An Approach to Estimate the Value of User Sessions Using Multiple Viewpoints and Goals

E. Menasalvas[1⋆], S. Millán[2], M.S. Pérez[1], E. Hochsztain[3], and A. Tasistro[4]

[1] Facultad de Informática UPM. Madrid Spain
[2] Universidad del Valle. Cali. Colombia
[3] Facultad de Ingeniería - Universidad ORT Uruguay
[4] Universidad de la República. Uruguay

Abstract. Web-based commerce systems fail to achieve many of the features that enable small businesses to develop a friendly human relationship with customers. Although many enterprises have worried about user identification to solve the problem, the solution goes far beyond trying to find out what navigator's behavior looks like. Many approaches have recently been proposed to enrich the data in web logs with semantics related to the business so that web mining algorithms can later be applied to discover patterns and trends. In this paper we present an innovative method of log enrichment as several goals and viewpoints of the organization owning the site are taken into account. By later applying discriminant analysis to the information enriched this way, it is possible to identify the relevant factors that contribute most to the success of a session for each viewpoint under consideration. The method also helps to estimate ongoing session value in terms of how the company's objectives and expectations are being achieved.

1 Introduction

The Internet has become a new communication channel, cheaper and with greater location independency. This, together with the possibility of reaching a potential market of millions of clients and reducing the cost of doing business, accounts for the amazing number of organizations that during the last decade have started operations on the Internet, designing and implementing web sites to interact with their customers. Though the Internet seems to be very attractive for both users and owners of web sites, web-based activities interrupt direct contact with clients and, therefore, fail to achieve many of the features that enable small businesses to develop a warm human relationship with customers. The loss of this one-to-one relationship tends to make businesses less competitive because it is difficult to manage customers when no information about them is available. Although many enterprises have been very worried about getting hold of the identity of the navigator, what is important is not the identity of the user but information about his likes and dislikes, his preferences, or the way he behaves. All this, integrated with information related to the business, will result in successful e-CRM.
The relationship with the users is paramount when trying to develop activities competitively in any web environment. Thus, adapting the web site according to user preferences

⋆ Research is partially supported by Universidad Politécnica de Madrid (project Web-RT)

B. Berendt et al. (Eds.): EWMF 2003, LNAI 3209, pp. 164–180, 2004.

is the unavoidable commitment that web-site sponsors must face and, when doing so, preferences and goals of the organization cannot be neglected.
Obtaining and examining the implicit but available knowledge about customers and site owners is the route to be followed to obtain advantages over other competitive Web sites and other communication channels. To analyze the available user navigation data so that knowledge can be obtained, web mining techniques have to be used. This is a challenging activity because the data have not been collected for knowledge discovery processes.
On the other hand, goal achievement has to be measured by the return on investment (ROI)[21]. Measuring the effectiveness of the site from this perspective is also challenging as it should comprise the different company's viewpoints [9]. Each viewpoint will correspond to a particular department or division of the company that will have, at each moment, defined the set of goals to achieve. Each goal will have a particular weight according to the global objetive of the company at a particular moment.
Some approaches (see section 2) have been proposed to measure the effectiveness of web sites but they only consider one viewpoint at a time. The challenge is to have a measure of the success of the site from each viewpoint considered, while also having a combined measure of the global success of the site at a particular moment.
The activity is challenging because the discrepancies among the different criteria used to evaluate business on the Internet often interfere with decision making and with the establishment of proactive actions. A project that has been successfully evaluated by one department is often classified as a failure by another. The first difficulty arises when trying to translate discovered knowledge into concrete actions [22] and trying later to estimate the effect of each action as ROI [12]. More difficulties arise on the Web since actions have to be taken repeatedly on-line as competitors are only a click away.
In this paper, we present an approach for measuring on-line navigation so that proactive actions can be undertaken. The approach is based on the evaluation of user behavior from different viewpoints and further action derivation. A method based on discriminant analysis is proposed so that in addition to obtaining an estimation of the value of the session, those factors that contribute most to the success of a session are identified. Since the exploitation approach is made in a dynamic environment and requires several coordinated and dependent actions, we propose to deploy the whole approach by means of a three-tier agent-based architecture in which agents for preprocessing, classification and proactive actions are identified. Agent technology allows applications to adapt successfully to complex, dynamic and heterogeneous environments, making the gradual addition of functionalities also possible [30].

The remainder of the paper is organized as follows. Section 2 presents the related work. Section 3 introduces the factors that have to be taken into account to estimate the value of a session when dealing with multiple viewpoints. The proposed methodology and a complete description of all the steps of the process presented in section 4. One of the key questions of the approach is the predictive method to estimate the value of the session. In this case, the method is based on discriminant analysis and is further explained in section 5. The architecture we proposed to deploy in our approach is shown in section 6. Section 7 presents experimental results and section 8 presents the main conclusions and further developments.

2 Related Work

Most enterprises are investing great amounts of money in order to establish mechanisms to discover Internet's user behavior. Many tools, algorithms and systems have been developed to provide Web site administrators with information and knowledge useful to understand user behavior and improve Web site results. Our work is related to three main topics: Web usage mining, Web agents and measures to evaluate site success.

2.1 Web Usage Mining

A way to evaluate the quality of a particular site is to determine how well a Web site's structure adjusts to the intuition of Web site users, represented in their navigation behaviour [3].

First, approaches to do this concentrated on analyzing clickstream data in order to obtain user's navigation patterns using data mining techniques [4] [16] [24] [26] [29]. However, Web mining results can be improved when they are enhanced with Web semantic information [2]. In this sense, several approaches, most of them based on ontologies, have been proposed in order to take site semantics into account. Berendt et al. [3] propose enriching Web log data with semantic information that could be obtained from textual content included in the site or using conceptual hierarchies based on services. On the other hand, Chi et al. [5] introduce an approach that, taking into account information associated to goals inferred from particular patterns and information associated to linked pages, makes it possible to understand the relationship between user needs and user actions. Oberle et al. [18] represent user actions based on an ontology's taxonomy. URL's are mapped to applications events depending on whether they represent actions (i.e. buy) or content.

However, most of these approaches and methods have concentrated on understanding user behaviour without taking Web business goals into account, whether these goals are diverse and dependent for different organizational departments. In [15] we propose an algorithm that takes into account both the information of the server logs and the business goals, improving traditional Web analysis. In this algorithm, however, the value of the links is statically assigned.
In this paper, we consider multiple business viewpoints and factors in order to understand Web user behaviour and act accordingly in a proactive way.

2.2 Web Agents

In this paper we introduce an architecture based on software agents. Taking into account that software agents exhibit a degree of autonomous behavior and attempt to act intelligently on behalf of the user for whom they are working [14][19], Web agents included in the architecture deal with all tasks of the proposed method.
In Web domain, software agents have been used for several purposes: filtering, retrieval, recommending, categorizing and collaborating.

A market architecture that supports multi-agent contracting implemented, with the use of the system MAGNET (Multi AGent NEgotiation Testbed), is presented in [6]. According to the authors, the system provides support for different kinds of transactions, from simple buying and selling of goods and services to complex multi-agent negotiation of contracts with temporal and precedence constraints. MARI (Multi-Attribute Resource Intermediary) [28] consists of an intermediary architecture that allows both buyer and seller to exercise control over a commercial transaction. MARI makes it possible to specify different preferences for the transaction partner. Furthermore, MARI proposes an integrative negotiation protocol between buying agents and selling agents.

Based on knowledge represented by multiple ontologies, in [20], agents and services are defined in order to support navigation in a conference-schedule domain. ARCH (Adaptive Retrieval based on Concept Hierarchies) [23] helps the user in expressing an effective search query, using domain concept hierarchies.

Most of these systems have been designed as user-side agents to assist users in carrying out different kinds of tasks. The agent-based architecture proposed in this paper has been designed taking the business point of view into account. Agents, in our approach, could be considered as business-side agents.

2.3 Measuring the Effectiveness of the Site

In spite of the huge volume of data stored on the Web, the relationship between user navigation data and site effectiveness, in terms of site goals when trying to design "good pages" or users, is still difficult to understand. Several approaches, models and measures have been proposed in order to evaluate and improve Web sites. Decision-making criteria related to the design and content of Web sites are needed so that user behavior is mapped onto the objectives and expectations of Web site owners.

Spiliopoulou et al. introduce in [25] a methodology useful for improving the success of a site and several success measures are proposed. The success of a site is evaluated based on the business goals. According to the authors it is necessary to identify one goal of the site at a time, in order to determine how different pages (action and target) on the site contribute to reach this goal. Besides, service-based concept hierarchies are introduced in order to transform Web site pages into action and target pages. User sessions are considered active sessions if they contain activities towards reaching the goal.
A set of metrics useful to evaluate the effectiveness of a Web market is proposed in Lee et al.[13]. The metrics definition, called micro-conversion rates, takes into account several shopping steps in online stores. Metrics have been integrated into an interactive visualization system and they provide information about store effectiveness.
Based on micro-conversion rates [13] and life-cycle metrics, Teltzrow and Berendt [27] propose and formalize several Web usage metrics to evaluate multi-channel businesses. Metrics are implemented into an interactive system that offers the user a visualization approach to analyze Web merchandizing.

3 Web Site Success Factors

Most of the approaches to evaluate the success of a Web site are stated in terms of efficiency and quality. In these approaches efficiency is generally measured by means of the number of pages accessed and served, duration of a session, and the action performed by the users (e.g. buy, download, query). Quality is measured by the response time, accessability of pages, and number of visitors of the site. The success of a company is measured by means of indicators such as profitability, costs, cost-effectiveness, ROI, gross sales, volume of business and turnover.
Thus, executive directors are often both amazed and disappointed by the differences between the criteria used to determine if the investment on a Web site is successful and the criteria used to evaluate the success of a project outside the Web. Department managers recognized that little consideration is given to the financial and commercial aspects of e-business projects [7] [8].
Due to the fact that the traditional criteria based on the ROI cannot be avoided when evaluating investments in Internet projects [21], some approaches that consider success both from the technological perspective (content, design of the Web pages) and the commercial perspective (achievement of company's goals) [9] have arisen.

Nevertheless, considering the commercial aspects of the site is not the only requirement for measuring the success of the site. The global success of a company is the result of the contribution of each department to the fulfilment of the company goals. It is not reasonable to think that a company has only one success criterion (independently of the fact that the Web is being used as channel). On the contrary, and particularly in the Web sphere, by its presence on the Web always tries to achieve more than one goal, depending on particular environment conditions. Thus, both the goals and their weight for different viewpoints have to be taken into account when measuring the success of the site.

3.1 Different Viewpoints, Different Goals

As it has been already mentioned (see section 2), previous approaches assume that the analysis concerns one objective at a time characterizing this objective as the goal of the site. However, the Web site is not the aim but the means to achieve a company's goals. The significance of goal achievement differs for each company and each department of a company. In retailing domains, it is often the case that the only viewpoint under consideration is merchandizing, so that success is measured as the number of purchases. But for the marketing department the number of times a product is accessed is the goal rather than the number of final purchases. Thus, success or failure is not a one viewpoint function but it depends on different viewpoints each of which defines its own success criteria and goals. The goals of a company then depend on the viewpoint of the different departments, sections or divisions of the same company, defining in each case its own weights for the set of goals. It is often the case that different departments will assign weights to different goals in a contradictory way. In any case, the importance of each goal will depend on the environment conditions.
All that has been stated above, can be equally applied to businesses that use the Web only as a communication channel. In this sense, for example, the greatest significance

for the marketing department will be assigned to the attractiveness and ease of use of the Web site by the user, while the department in charge of the design of the pages will give more importance to the site's design. In contrast, the sales department's most significant goal will be increasing the number of purchases.
Under any circumstances, the global success criteria will depend on the particular conditions. Hence, the goal established as: fulfill the user's needs at the same time as the company's goals are fulfilled can be measured as a function that depends, at least, on the following factors:

- Goals to be achieved
- Viewpoints that are considered
- Environment conditions
- User information on navigation and satisfaction
- Pages content

We formally express these factors in the following definitions.

Definition 1 *The set* $Goals = \{g_1, g_2, \ldots, g_s\}$ *represents the goals of a company at a particular moment.*

Definition 2 *The set* $Viewpoints = \{v_1, v_2, \ldots, v_n\}$ *represents different points of view (e.g., marketing, sales).*

Definition 3 *We define* w_{ij} *as the function that assigns weights to each goal, where* i *represents the* i^{th} *viewpoint and* j *the* j^{th} *goal. We assume that weights are going to be assigned to different goals depending on the viewpoint and depending on the environment conditions. As we are dealing with a weight function, it has the following properties:*

- $0 <= w_{ij} = w(g_i, v_j) <= l$ *where* $g_i \in Goals$ *and* $v_j \in Viewpoints$
- $\sum_{i \in Goals} w_{ij} = 1$

The set of goals G is made of the goals that the company (every department or just one particular department or division), aims to achieve. The importance of the different goals for the different viewpoints is reflected in the weight the goal has been assigned for each viewpoint. Thus in a way it is possible to say that the weight assigned to a goal is conditioned by the viewpoint under consideration. For each viewpoint the sum of the weights assigned to all the goals is always 1 (it may happen that a particular goal has zero as the weight assigned by a viewpoint meaning that this is not a goal of this viewpoint at this moment), $\sum_{i \in Goals} w_{ij}/v_j = 1$.
The example in table 1 illustrates this particular fact.

In the example of Table 1, goal g_2 is only taken into account by viewpoint number 3. Thus, for the rest of the viewpoints its weight will always be zero.
These assignments will be made at each particular moment by the experts of each department corresponding to different viewpoints. Notice that these assignments may change, for example, because of inside decisions (the company promoting a particular product or section), by actions taken by the competition (a competitor launching a new product), or simply by an environmental event (political events, ...). In order to be proactive the assignments need to always be kept update. Notice that these assignments could also be made on the basis of different user profiles.

Table 1. Example of relationship between goals and viewpoints

Goal	$v_1 = Marketing$	v_2 = Commercial	v_3=Web Design
g_1 = Long sessions	0.3	0.1	0.3
g_2= Awards in site design	0	0	0.6
g_3= Increasing purchases	0.1	0.4	0
g_4= Purchase different products	0.2	0.2	0
g_5= Access different products	0.3	0	0.1
g_6= Session Profit > 0	0.1	0.3	0

3.2 Mapping Pages into Business Goals

Business experts are responsible for the establishment of viewpoints, goals and weights. Once these have been established, the challenge is to add this information to the information about users and their navigation patterns. The only information about user behavior that is available is stored in the Web log (clikstream). The gap to be filled is then the mapping of the business information to the clickstream. In order to capture the business information and to integrate it with the user navigation, we propose pages to be enriched with semantic information related both to the content (as already proposed in [2]) and to the business goals and viewpoints.
In our case, for each site and for each particular viewpoint we define the set of semantics (actions and/or contents) that are relevant to be studied by the site. Notice that this set of semantics can be modified through the site's life time depending on the factors that are relevant to be analyzed at each moment. The relevant information is modelled with the use of ontologies, taxonomies and databases. The way in which this is implemented depends on the technology underlying the site. For example, a site could enrich XML pages themselves with this information. In the case of pages being dynamically generated the enrichment could be done when accessing the database to build the page. The construction of the ontologies and databases containing semantics is outside the scope of this paper as we are concerned with the way we use this information to obtain a predictive model that can compute the value of a session. Thus the important point is that information about the business is obtained and later mapped against the pages.
Possible information to enrich sessions is the length of the session, whether the user has accessed pages in which awards were given, whether he has clicked on a link in the right or left handside part of the page, and whether he has downloaded certain information.

4 Estimating the Result of a Session

We propose to estimate the result of a session while the user is navigating. Our proposal is to map information about navigation onto business information to have a measure of how the goals are being achieved and consequently act. To estimate the result of ongoing sessions we propose a method based on discriminant analysis. The method is twofold: on the one hand, relevant factors to establish the success of a session are obtained and on the

other, a measure of the success or failure of the session is obtained. This approach will help classify users on-line taking into account different viewpoints and, consequently, act according to their behavior and the weights each goal has been assigned. The process we propose to undertake includes the following steps:

- Classify historical user sessions as success or failure for the different viewpoints.
- Enrich the information of these sessions with information about the business.
- Apply discriminant analysis to obtain both relevant factors for the success of a session and a method to estimate the result of ongoing sessions.
- Apply the on-line method obtained to estimate the result of the session.
- Act consequently to the estimated result.

4.1 Classify Historical Sessions

The first requirement for obtaining a predictive model is to have historical examples already classified. In our case, the innovative aspect is that sessions have to be classified according to the different viewpoints under consideration.
In classifying sessions we propose two approaches that have been used in the case-study:

- Expert-driven classification: In this approach the expert establishes some criteria for the success of a session, e.g. having asked for information about a certain product. Later, all historical sessions are classified according to the established criteria. In our approach, criteria for the different departments have to be established.
- Improved-expert driven classification. In the second approach the expert is given a set of sessions and instead of giving the criteria to classify them, he classifies them. Then using a classification method (decision tree in our case) the rest of the historical sessions are classified. The main problem of this approach is that the expert is rarely available to manually classify past sessions and if available he will luckily classify a very small set of sessions. Thus, the reliability of the classification method will be under question.

For the future, we propose to use an agent that records, once the session is over, its value (real one) from all the viewpoints considered. These sessions are the input for future improvement of the estimation procedure. This task is performed in the refining stage, as shown in section 6.

4.2 Enrich the Information on the Sessions

The set of sessions already classified enriched with information about the business. Consequently, we obtain a table called **enriched session table**, in which each tuple contains the session identifier and the business information that is taken into account,i.e., information related to action, contents, design of the pages, etc. For each piece of information taken into account, the session can only take values 0 or 1 to specify whether that action occurred during the session. As the approach is multi-goal, in this table we keep, for each department (viewpoint), if the session fulfilled each goal. Nevertheless, we need a global value of the session for each department. In order to obtain this global value, weights of the goals are used. And the resulting values are stored in a new table, named

Results table. Consequently, *Results table* contains information related to the success or failure of each session from each point of view considered.
The innovative aspect of the approach is that several viewpoints and goals are considered at a time. This way, for each session and for each goal we will have a measure of how successful the session occurred to be. Once this information is available, we propose to apply discriminant analysis to obtain the factors that contribute most to the results for each viewpoint and for each goal.

5 Discriminant Analysis Application

We applied a stepwise multivariate predictive model to the set of sessions already classified and enriched. The basic strategy in discriminant analysis is to define a linear combination of the dependent attributes. In our case, the dependent attributes are semantic actions or concepts $s_1, s_2, \ldots, s_t$ that have been used when enriching the sessions. Notice that the aim of our approach is to determine which user actions (s_i) contribute to achieve the department goals. Thus the equation will have the form:

$$L = v_1 s_1 + v_2 s_2 + \ldots + v_t s_t.$$

Once this equation is obtained, the success or failure of a session will be established on the basis of the value of L obtained for that session:

- if $L >= 0$ session success is predicted
- if $L < 0$ session failure is predicted

For simplicity reasons we use here a dichotomic classification. Nevertheless, the extension to deal with more classes is straightforward and can be found in the references. Discriminant analysis is useful in finding the most relevant semantic concepts. In each step of the stepwise discriminant technique the importance of each attribute included can be studied. This is important not only because we will have the equation to estimate the value of a session in the future but also because the method provides the analyst with criteria to understand which actions are most relevant for the success of a session from each viewpoint. The computation of the discriminant function has been done according to [11], [1], [17]. The model estimates the coefficients (values) of each attribute considered (semantic) for each pair of goals and viewpoint considered.
>From this result, it is possible to establish the relationship between semantics and the result of the session. Those coefficients with higher positive value, are associated with sessions successfully ending while those taking negative values are associated with failure sessions.
Hence the proposed approach, provides both the predictive model and the procedure used to establish actions to make a session successful from a certain site viewpoint. Notice that the number of relevant concepts ($N1$) will be much less than the number of pages N ($N1 << N$) in a Web site, so that the problem of analyzing user session decreases in complexity, improving the performance of the methods used to analyze them.

5.1 Applying the Algorithm Online

Once the functions to estimate session value have been obtained, they can be applied on-line to decide the action to be undertaken depending on the estimated result of the session. This activity is again challenging as the result depends on the moment that we apply the model. Due to the fact that it is difficult to predict when a session is to be terminated, let alone finding out how many pages the user will visit, the timing for applying estimation procedures can bias the results.
In our case, we propose to use the algorithm proposed in [10]. According to this algorithm, "breaking" pages in which the user will make a decision can be recognized in the navigation pattern of a user. In the architecture proposed here there will also be agents in charge of this task and according to the breaking pages, the agents will decide the moment of application of the estimation procedure for each on-going session. As a result of the application of the estimation procedure, we will obtain a value of goal achievement for the session and for each viewpoint considered. Furthermore, making use of the weight assignment policies at each moment, another agent will decide on the global value of the session and the action to be taken, if any (i.e., action agents). The action to be taken has to be decided by the business experts and we assume that the information on these actions is stored somewhere in the system and accessed by action agents. Notice that these actions could depend on the user profile if this information were available in the system.

6 Architecture Overview

Web Mining tasks have often been implemented by agents. The agent paradigm offers desirable features such as autonomy, that is, the ability of acting itself and on behalf of others and proactivity, that is, the ability of acting in anticipation of future problems, needs or changes. These characteristics are very suitable in the scenario described in the previous sections because of its dynamic idiosyncrasy. Furthermore, the agent paradigm makes dynamic changes of functionalities feasible. Since Web Mining tasks evolves quickly, we consider that this paradigm is the most suitable one.
Thus, a multiagent architecture is proposed, which is composed of three different layers:

1. Semantic Layer. This layer contains agents related to the logic of the algorithm or method used in the ongoing session result estimation.
2. Optimization/Decision Layer. Corresponds with the agents responsible for optimizing or making decisions depending on the estimated value.
3. Service Provider Layer. This layer contains agents that provide several services other of the agents. These services are generic and independent of the other layers. It also offers an interface, which will be used by any agent asking for a service.

6.1 Semantic Layer

This level is fed directly by the session value estimation methods. The agents of this layer deal with the concepts value estimation and its usage in subsequent sessions. This layer is a multiagent subsystem composed of different specialized agents. They are:

- Preprocessing agents: These agents are responsible for enriching the sessions.
- Classification agents. Classification of a session is made according to an expert, but the criteria used can be automated through the usage of previous classifications and expert knowledge.
- Estimation agents. These agents apply the stepwise multivariate predictive model. In a first phase, this operation is made offline. Nevertheless, the algorithm must be applied online in current Web usage data to estimate the result of a session. Another task that has to be performed by these agents is obtaining the breaking points to decide when the estimation procedure needs to be applied.
- Refining agents. Classification is a continuous task. It is necessary to refine the algorithm with new information (new sessions, new business criteria). This kind of agents must communicate with the estimation agents in order to inform them about changes.

As we can see, using agents in the semantic layer provides adaptivity to the algorithm.

6.2 Optimization/Decision-Making Layer

This layer includes agents that make decisions depending on the information supplied by the semantic layer. We define a generic action agent as an agent template for building agents which allow actions to be taken according to session values. Although it is possible to build other kinds of agents using this template and the needs of the Web site, we have defined the following agents with the aim of optimizing the accesses and personalization usage of the Web site:

- Prefetching agents. These agents prefetch most probably next visited Web pages, depending on the session value. Sessions with a higher estimated value are given higher priority. This way, the Web session load is more efficient and the user feels more comfortable in the Web site.
- Adaptive agents. These agents are responsible for building offers adapted to the preferences of the users. These offers may be shown as popups or Web pages. Any other kind of personalization can be added to the logic of these agents.

6.3 Service Provider Layer

This level includes generic services used for assisting other layer agents. These agents are:

- Data retrieval agents. The goal of these agents is to retrieve data from different information sources. These sources are heterogeneous (e.g., databases, files) and they have different types of information and access requirements. Therefore, these agents can delegate on specialized agents for different sources.
- Locator agents. They are in charge of connecting agents. In order to locate a given agent, the locator agents use its category, that is, the type of agent. If there is no available agent of this type, the system launches a new agent for serving this request.

It is possible to add more services to the architecture. The procedure to do so is to implement the corresponding kind of agent and define the corresponding interface with the rest of the architecture.
Figure 1 represents the three layers of the proposed architecture and the relationship between layers. Notice, that there are both internal and external relationships among different kinds of agents. The different information sources are also shown in the graphic.

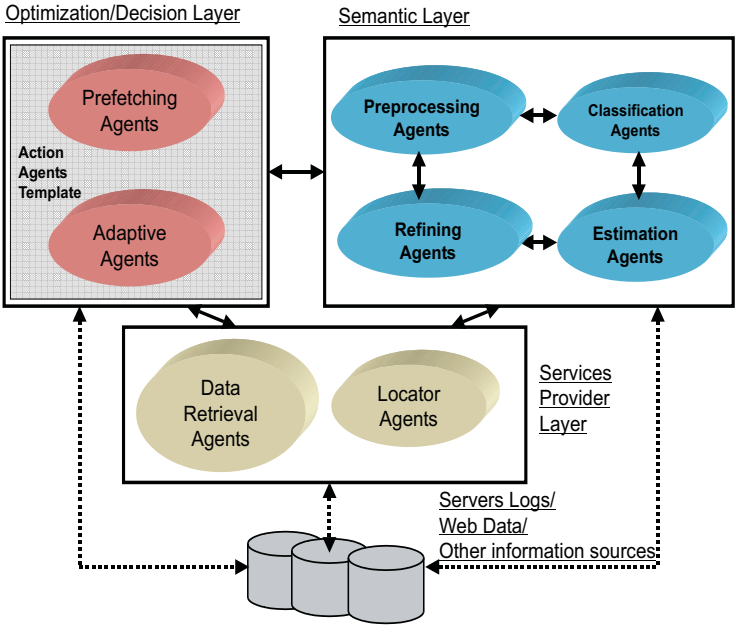


Fig. 1. Web-behavior agent-based architecture layers

7 Case Study – Experimental Results

A company developing its activity both through the Internet and in the traditional way, was taken as the example for the case study. The company under consideration was having difficulties to define success criteria. The departments involved in the study were:

- Marketing
- Commercial
- Web Design
- Technological
- Manufacturing

The e-commerce site contained 2500 pages. We present here the results after analyzing sessions on the server for a month. After filtering out irrelevant entries, the data were segmented into 38058 sessions. The discriminant model has been obtained through a two-step process of test and training (80%)
Table 1 shows the weights of the goals for each viewpoint under consideration. For clarity reason, we only present in this section the results for the commercial department. The main goal of the analysis was applying the proposed method so that the following actions could be undertaken:

- Establishment of concepts or actions behind the visited pages (semantics) that are more relevant for the success of a session (i.e., those that contribute most to the achievement of goals).
- To provide on-line information to automatically make decisions on what motivates the user (e.g., automatic prefetching, online offers and discounts).

The results of the study found discrepancies between departments and highlighted that the cause of problem was the fact that they were trying to measure the efficiency of the site from a single viewpoint. Viewpoints were identified and success criteria were defined to be integrated in the procedure that estimates the result of a navigation session. The method used to evaluate the sessions combines qualitative and quantitative elements. Qualitative methods included meetings with the managers of the five divisions involved. The aim of interviews was to find the main goals of each department and the importance that each department assigns to each goal.
Some the goals identified include increasing the number of purchases (both in number of purchases and in variety of products in the baskets), awards for the design of site pages, use of contact points at the site by the users.
The quantitative elements were used to measure the achievement of the proposed goals. The first step was to eliminate from the set of goals those that were not directly reflected in the sessions. For example, the goal: “obtaining awards for the design of Web pages” was eliminated.
In a second step, semantic elements were identified to enrich sessions. As most of the pages in the site were dynamically generated, the semantic enrichment in this case was done as the page was generated.
A total of 94 elements were identified to enrich the pages. During the preprocessing and data transformation phases, only 9 out of these 94 elements, were considered relevant. These 9 elements are shown in table 2.

The value of each concept has been established. Results are shown in table 3. The model predicts 93.3 percent of the successful sessions and 82,2 percent of the failure sessions.
The biggest advantage of the proposed method is that helps to determine the importance of each concept for a viewpoint with a high predictive power. A concept is consider to have high predictive value when the model properly classifies more than 70 percent of the sessions.
For the commercial department, the predictive accuracy of the model can be seen in table 3. In this table, we can see that 93.3% of the successful sessions and the 82.2% of the failure sessions were correctly predicted.

Table 2. Examples of semantic elements used to enrich pages

semantic-id	semantic-description
s1	advertising
s2	news
s3	food purchase
s4	cleaning products
s5	cosmetics
s6	download travel information
s7	travel booking
s8	download music information
s9	asking information about promotions

Table 3. Summary discriminant table

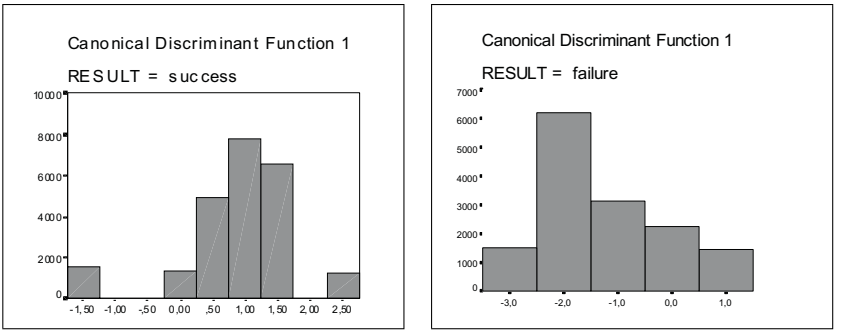
	Predicted	
Observed	Success	Failure
Success	93.3	6.7
Failure	17.8	82.2

The discriminant function obtained was:

$L = 0,431s_1+, 0,99s_2 + 0,126s_3 + 0,751s_4 - 0,444s_5 - 0,107s_6 + 0,119s_7 + 0,742s_8 + 0,306s_9$

The proposed analysis helped to establish the relevant factors to take into account when analyzing sessions.
The results of the discriminant function have been depicted in an histogram, where X-axis represents the value of the discriminant function (L) and the Y-axis represents the function of the empiric density. Thus, the area of each rectangle represents the relative frequency of values of the discriminant function for each interval. (Figure 2(a) represents the results for successful sessions and figure 2(b) represents the histogram for failure sessions). The in-depth analysis of the wrong classified sessions helped the site sponsor recognize market niches. Notice that it is more difficult to describe the wrongly classified examples as they resides at the edges.

In 83% of the cases, the number of relevant concepts able to classify a session was 20% or less out of the 94 initial concepts. Finally, when goals were weighted according to each point of view taken into account, an aggregate perception of all goals was obtained for each point of view (i.e., the same session was successful for the Department of Marketing but not for the Sales Department).



(a) RESULT = success. (b) RESULT = failure.

Fig. 2. Canonical Discriminant Function 1

8 Conclusions

The global success of a company is the result of the contribution of each department to the fulfilment of the company goal. Based on this fact, in this paper we have presented a method to estimate the value of a navigation session according to different viewpoints. This is the innovative aspect of the approach as it makes it possible to analyze user behavior by a global measurement related to all the relevant goals of the company. Besides, an agent-based architecture has been defined with the aim of providing dynamism to method deployment.

The main drawback of the method proposed in this paper, is that it requires a set of sessions already classified by the experts of each department. To solve this problem we have also proposed a semi-automatic method to classify sessions but the methods still depends heavily on the expert.

The method has shown to have promising results in the e-commerce site that has been used in the case-study. The results highlighted some market niches and helped the organization to find the factors that made sessions end successfully.

One important factor to measure the success of a session that has been tackled by the proposed approach, is that of dealing with the different goals and viewpoints of the company. However, goals to be achieved by a company also depend on the customer typology, this is to say, for example, that goals to be achieved for very loyal customers may differ from those for the customer who visit the site for the first time.

On the other hand, the proposed method estimates the value of a session based on the presence of some relevant factors the order in which the events occur.

Some open issues that can be developed and addressed by multiple alternatives have been the motivation of current research for improving the proposed method and forecoming work.

Acknowledgments. The research has been partially supported by Universidad Politécnica de Madrid under Project WEB-RT and Programa de Desarrollo Tecnológico (Uruguay).

References

1. R. E. Anderson, R. L. Tatham, and W. C. Black. *Multivariate Data Analysis (5th Edition)*. Prentice Hall College Div, Mar 1998.
2. B. Berendt, A. Hotho, and G. Stumme. Towards semantic web mining. *ISWC 2002, LNCS 2342*, pages 264–278, 2002.
3. B. Berendt and M. Spiliopoulou. Analysis of navigation behavior in web sites integrating multiple information systems. *The VLDB Journal*, 9:56–75, 2000.
4. J. Borges and M. Levene. A heuristic to capture longer user web navigation patterns. In *Proc. of the First International Conference on Electronic Commerce and Web Technologies, Greenwich, U.K.*, pages 27–34, September 2000.
5. E. H. Chi, P. Pirolli, and J. Pitkow. The scent of a site: A system for analyzing and predicting information scent, usage, and usability of a web site. In *Proc. of CHI '2000 The Hague, Amsterdam CHI Letters*, 2000.
6. J. Collins, W. Ketter, M. Gini, and B. Mobasher. A multi-agent negotiation testbed for contracting tasks with temporal and precedence constraints. *Int. Journal of Electronic Commerce*, 7(1):35–57, 2002.
7. Jaap Gordijn and Hans Akkermans. Does e-business modeling really help? *Proceedings of the 36th Hawaii International Conference On System Sciences,(HICSS-36 2003)*, January 2003.
8. Jaap Gordijn and J.M. Akkermans. Value-based requirements engineering: exploring innovative e-commerce ideas. *Requirements Engineering*, 8(2):114 – 134, July 2003.
9. Jaap Gordijn. Why visualization of e-business models matters. *16th eCommerce Conference eTransformation panel Business models & the mobile industry: Concepts, Metrics, Visualization and Cases*, June 2003.
10. M. Hadjimichael, O. Marbán, E. Menasalvas, S. Millan, and J.M. Peña. Subsessions: a granular approach to click path analysis. In *Proceedings of IEEE Int. Conf. On Fuzzy Systems 2002 (WCCI2002), Honolulu, U.S.A.*, pages 878–883, May 2002.
11. D. G. Keibaun, L. L. Kupper, and K. E. Muller. *Applied Regression Analysis and Other Multivariable Methods*. PWS-KENT Publishing Company, USA, 1988.
12. Ron Kohavi, Neal J. Rothleder, and Evangelos Simoudis. Emerging trends in business analytics. *Communications of the ACM*, 45(8):45–48, August 2002.
13. J. Lee, M. Podlaseck, E. Schonberg, and R. Hoch. Visualization and analysis of clickstream data of online stores for understanding web merchandizing. *Data Mining and Knowledge discovery*, 5(1/2), 2001.
14. P. Maes. Agents that reduce work and information overload. *Communications of the ACM*, 37(7), 1994.
15. E. Menasalvas, S. Millán, and E. Hochsztain. A granular approach for analyzing the degree of afability of a web site. In *Proc. of International Conference on Rough Sets and Current Trends in Computing RSCTC2002*, October 2002.
16. B. Mobasher, H. Dai, T. Luo, M. Nakagawa, and J. Witshire. Discovery of aggregate usage profiles for web personalization. In *Proceedings of the WebKDD Workshop*, 2000.
17. D. Morrison. *Multivariate Statistical Methods*. Duxbury Press; 4th edition, 2002.
18. D. Oberle, B. Berendt, A. Hotho, and J. Gonzalez. Conceptual user tracking. *Advance in Web Intelligence. LNAI 2663. AWIC 2003*, pages 155–164, May 2003.

19. M. Papazoglou. Agent-oriented technology in support of e-business. *Communications of the ACM*, 44(4), 2001.
20. T. R. Payne, R. Singh, and K. Sycara. Browsing schedules - an agent-based approach to navigating the semantic web. *ISWC 2002, LNCS 2342*, pages 469–473, 2002.
21. Gregory Piatetsky-Shapiro. Interview with Jesus Mena. http://www.kdnuggets.com/news/2001/n13/13i.html, 2001. Last update: July 2003.
22. M Schoop, A Becks, C Quix, T Burwick, C Engels, and M Jarke. Enhancing decision and negotiation support in enterprise networks through semantic web technologies. *R. Tolksdorf, R. Eckstein(eds.): XML Technologien für das Semantic Web - XSW Berlin Lecture Notesin Informatics*, Volume P-14:161–167, 2002.
23. A. Sieg, B. Mobasher, S. Lytinen, and R. Burke. Concept based query enhancement in the arch search agent. In *Proceedings of the 4th International Conference on Internet Computing (IC'03), Las Vegas*, June 2003.
24. M. Spiliopoulou, C. Pohle, and L. Faulstich. Improving the effectiveness of a web site with web usage mining. In *Proceedings of the WEBKDD99*, 1999.
25. M. Spiliopoulou and C. Pohle. Data mining for measuring and improving the success of web sites. *Data mining and Knowledge Discovery*, 5(1-2), 2001.
26. J. Srivastava, R. Cooley, M. Deshpande, and P. Tan. Web usage mining: Discovery and applications of usage patterns from web data. *SIGKDD Explorations*, 1:12–23, 2000.
27. M. Teltzrow and B. Berendt. Web-usage-based success metrics for multichannel businesses. In *Proceedings of WebKDD 2003 Workshop August 27h, 2003, Washington DC, USA*, 2003.
28. G. Tewari and P. Maes. Design and implementation of an agent-based intermediary infrastructure for electronic markets. In *Proceedings of the EC'00, Minneapolis, Minnesota*, October 2000.
29. G. Wolfang and S. Lars. Mining web navigation path fragments. In *Workshop on Web Mining for E-Commerce - Challenges and Opportunities Working Notes (KDD2000), Boston, MA*, pages 105–110, August 2000.
30. Michael Wooldridge and Nicholas R. Jennings. Intelligent agents: Theory and practice. *Knowledge Engineering Review*, 1995.