area
- Absolute Values of Detected Sales Peaks
- Position of the Highest Peak
- Number of Peaks Detected
- Distance between the two Highest Peaks Detected

Further research involving the use of neural networks to ensure self-learning during the fuzzy reference classes definition could be considered as a possible extension of this work.

In the proposed architecture variables' fuzzyfication can be either enabled or disabled for the purposes of evaluating its effectiveness.

As far as the mathematical statistical techniques are concerned, the tool provides the option of following either the Bayesian or the Dempster Shafer approach.

OUTCOME TESTING THROUGH SIMULATION

The Fuzzy Data Fusion method can be interfaced to a business simulator for outcome testing and verification purposes. For instance, the identified sales trends and consumers' spending patterns may be used to generate an initial scenario, and their actual developments may be tested over time in a dynamic simulation environment, responsive to dynamic changes in the external influencing factors (e.g. fluctuations in the market price of goods).

The proposed simulation model generates projections of all the relevant sales indicators by tracking their temporal dynamics, for each product group, in the real context considering the limited availability of resources (i.e. lack of personnel and equipment failures) and the probability of product stock-outs. The stochastic nature of these variables can be accounted for using the Montecarlo technique: each random phenomenon can be represented through repeated extractions out of the corresponding probability distributions. The simulator can also be used to test inconsistent scenarios. This type of test simulates and verifies hypotheses corresponding to erroneous information, false alarms and special consumers' behaviors. The system was developed in C++ to ensure maximum operating efficiency. As shown in figure 1 the simulator not only reproduces real sales projections building off identified trends, but it also brings out critical scenarios that can then be further analyzed. The simulator generates both the set of input for the Demonstrator and the reference data for performance evaluation.

DEMONSTRATOR DEVELOPMENT

As previously mentioned for the Simulator, the Demonstrator is also directly integrated into the model architecture. After completing the data acquisition phase from the integrated Databases, the demonstrator continues the elaboration and supports detection/tracking and the classification of the acquisitions obtained.

The Bayesian statistical-mathematical methodology also includes a priori estimates regarding the probability of encountering a specific trend in the scenario, while all the information about tolerances and reliability of the data sources are read off the same file

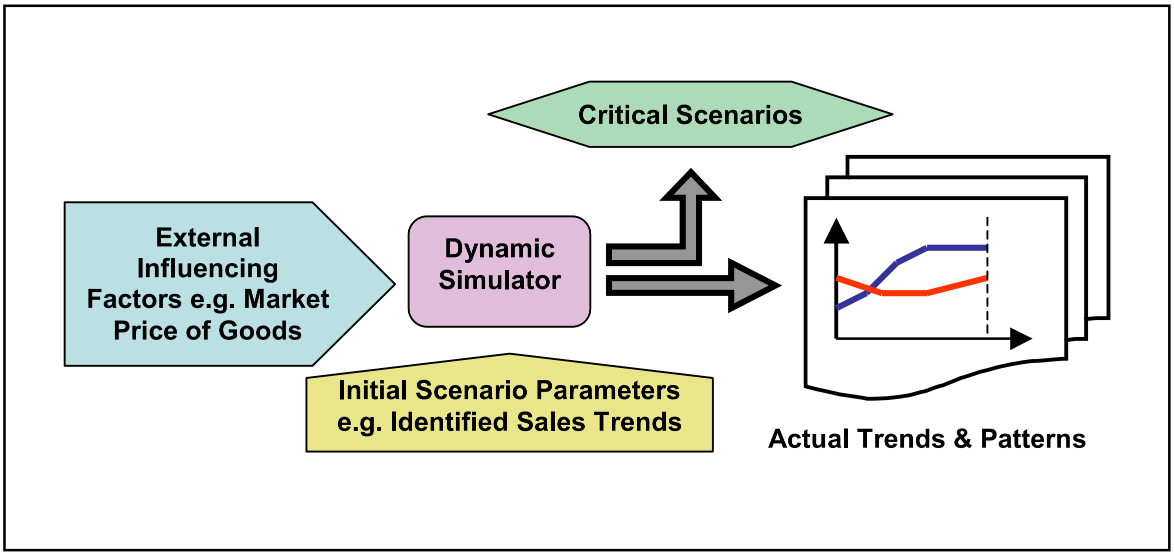

Figure 1: Functional Schematic of the Dynamic Simulator

that the simulator uses to reproduce the randomness of the observations. In addition to these static database objects, the demonstrator also interacts with a third category of objects, namely, the data provided by the simulator.

The demonstrator first identifies the particular trend indicator that each piece of information made available belongs to. It then attempts to combine the new observation with any of the previous ones to improve the estimate of the corresponding characteristics, especially data on rates of change over time, to support trend identification and pattern definition.

The demonstrator processes the information with the Kalman Filter for the current indicator's value and calculates an estimate of its new value, assuming that the information required to process it is available.

As shown in figure 2, the demonstrator fuses the new observed value with the estimated one through the Kalman Filter and maintains the current range estimate as the valid reference.

The next step for the demonstrator is the fusion needed for trend identification. The procedure is the same as the previous one, even though it is more difficult: the various characteristics are processed, in this case for identification, using the Dempster-Shafer method or the Bayes rule, and the fuzzyfication classes, if necessary. They are based on the fact that they measure the membership of such classes, instead of using binary logic (true/false, or membership probability 0/1), as probabilistic membership in the tolerance segments through *membership functions*, built around the existing segments. If several points are present, it will be determined which one is "more" of a member and which one is "less", with triangular linear membership functions.

VERIFICATION AND VALIDATION OF THE LOGICAL MODEL

Demonstrator validation is supported by graphical functions and by detailed reports. In fact, once the data concerning the simulation scenario are loaded and the processing performed through the demonstrator, we obtain the results of the gating and association techniques. Then the simulated scenario can be graphically represented in the main window. If the demonstrator has not yet been run, the graph will represent the real values meaning only the observed values, but not the demonstrator estimates since they have not yet been generated according to previously defined graphic settings. The display can be used to check the coherence of the results with the data set and simulated scenario, using different colors to represent different indicators.

Some parameters and methods used by the demonstrator can be configured to check performance. In fact, it is possible to choose various types of data sources and sales indicators to be implemented in the data fusion and the resolution algorithm for data association and identification of trends and patterns. Similarly, the processing algorithm can be chosen: the *Bayes* logic or the *Dempster-Shafer* method (in a mutually exclusive way) and finally, decide whether or not to use *Fuzzy Logic*.

By applying experimental techniques and statistics, an analysis was performed on the product groups considered significant for testing on the demonstrator for the implementation of experimental analysis techniques (DOE Design of Experiments). The purpose of such an analysis was to verify the influence of various factors on some objective functions representing the efficiency of the Data Fusion

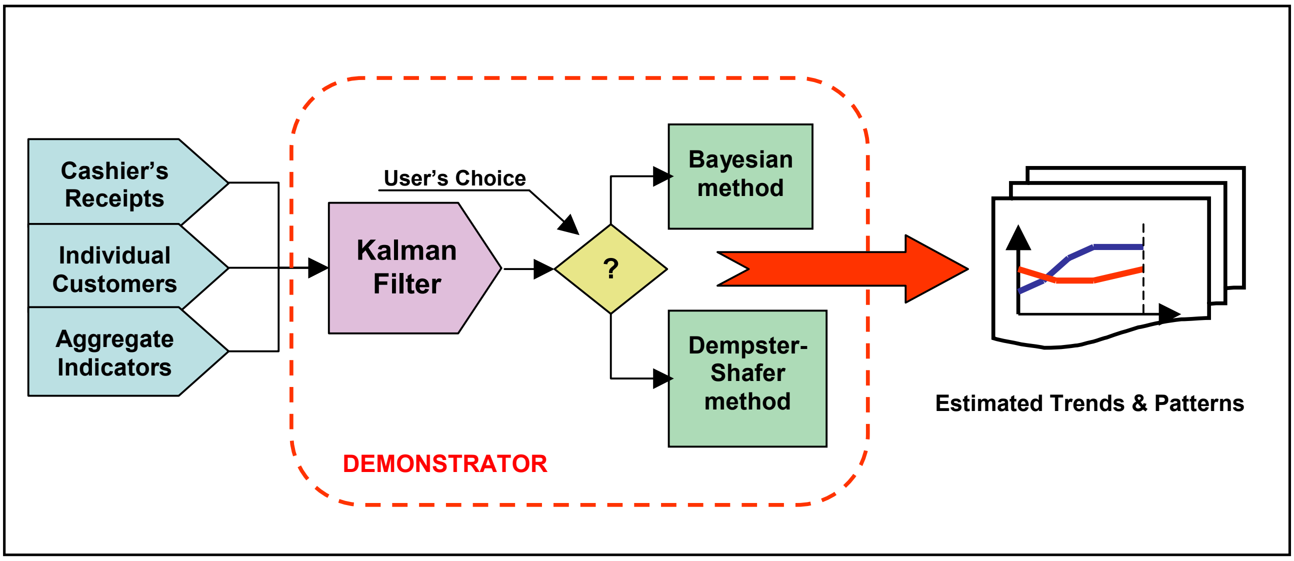

Figure 2: Functional Schematic of the Demonstrator

architecture developed to associate data and identify trends and patterns.

These methods involve an analysis of the influence of various independent variables, and their higher-order combinations, through the development of an experiment based on a factorial design. All the tests corresponding to the combinations of independent variables in the various levels will be performed in this design.

The experiment was based on a set of functions offered by the demonstrator that may or may not enable certain algorithms or process specific information

The objective of the analysis was to determine the influence of the factors and their interactions on three objective functions, and specifically:

- *Unreliability* Index: ratio between the number of trends not recognized and the number of trends for a given scenario.
- *Robustness* Index: the lowest probability assigned to a trend correctly recognized for a given scenario.

Preliminary sensitivity analysis shows that the use of Fuzzy Logic has the greatest impact on reliability. This demonstrates the accuracy of the basic Data Fusion concept: the proper fusion of data coming from various sources generates significantly greater benefits than the generalization of information from single sources.

Similar considerations apply to the test results relative to the influence on the minimum probability of correct identification.

In addition to this sensitivity analysis, further testing may involve response surfaces generated using the RSM (Response Surface Methodology) in which regressive meta-models were built based on the experimental data of factorial designs. These surfaces are obtained by evaluating the influence of the activation or non-activation of the various independent variables already considered in the sensitivity analysis in terms of use level, i.e. by varying the number of detections processed with that methodology.

CONCLUSION

This work explores new areas of research in the field of data fusion. In particular, this research examines the potential of advanced artificial intelligence techniques in the data fusion application context, focusing on fuzzy logic techniques, which are widely employed in information processing and in the formalization of logical rules.

The large number of applications developed at the international level show that the effectiveness of the integration of fuzzy techniques in the field of data fusion has been scientifically recognized. Experimental analysis confirms these observations, however, there are key aspects of this project which make the development of a simulator/demonstrator and the related research quite innovative.

The development and use of the demonstrator and simulator has shown the benefits of applying fuzzy logic techniques to the fusion of data coming from multiple heterogeneous sources. These benefits include greater robustness and higher reliability. Further improvements in performance outcomes may be sought including genetic algorithms or other artificial intelligence techniques to produce efficient decision support tools, while the use of neural networks may be tested for system self-learning.

The existing architecture was tested in the field using simulation and design of experiments as tools to verify theoretical scenarios. The corresponding results have shown that the fusion of data from multiple heterogeneous sources significantly influences performance in terms of efficiency and robustness of trend identification and pattern definition.

Further work will involve more in-depth experiments on the tool, with particular attention to the simulation model in order to enhance its efficiency and functionality; the purpose of this additional testing will be to facilitate the process of creating complex scenarios including the functions to be used in an analysis of such scenarios.

Ideally, the platform developed will provide the context to experimentally test various architectures and methodologies related to the possible scenarios and external/boundary conditions.

REFERENCES

Anthony R.T. (1995) "Principles of Data Fusion Automation", Artech House, Norwood, MA

Blodgett D. (2001), "COMDAT I, Command and Control Functional Architecture", Combat System and Engineering, n°65200801, April 2001

Brooks R.R., Iyengar S.S. (1998) "Multi-Sensor Fusion", Prentice-Hall, Upper Saddle River, NJ

Bruzzone A.G. (1995) "Fuzzy Logic and Genetic Algorithms Applied to the Logistical and Organisational Aspects of Container Road Transports", *Proc. of ESM95*, Praha, June 5-7

Bruzzone A.G., Colla G., Mosca R., Scavotti A. (1998) "Simulation as Verification Support for Qualitative Risk Analysis in Industrial Facilities based on Fuzzy Logic", *Proceedings of Emergency Management Conference*, Boston, 4-9 April

Bruzzone A.G., Cotta G., Cerruto, M.E. (1997) " Simulation & Virtual Reality To Support The Design Of Safety Procedures In Harbour Environments ", *Proc.of ITEC97*, Lausanne (CH) Bruzzone A.G., Giribone P. (1998) "Artificial Neural Networks as a Meta-Modelling Techniques for Car Component Industrial production", *Proc. of Neurap 98*, Marseilles

Bruzzone A.G., Giribone P. (1998) "Decision-Support Systems and Simulation for Logistics: Moving Forward for a Distributed, Real-Time, Interactive Simulation Environment", *Proceedings of the Annual Simulation Symposium IEEE*, Boston, 4-9 April

Bruzzone A.G., Giribone P., Mosca R. (1996) "Simulation of Hazardous Material Fallout for Emergency Management During Accidents", *Simulation*, vol. 66, no.6, June, 343-355

Bruzzone A.G., Kerckhoffs (1996) "*Simulation in Industry*", SCS Europe, Genoa, Italy, October, Vol. I & II, ISBN 1-56555-099-4

Bruzzone A.G., Mosca R. (1998) "Special Issue: Harbour and Maritime Simulation", Simulation, Vol.71, no.2, August

Bruzzone A.G., Uhrmache A., Page E.H. (1999) "Web-Based Modeling and Simulation", SCS, San Francisco, CA, ISBN 1-56555-156-7

Bruzzone, A.G. (1994) "Potential of Reduced Experimental Designs in Object-Oriented Modeling of a Manufacturing Department for Film Production", *Proceedings of ESM94*, Barcelona, SP, June 1-3

Cox E. (1994) "The Fuzzy System Handbook", AP Professional, Chestnut Hill, MA

Dubois D. Prade H., (1980)"Fuzzy Sets and Systems: Theory and Applications", Academic Press, New York

Endsley, M.R. (1995) "Toward a Theory of Situation Awareness in Dynamic Systems" Human Factors, Vol.37 No. 1 pp.32 – 64.

Gagnon L., Oppenheim J.R., Valin P. (1998) "R&D Activities in Airborne SAR Image Processing/Analysis at Lockheed Martin Canada", *Lockheed Martin Technical Report*, Montreal

Giribone P. & Bruzzone A.G. (1995) "Neural Networks And Fuzzy Control For Bulk Terminal Operating Management", *Proc. of Summer Simulation '95*, Ottawa, July 23-26

Giribone P. & Bruzzone A.G. (1995) "Object-Oriented Modeling of an AGV System for a Harbor Environment", *Proceedings of WMC95*, Las Vegas, Nevada (USA), January 15-18

Giribone P., Bruzzone A.G. (1999) "Artificial Neural Networks as Adaptive Support for the Thermal Control of Industrial Buildings", International Journal of Power and Energy Systems, Vol.19, No.1, 75-78

Giribone P., Bruzzone A.G. (1997) "Design of a Study to use Neural Networks and Fuzzy Logic in Maintenance Planning", *Proceedings of Simulators International XIV*, SMC'97, Atlanta, Georgia, April 6-10

Hollenback J.W., Kaminski P.G. (1996) "M&S VV&A DoD Directive 5000.61 Best Practice", DMSO, Washington

Hung T. Nguyen, Elbert A. Walker (2000), "A First Course in Fuzzy Logic", Chapman & Hall/CRC

ISIF (2000) "Proceedings of the 3rd Annual Conference on Information Fusion", Lockheed Martin Canada

ISIF (2001) "Proceedings of the 4th Annual Conference on Information Fusion", Lockheed Martin Canada

Jameson S. (2001) "Architectures for Distributed Information Fusion To Support Situation Awareness on The Digital Battlefield" *Proceedings of The Annual Conference on Information Fusion* "Fusion 2001", LM Canada

Jang J.S.R., Gulley N., (1995) "Matlab Fuzzy Logic Toolbox", The Math Works Inc., Natick, MA

Kadar I., Libby V. (1994) "Signal Processing, Sensor Fusion, and Target Recognition III", SPIE, Bellingham, Washington

Klein L.A. (1999) "Sensor and Data Fusion Concepts and Applications", SPIE, Bellingham, Washington

Lefebvre E., Helleur C., (2001) "Multisource Information Adaptive Fuzzy Logic Correlator for Recognized Maritime Picture" *Proceedings of The Annual Conference on Information Fusion* "Fusion 2001", LM Canada

Merkuriev Y., Bruzzone A.G., Novitsky L. (1998) "*Modelling and Simulation within a Maritime Environment*", SCS Europe, Ghent, Belgium, ISBN 1-56555-132-X

Molina Lopez J.M., Jiminez Rodriguez F.J., Casar Corredera J.R. (2001), "Negotiation in a Multi – Agent System of Surveillance Sensors Based on Fuzzy Priorities" *Proceedings of The Annual Conference on Information Fusion* "Fusion 2001", LM Canada

Molina Lopez J.M., Jiminez Rodriguez F.J., Casar Corredera J.R. (1995), "Fuzzy Reasoning for Multisensor Management" *Proceedings of IEEE International Conference on Systems, Man and Cybernetics 1995, Vol. II*

Montgomery D.C. (1997) "Design and Analysis of Experiments", Wiley

Shahbazian E. Bergeron P., Duquet J.R., Jouan A., Valin P. (1999) "Data Fusion Applications for Military and Civilian Purpose Developed on DND/LM Canada Decision Support Testbed", *Lockheed Martin Technical Report*, Montreal

Shahbazian E., Blodgett D., Labbè P. (2001) "The Extended OODA Model For Data Fusion Systems", *Proceedings of The Annual Conference on Information Fusion* "Fusion 2001", LM Canada

Shahbazian E., Duquet J.R., Macieszczak M., Valin P. (1998) "A Generic Expert System Infrastructure for Fusion and Imaging Decision Aids", *Proceedings of EuroFusion98*, Bournemouth, UK

Simard M.A., Lefevbre E., Helleur C. (2000) "Multisource Information Fusion Applied to Ship Identification for the Recognised Maritime Picture", *Proceedings of SPIE2000*, Orlando

Simard M.A., Valin P., Lesage F. (1998) "The ID Box: A Multisource Attribute Data Fusion Function for Target Identification", *Lockheed Martin Technical Report*, Montreal

Stone L.D., Barlow C.A., Corwin T.L. (1999) "Bayesian Multiple Target Tracking", Artech House, Norwood, MA

Valin P., Jouan A., Bosse E. (1999) "Fusion of Imaging and Non-Imaging Sensor Information for Airborne Surveillance", *Lockheed Martin Technical Report*, Montreal

Waltz E., Llinas J. (1990) "Multisensor Data Fusion", Artech House, Norwood, MA

Wang L.X. (1997) "A Course in Fuzzy Systems and Control", Prentice Hall, Upper Saddle River, NJ