SimMarket: Multiagent-Based Customer Simulation and Decision Support for Category Management

Arndt Schwaiger and Björn Stahmer

DFKI GmbH
Stuhlsatzenhausweg 3
66123 Saarbrücken
{Arndt.Schwaiger, Bjoern.Stahmer }@dfki.de

Abstract. A key to an optimal assortment of goods and pricing of individual items in a store is the knowledge about potential customer's behaviour. In this paper we present the simulation of individual customers based on a *multiagent system* which models the important elements and external influences as single agents. An agent can be member of several agent groups which are represented as *holons*. We model each individual customer as an agent which behaves according the customer's individual preferences. These preferences are extracted from real world data, such as customer cards, sales data and interviews. The customer's shopping behaviour is represented in *behaviour networks* (Bayesian nets) which are stored in the customer agents' knowledge bases. The behaviour of a representative group of customers induces the overall sales figures, which support decisions what to sell at which price. The presented concepts are based on ideas of Joachim Hertel from DACOS and Jörg Siekmann from the DFKI. They are implemented as a prototype, which provides, after further evaluation, the basis for a new and final system to be used by retailers.

1 Introduction

Increasing globalisation helps international retail chains to gain more and more market share and more pressure is put onto local competitors. This is especially true for supermarkets and drugstores which have to develop new strategies to remain competitive. An essential part of the competitive position is the optimization of assortments and prices. This is often done by Category Managers, who are making the decisions concerning product prices, promotions, placement and assortment of goods which are crucial for the success or failure of an individual store.

In the daily decision making process there are many different alternatives and combinations from which the person in charge has to map out an optimal strategy. The biggest problem is the enormous amount of alternatives and the complexity of internal relationships and external influences, such as competitor's prices and promotions, market trends, the economic situation and weather conditions. An average supermarket sells several thousand different items which could be selected out of up to half a million of products. These numbers alone illustrate the "combinatorial explosion" of possible combinations quickly exceeding human capabilities. In addition, most of the changes concerning a single item, such as price increase or reduction are

M. Schillo et al. (Eds.): MATES 2003, LNAI 2831, pp. 74–84, 2003.
© Springer-Verlag Berlin Heidelberg 2003

likely to interfere with the sales of other items. The decision processes are usually supported by a computer but the support is mostly limited to the provision of information about current market figures and the generation of some sales forecast which are unsuitable for "*what-if*"-scenarios.

A number of approaches have been considered, e.g. regress analysis [14] in economics, which provides means for a more accurate forecast. More recent approaches in computer science include the use of neural networks [14, 13]. Yet, all these approaches have significant disadvantages: regress analysis provides us with acceptable quantitative results; however, it allows for just a few factors to be included in the analysis and can only treat linear dependencies. Neural networks can cope with a much higher number of influencing factors and while single-layered networks can only model linear dependencies, a multi-layer neural network can even model non-linear dependencies. The problem with neural networks is their fast rise in complexity which renders them unsuitable when augmenting the number of factors. As shown in [1] the possible existence of local minima and the large dimensional degree of the weights makes it increasingly difficult to train large multi-layer networks. Furthermore, even a small neural network becomes a black box in the sense that it takes an input and returns an output without providing the user with any insight on how the various factors correlate.

But more important is the fact that both approaches simply predict future values out of past experience without a deeper understanding of how the individual factors correlate. To sum up these approaches can support the decision making process only up to a certain extent resulting in suboptimal solutions to be optimized by man.

A fully satisfactory approach to this problem must be able to model and simulate the individual customer's behaviour as a reaction to both internal and external influences. Given perfect information on the individual behaviour of a customer we could then predict the overall behaviour of a group by considering the sum of the predictions for all individuals. For a large group it would even suffice to consider a representative subgroup to determine the likely overall reactions to influences like promotions or changes in price or location. The SimMarket approach does just that. It simulates a representative set of individual behaviours by modelling each individual customer through an individual software agent. Note that the behaviour is determined from real world data by using special learning algorithms to extract the relevant information where "real world" refers to data gathered by a store on its human customers. The agent architecture uses also known models from psychology and economics. It models the relevant entities and external factors of a supermarket in order to create a test scenario, where the agents react to the scenario according to their individual behaviour patterns. As these patterns are derived from the actual behaviour of individual human customers the system is capable of predicting realistic results. The data collection has been arranged by DACOS and is done in a partnership with three modern German retailers (tegut, dm-Markt and Globus) to ensure the desired level of realism. For example, the (anonymised) data contain information about customer card usage, sales, distribution, detailed information about individual products or promotions. We can furthermore include the experience of the responsible assortment and pricing managers in the system to guarantee realistic evaluation and testing of the system.

Current research on agent based customer modelling and simulation includes the following: *Consumat* by W. Jager [2] and the *SIMSEG* project [3] of the Vienna University of Economics and Business Administration as they share some conceptual ap-

proaches with SimMarket. The Consumat model is an agent-based implementation of psychological meta-models based on a number of customer behaviour relevant theories which unfortunately uses an abstract representation of the actual act of purchasing due to which it is not suited for practical use.

The SIMSEG project is a simulation environment for the analysis of market partitioning and strategies for positioning based on an abstract model which has to be designed manually. Its primary aim is to determine the results of corporate image chance with respect to the shopping behaviour on a larger level.

The area of user modelling which has a long history at DFKI [8], is closely related to the area of customer modelling. Multiagent based user modelling provides a rich basis of concepts for representing patterns of human behaviour in agents. Schaefer [9] successfully uses dynamic Bayesian networks [10]to model users of dialog systems and to personalise the application flow.

2 SimMarket Architecture

The aim of the SimMarket project is to develop a realistic representation of a single store or supermarket including the goods and customers to be modelled as agents.

We use a *holonic* multiagent system (MAS) which supports agent groupings as *holons* so that each individual agent can be a member of different agent groups. A *holon* is an agent that represents a group of agents whereas each group member keeps its individuality [4]. In addition the MAS is connected to a *data warehouse*[1] which stores real data from the related supermarket.

Fig. 1 shows the current architecture where each element is modelled as an agent. The whole supermarket is also represented by a *supermarket agent* which provides detailed information about the specific store, such as assortment of goods, item prices and promotions, shelf layouts, sales data, costs and so on.

The *customer agents* model the individual customers and their personal shopping behaviour. The system is able to represent each customer individually as well as an arbitrary number of (possibly not disjoint) customer groups and subgroups. The largest group of customers, the group of all customers, is represented by a *meta customer agent*.

Similarly all items are represented as *item agents* which are grouped into predefined commodity groups modelled by *commodity group agents*. In addition, items can be dynamically classified in order to generate item groups such as private brand items, all items which are lower than *n* Euros and so on. The group of all items is represented by the *meta item agent*.

The architecture of the MAS supports also *competitor agents* which model relevant competitors. These agents provide information about the competitor's prices, promotions and strategies. They may have links to common customers and items.

In forthcoming versions we want to add *manufacturer agents* which mainly provide information about the current range of products, their prices and promotions as well as details about forthcoming products.

[1] „A Data Warehouse is a subject-oriented, integrated, non-volatile, and time variant collection of data in support of managements decisions. " – Bill Inmon (1996)

Important external effects and influences e.g. current season, economic situation and weather conditions are represented by *environment agents*. These are information agents which collect and administrate the relevant data from external services or the internet. They generate simple forecasts or (if provided with insufficient data) request more detailed prognosis from external services (such as e.g. the daily weather forecast).

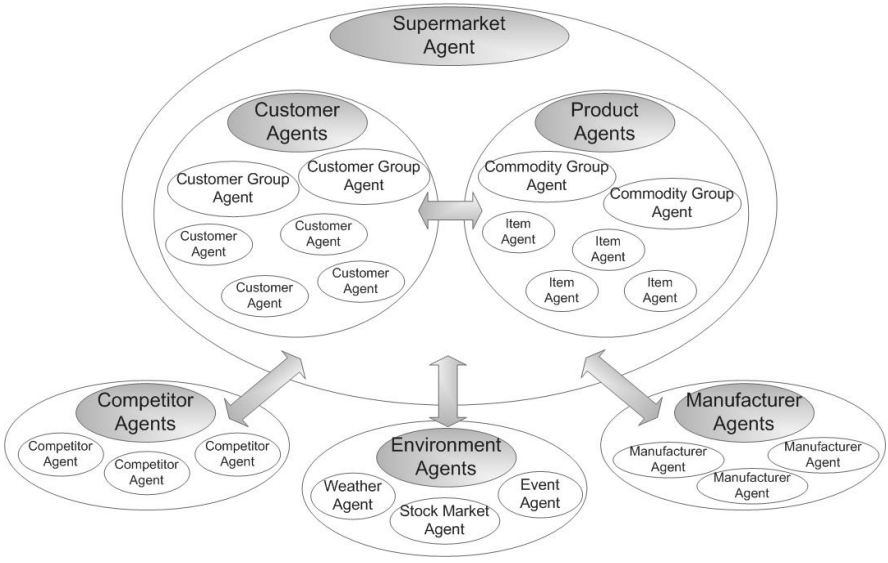


Fig. 1. The Holonic Multiagent Architecture of SimMarket

3 Customer Agent Architecture

The predictions of the shopping behaviour and preferences of customers are represented in the *customer agents*. Any individual customer that is known to the supermarket through empirical surveys, the use of customer cards, electronic payment or direct contact is represented as an individual customer agent.

Fig. 2 shows the model of a customer agent which administrates the knowledge and information about a related customer and his behaviour. The knowledge consists of a *personal profile* (information about age, income, gender, domicile etc.) and a special model of the customer's individual buying preferences.

We consider the customer's behaviour as an interaction of many different factors e.g. characteristics, addictions, needs and environmental influences. These factors may correlate or even contradict each other. For example, it is interesting whether the quality sensitivity outweighs the price sensitivity for a certain offer or not, i.e. whether a very low price can "persuade" a customer to purchase a low grade product or not.

Metaphorically speaking the characteristics of a customer compete for the greatest influence on the customer's decision. For this reason it is important to consider the customer's characteristics and factors themselves as *feature agents*. Each feature agent represents a single factor of the customer's behaviour, such as price and promotion consciousness or brand loyalty. Collectively, these agents form a multiagent system of their own, the *society of feature agents*. From this point of view the customer's (shopping) behaviour is the result of the interaction of its feature agents, a view which conforms with Marvin Minsky's "society of mind" [4].

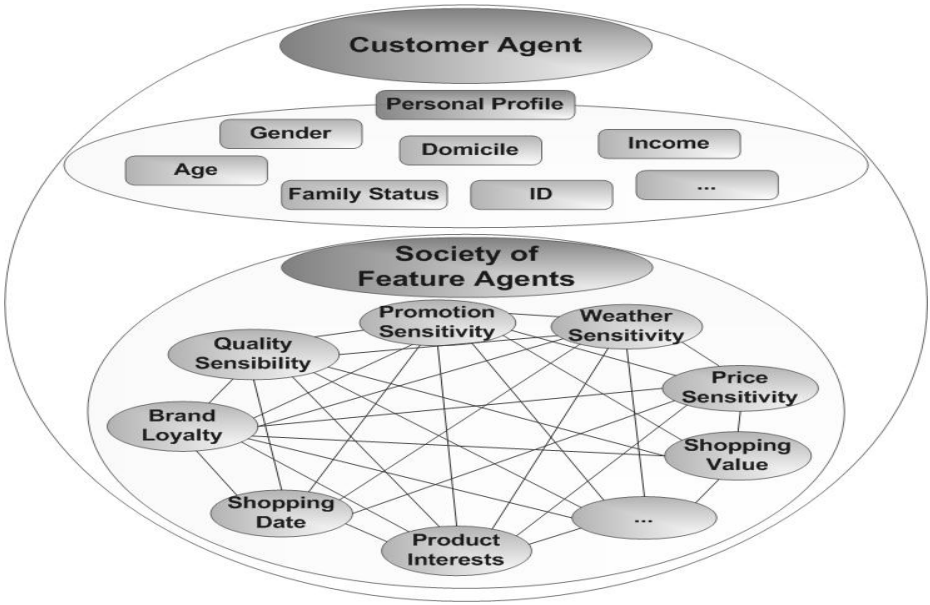


Fig. 2. The Customer Agent Architecture including the *Society of Feature Agents*

3.1 Customer Agent Groups

According to the literature, the classification of a customer can be based on a variety of criteria or similarities, e.g. there are many different concepts and models of customer groups developed in marketing [12] and psychology, such as the "Sinus-Milieus" from Sinus Sociovision [6] or the "Euro-Socio-Styles" from CCA/Europanel [7].

In order to support complex groupings of customers in our system, each customer agent can be a member of one or more customer agent groups (see Fig. 3), which are represented as holonic agents. Thus, we are able to simulate a certain group of customers either by simulating all related individual group members (higher accuracy) or by simulating the average group behaviour (higher computational efficiency) depending on situation at hand.

Customer groups can be built beforehand or dynamically at runtime where the criteria can vary, for examples similarities concerning the personal profile (same age, gender or domicile etc.) or patterns of shopping (similar baskets of goods, same product interests, similar price consciousness etc.) or a combination (Fig. 3).

Evaluating the actual distribution of customers over *typical* customer groups by comparing the shopping data of individual customers to *typical* shopping baskets is our next goal. This will allow for a detailed ratio analysis of the customer groups of a supermarket and thereby to fine-tune the weighing of the group based simulation results.

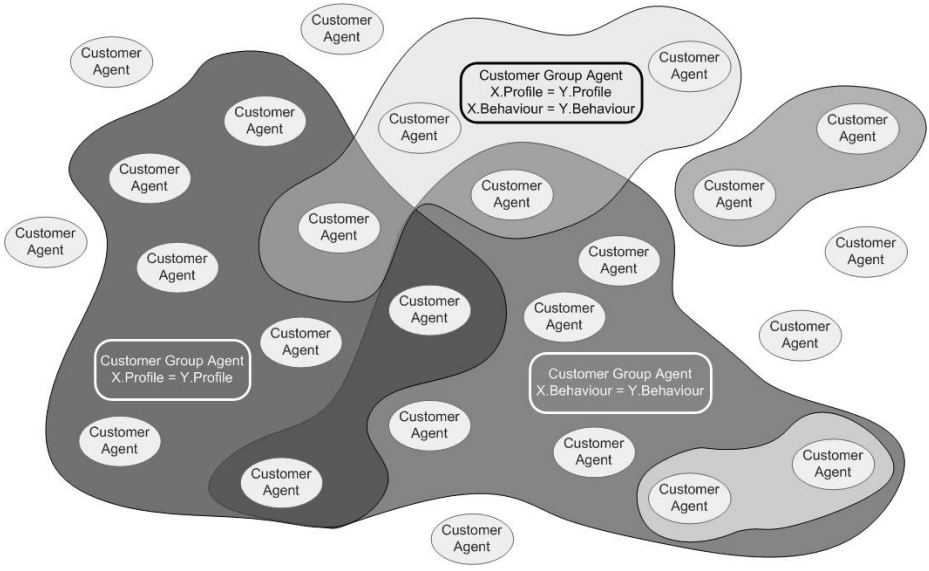


Fig. 3. Holonic Grouping of Customer Agents Concerning Different Criteria

3.2 Behaviour Networks

The empirical data, i.e. the characteristics of individual customer shopping behaviour, was primarily extracted from customer cards data, questionnaires and from general marketing knowledge. It is represented in behaviour networks, i.e. Bayesian nets. The quality of the customer simulation and prognosis depends mainly on how exact and realistic these characteristics and thus the shopping behaviour of individual customers can be modelled.

Our approach is based on a rule system where customer data is encoded in a set of rules which are associated with conditional probabilities. The rules are stored in a specific network structure - the behaviour network - to represent the dependencies between them. Bayesian networks are an example of such a network formalism. Currently, we model the behaviour of a customer by generating a specific Bayesian network for each relevant commodity group. We extract these networks by data mining patterns of behaviour from individual customer data (see Fig. 4). As a source for

modelling the networks, we use anonymous customer cards and sales slip data of the retail company Globus.

In addition, we integrate the individual dependencies of single customers of external influences, e.g. weather, economic situation or seasons into the behaviour networks.

In later versions we will encode the customer's attitudes and habits in a matrix of feature agents. Every feature agent contains - as part of his knowledge base - a specific feature network which simulates the characteristics of a customer with respect to this feature.

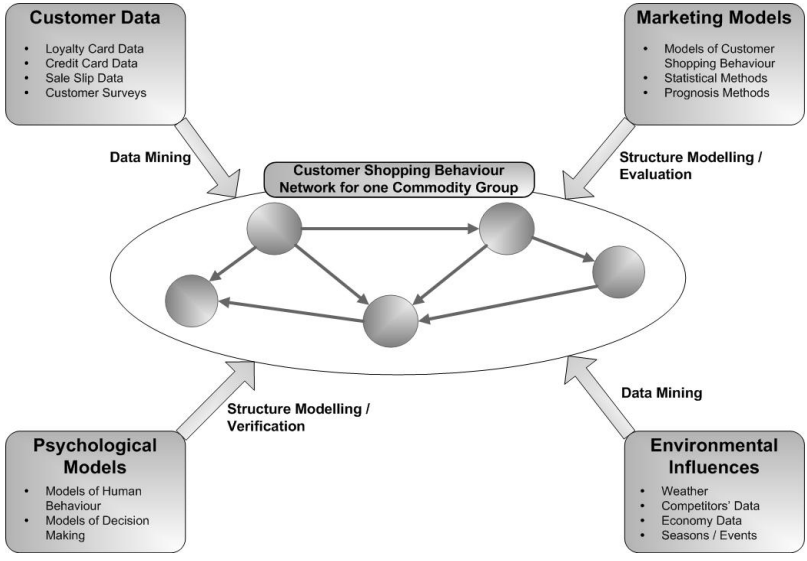


Fig. 4. Generation of the Customers' Behaviour Networks

Fig. 5 shows the user interface of the *SimMarket Customer Manager,* presenting a simple example of a Bayesian network which models the behaviour of a customer concerning a specific commodity group "Brötchen", i.e. bread roll. In this example, we are interested in the customer's price consciousness and his favoured shopping dates. On this account we create a behaviour network with the following nodes representing the price consciousness: "low-priced units", "mid-priced units", "high-priced units" and "value", i.e. how many low-priced, mid-priced and high-priced units the customer is expected to buy depending on the shopping date which in turn is modeled by the nodes "month", "part of month" and "weekday".

The customer's precise individual dependencies are data mined from his real shopping data and can be used for simulation. Therefore the date nodes have to be set accordingly to the desired date. Then the propagation algorithm [11] of the Bayesian network infers the effects on the dependent nodes. The result is summed up as expected values of the numbers of units which the customer is expected to purchase. E.

g. he buys 1 low-priced and 5 high-priced products of the related commodity group on the first Friday in June.

Similarly, it is possible to model more characteristics of the customer's personal behaviour to improve the simulation quality. In a later version each feature agent will contain at least one behaviour network for the related feature. The dependencies between the different characteristics are modeled by negotiation of the feature agents.

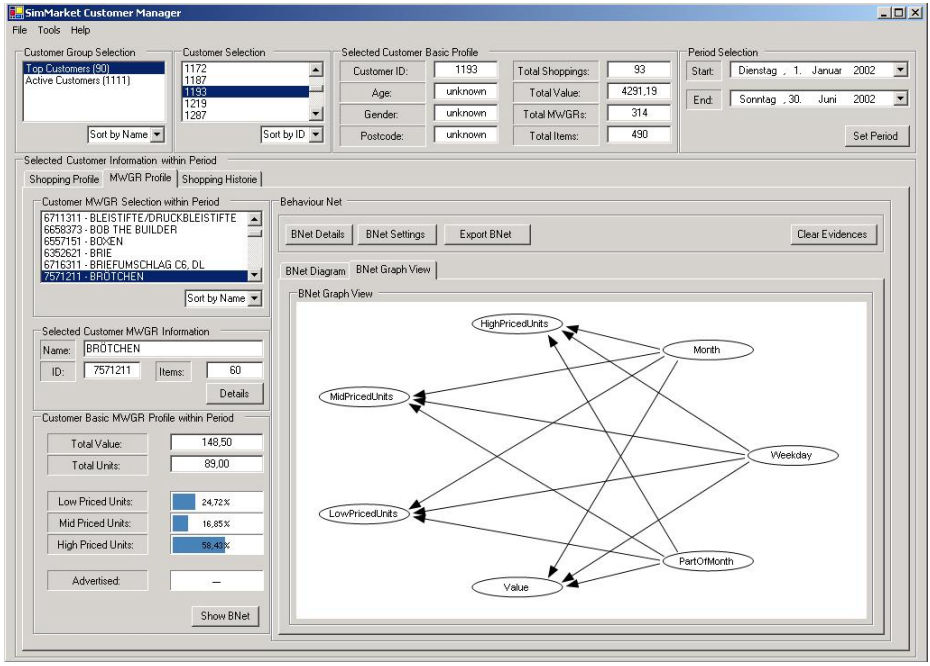


Fig. 5. The SimMarket System with a simple Bayesian Network

4 Agent-Based Simulation of Customer Behaviour

The decision about assortment, prices and promotions is based on assumptions concerning *future external influence factors* as well as the *current situation* and the result is a set of *internal changes* and *promotions* likely to lead to the company's business objectives.

Finding optimal sets of changes and promotions is the main challenge of the SimMarket project. The problem is that the company's primary objectives can vary between e.g. profit maximisation, cost reduction and increasing the number of customers.

Fig. 6 illustrates the simulation process of SimMarket at the level of a supermarket, i.e. regarding all possible decisions of a specific store.

First, a possible future scenario for the store is defined in accordance with the current situation on the basis of predicted external influences and planned internal changes. Based on this scenario a prediction about the possible impact of planned changes is generated by agent-based simulation. From the result the most important key figures are computed to indicate the degree to which the defined business objectives will be accomplished. To allow comparison and evaluation of the different scenarios more than one simulation run is computed. Finally the combination of changes and promotions is chosen, which has the highest probability to fulfil the defined business objectives.

In the actual simulation, the customer agents and item agents are presented with the current scenario. Their response indicates the individual preferences with respect to all affected products. When "questioned" the agents return the amount of units for each individual product, which the customer is expected to purchase in the scenario under consideration.

This is currently based on the specific behaviour of the commodity group networks with all affected networks being configured with the specified values of the influence factors for each defined action. The propagation algorithm of the behaviour network computes the expected values for all purchased products of the commodity group. Items are not represented as single units in the network but as a set of abstract attributes. Therefore, the "answer" of an agent is not a list of products which its customer is expected to buy but instead an abstract description of them, e.g. "five high-priced brand products" of a specific commodity group.

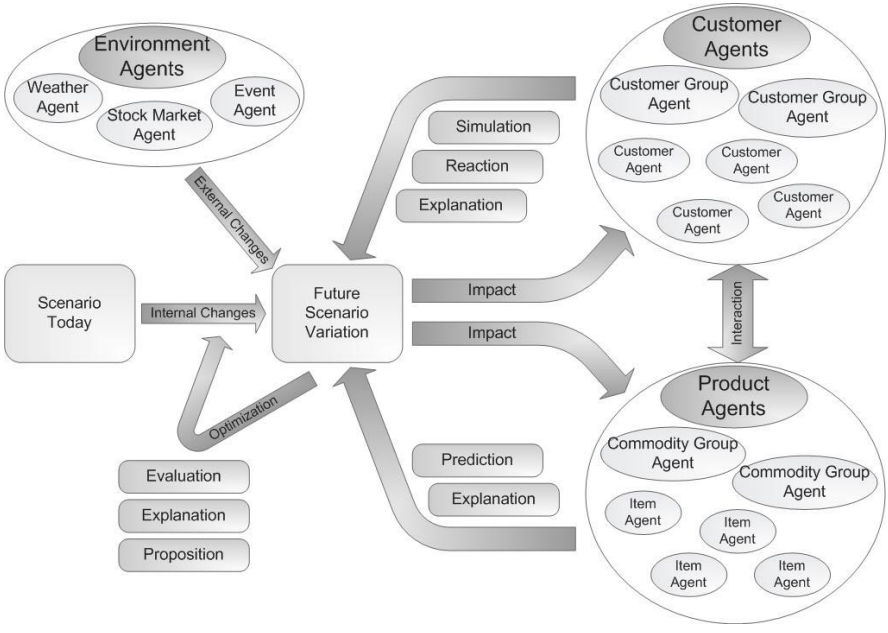


Fig. 6. The Multiagent-Based Simulation of Customer Behaviour

In later versions the simulation will comprise several steps. Initially, all relevant item agents give a description of their items in the form of an attribute list containing

e.g. price category, placing, quality and promotions. These attributes are then matched with those of the feature agents and the feature agents are confronted with these item descriptions. The relevant feature agents react internally with quantified acceptance or rejection and the feature agents then "negotiate" the overall reaction of the customer agent with respect to the relevant commodity group. Again the result is a list of item descriptions which the customer is expected to buy. The list of attributes can be mapped to specific items. Based on this the expected amount of units sold per product is computed. A quantified prognosis about the behaviour of all customers is obtained by summing up the expectations.

Using this approach it is also possible to integrate *virtual customers* into the simulation, representing typical customer groups instead of individual customers e.g. young families, students or senior citizen. The quantitative distribution of these prototypes of customer groups is determined by an analysis of sales slip data or questionnaires and is then used for the simulation.

The use of behaviour networks serves also an explanatory purpose as it allows a new kind of evaluation: besides computing scenario rankings and key figures it is also possible to give detailed verbal and graphical explanations for the simulation results.

For example a direct connection between increasing the price of a product and loss of customers to a discounter may to be due to the extreme price sensitivity of a certain group and this is what the system reveals.

A selective simulation of customer groups is possible since the customer groups are modelled as holons and the system explicitly forecasts the effects on specific customers and/or customer groups. Thus, a simulation is highly scaleable as the system can optionally simulate all, only some or just the individual customers.

5 Conclusion and Outlook

SimMarket is a MAS-based approach which models a real supermarket with all items, customers and relevant external influences, where the holon-paradigm supports arbitrary dynamic agent groupings.

A model based on a combination of a top-down (predefined general agent structure) as well as a bottom-up (individual behaviour and agent groupings based on real data) approach is presented. The extraction of individual shopping behaviour of single customers from the sales records of supermarkets and its representation as behavioural networks in the knowledge bases is the core of SimMarket. It is the basis for simulation of customer behaviour (regarding assortment, price and promotion changes) and supports the assortment and pricing manager in his decision making process.

Future research will improve the model of the customer agents by enhancing the society of feature agents. By defining adequate measurements of similarity, customers are divided into groups and represented as holons. The simulation results will be empirically evaluated in comparison to realistic sales data of the past to predict the effects on the future.

References

1. Russell, S., and Norvig, P.: *Artificial Intelligence: A Modern Approach*. Second Edition. Prentice Hall, 2003.
2. Jager, W.: *Modelling Consumer Behaviour*. Rijksuniversiteit Groningen, Dissertation June 2000.
3. Buchta, C., and Mazanec, J.: SIMSEG/ACM – A Simulation Environment for Artificial Consumer Markets. Working Paper Nr. 79, May 2001.
4. Fischer, K., Schillo, M., and Siekmann, J. (In print): Holonic Multiagent Systems: The Foundation for the Organization of Multiagent Systems. Proceedings of the First International Conference on Applications of Holonic and Multiagent Systems (HoloMAS'03).
5. Minsky, M.: The Society of Mind. Simon and Schuster (Touchstone), 1986.
6. Sinus Sociovision, Paris, France, www.sociovision.com.
7. Europanel, London, UK, www.europanel.com.
8. A. Jameson. *Systems That Adapt to Their Users.* Tutorial presented at IJCAI 2001, www.dfki.de/~jameson.
9. R. Schaefer. Benutzermodellierung mit dynamischen Bayes'schen Netzen als Grundlage adaptiver Dialogsysteme. Dissertation, University of Saarland, 1998.
10. Judea Pearl. Probabilistic Reasoning in Intelligent Systems. Morgan Kaufmann Publishers INC. 1988.
11. R. E. Neapolitan. *Probabilistic Reasoning in Expert Systems: Theory and Algorithms*. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York. 1998.
12. W. Kroeber-Riel, P. Weinberg. *Konsumentenverhalten*. 7. Aufl., (Vahlen) München 1999.
13. Hean-Lee Poh, Jingtao Yao, Teo Jasic. Neural Networks for the Analysis and Forecasting of Advertising and Promotion Impact. Intelligent Systems in Accounting, Finance and Management, Vol. 7, No. 4, 1998.
14. Frank M. Thiesing. *Analyse und Prognose von Zeitreihen mit Neuronalen Netzen*. Dissertation, Mai 1998.