TRIGGERING RULES FOR CONVERSATIONAL AGENTS IN TRADING SITUATIONS

Grzegorz Dziczkowski, Arnaud Doniec and Stéphane Lecoeuche
Univ. Lille Nord de France, Ecole des Mines de Douai, IA, Douai, France

Keywords: Conversational agents, Data mining, Web intelligence, Behavior analysis.

Abstract: This paper describes a methodology to establish behavior rules for conversational agents on commercial web sites. Our work is a contribution to a recent research field: *agent mining* (Cao, 2009) which results from two interrelated research area: Agent/Multi-agent system and Data Mining. The proposed methodology is based on behavior analysis of e-commerce clients and customers' segmentation. Our proposal has been applied on a real commercial web site to construct the triggering rules of a virtual seller agent.

1 INTRODUCTION AND ISSUE

Electronic commerce (e-commerce) has become very popular as the World Wide Web has grown, with many websites offering on-line sales and e-commerce activity undergoing a significant revolution.
However, on any commercial web sites, the conversion rate (the ratio of visitors who convert casual content views to a commercial transaction) is very low. Many reasons can be put forward to explain this observation. One of them is the lack of humanity in the customer relationship.

The work described in this paper is a part of a project which aims to study the implementation of innovative, intelligent and automatic solutions to enhance on-line sales.
Project challenge is to build an intelligent agent which is able to understand, to reason and to decide in an autonomous way. The proposed agent has to be able to interact with the visitors performing proposals of relevant marketing messages at the right time and giving useful advises. This virtual seller has to be a conversational agent with the ability to:

- perceive and analyse the behavior of a new visitor,
- recognize an existing client,
- interact and communicate with clients through text and using a human representation (avatar),
- understand the need of clients and decide upon which action to take in collaboration with existing information systems (CRM tools for instance).

This paper focuses on the triggering of the agent. In other words, we try to answer the questions: *when and why should the virtual seller appear to a customer?* Our aim is to construct, for the agent, a set of behavior rules which control the appearance and the beginning of the discussion with the clients. We base our works on an analysis of customers' behaviors: we propose to identify great categories of clients and to associate some triggering rules to them.

2 LITERATURE REVIEW

By implementing new techniques to analyze more precisely user' actions, it becomes possible to extract strong correlations between web site pages and typical behavior during a time period. The main difficulty in the field of Web Usage Mining is getting numerical vectors describing the navigation of users.

The first step to perform analysis of customer behavior is pattern discovery. A variety of machine learning methods have been used for pattern discovery. The approaches that most often appear in literature are: clustering, classification, association discovery, and sequential pattern discovery. We focus our research on clustering methods. For this reason clustering is the only approach described in more details in this paper.

Generally clustering aims to divide a data set into groups that are very different from each other and whose members are very similar to each other. This method has been used for grouping users with common browsing behavior (Srivastava et al., 2000). As the customer may belong to more than one group the clusters should be able to overlap. (Yan et al., 1997) use the Leader algorithm to cluster user ses-

Dziczkowski G., Doniec A. and Lecoeuche S..
TRIGGERING RULES FOR CONVERSATIONAL AGENTS IN TRADING SITUATIONS.
DOI: 10.5220/0003183603820387
In *Proceedings of the 3rd International Conference on Agents and Artificial Intelligence* (ICAART-2011), pages 382-387
ISBN: 978-989-8425-40-9

sions. Each user session is represented by an n-dimensional features vector, where n is the number of Web pages visited during the last 30 minutes in the session. The computation of weight is based on a different parameters (like the number of times the page has been accessed, the amount of time the user spent on the page). A partitioning clustering method is employed by (Cadez et al., 2000), which visualizes user navigation paths in each cluster. In this system, users' sessions are represented using categories of general topics for Web pages. A number of predefined categories are used as a bias, and URLs are assigned to them, constructing the user sessions. The Expectation Maximization (EM) algorithm, based on mixtures of Markov chains is used for clustering user sessions.

An extension of partitioning clustering methods is fuzzy clustering that allows the presence of ambiguities in the data, by 'distributing' each object from the data set over the various clusters. Such a fuzzy clustering method is proposed in (Joshi and Joshi, 2000) for grouping user sessions, where each session includes URLs that represent a certain traversal path. The Web site topology is used as a bias in computing the similarity between sessions. The site is modelled using a tree, where each node corresponds to an URL in the site, while each edge represents a hierarchical relation between URLs. The computation of the similarity between sessions is based on the relative position in the site tree of the URLs included in the sessions.

Model-based clustering methods have been also used in (Paliouras et al., 2000). A probabilistic method, a neural network Self-Organizing Maps, and a conceptual clustering method, are exploited in order to construct user community models (i.e. models for groups of users with similar usage patterns). Community models are derived as characterizations of the clusters and correspond to the interests of users' communities.

Despite of the variety of clustering methods that have been used for Web usage mining, no work has been done on the comparison of their performance. The reason for this is the inherent difficulty in comparing clustering results, due to the lack of objective criteria independent of the specific application (Pierrakos et al., 2003).

3 SELLER AGENT TRIGGER

In order to determine the onset triggering factors of seller agent for each customer, we are interested in at least 2 different aspects, i.e.:

- how can the agent characterize clients' behaviors, and
- how will the agent's appearance be performed.

The first aspect listed above is in the field of Web Usage Mining. Research has focused on methods, based on Web Mining and Machine Learning algorithms, to automatically analyze different data (from questionnaires, observation of sales areas, interviews with clients and vendors, loyalty programs or internet data such as web server logs) in order to obtain the most relevant information.

To answer the second aspect, we will study the seller agent trigger based on customer navigation analysis.

It means that our system has to categorize a user who is surfing on the commercial website. The seller agent should evaluate user expectations and motivations from his apparent behavior and habits. The source of data is limited to the data we can collect (e.g. visited pages, request, connection time, the use of site tools, etc.). The main problem is to determine the kind of client's behaviors we could regroup based upon navigation information and how user's classes can be evaluated. The proposed methods also take into account the visit context like the year period (vacation, sales period, etc.), the day of week or the hour.

By the expression *creation of customers' behavior* we mean the analysis of customers' navigation traces on the commercial website in order to detect the most significant and common profiles - clustering. The creation of the triggering rules set is based directly on the customers' behavior types detected.

To perform acceptable triggering rules of the agent we should first identify the web users, then create and analyze the general customers' behavior and at last define the set of rules of virtual agent triggering.

We concentrate our research on two sorts of rules of virtual seller agent trigger: the *specific* and the *general* rules. The *general* rules are divided into *direct* rules and *transition* rules.

As an example of *specific* rules, we can consider the cases where a user enters on support pages of ecommerce website, or when the client try to buy several products of the same family with incoherent parameters (like for example a size of a duvet cover may be different from size of a flat sheet). In first case we can lunch the virtual seller in order to perform assistance, in the second case the agent can suggest the possibility of confusion of products features. These rules could satisfy a customer but as they are too *specific* and detailed, the number of agent triggers based on these rules will be the most of time low. Even if the *specific* rules will have high precision it seems impossible to predict all the situations they can be adapted, it means that the recall will be low.

For this reason we have to develop *general* rules which should be more global. As we already described above we will establish and analyze customers' behaviors. We predicted the set of behaviors' *types*. As an example of *general* rules we can consider the situation when customer changes his behavior from one *type* to another *type* during his navigation. We assume that these changes may indicate that the client gets lost, needs some help, or is confused, we can also detect that a customer checks for more details or compares the prices/products/products' features ... (conclusion - that the customer could be interested to talk with seller, even with the virtual seller). Nevertheless, the change of customer' behavior can show that a client lost or earned his interest in products proposed on the website what also is an important criterion to seller triggering. On the opposite side, when the client follows a common *type* of navigation (behavior) we can assume that the client knows exactly what he wants and he pursues the objective. The general rules implementation requires the introduction of client behavior clustering. If a rule refers to one cluster we have a *direct* rule, and if a rule refers to the changes of clusters during navigation we have a *transition* rule.

As we can infer from this section, the most important part in our research is the customers' behavior analysis.

4 PROPOSED METHODOLOGY

The proposed methodology to perform the set of virtual seller triggering rules consist in 3 steps:

- feature selection (from raw logs files)
- performing the clustering (and assign commercials labels to clusters)
- establish the set of triggering rules (specific and generals rules)

The main idea of analysis the customers' behavior and transition of clusters during user navigation was the limitation the customers' sessions to 10, 15 and 20 actions. Each action corresponds to one page view by the users. The session limitation means that for example one user session with 40 actions was changed on 3 sessions with 10, 15 and 20 actions. Then the clustering was done separately for each limitation (10, 15, and 20). The choice of actions numbers in new limited session was due to the two reasons. First it was not possible to perform a pertinent clustering using less than 10 actions for user session because the cluster was not enough significant and differences between clusters were negligible. Second reason was more commercial - the agent should not wait to much time before appear in front of customer and only 10% of sessions are longer than 20 actions.

After repeating this treatment for each session, we performed the clustering using different features for sessions' limitation. The pertinent clusters (from statistic point of view) was then analyzed with marketing experts of our commercials partners in the goal to establish the best clustering from marketing point of view. During this state the commercial labels was assigned to the clusters in the goal of prediction the set of triggering rules.

At the end the specific and general rules was created. The details of each step are described below. During our research we were in possession of our commercial partner's data for the one entire year period (120 Gb). For our learning base we use the sample of data of one month. We choose the month of April due to the miss of any marketing actions. The database for one month represents more than 300 000 of sessions with more then 10 actions performed. On account of the scale of the database the treatment is time consuming.

The logs files delivered from our partners were in the form of NedStat logs files. The main difference between this format and Extended logs files is that to each client the unique Id is assigned (based on cookies).

Before selecting the navigation features, the hierarchy of web site was performed. We divided the site on 7 different universes: store (the main universes with for example products list), quick order (direct purchases by entering catalogue reference), shopping cart (purchase), sales, consulting (customers questions, FAQ), condition (terms of sale, shipping), various (all others like for example home page). The universe store was divided on three levels of hierarchy: section, subsection and sub subsection. Generally the final product page corresponds to sub subsection.
Example:

```
the final page of product:  Tablecloth XYZ
--> universe:  Store; section:  Table;
subsection:  Tablecloth; sub subsection:  Tablecloth XYZ
```

Based on this hierarchy, we selected 36 session features witch describe customer navigation of our commercial partners web site Table 1.

The set of triggering rules depends directly from the navigation feature selection. For this reason, it seems necessary to describe them more precisely: *"User ID"* describes the id based on cookies, unique id for a user; *"Session ID"* designs the session id during one day, each session is considerate as closed after 30 minutes without any action; *"Purchase"* is Boolean value which shows if the customer already made a purchase during his action; *"Reduction*

Table 1: All the session features possible to retrieve.

User ID	Session ID
Day/Month/Year	Hour of begin
Hour of end	Purchase
Total amount	Nb products bought
Nb references bought	Reduction code
Knew customer	Source of navigation
Total time	Time universe (1-7)
Nb total pages seen	Nb pages universe (1-7) seen
Nb universes changes	Nb sections changes
Nb subsection changes	Nb subsubsection changes
Nb of section seen	Nb of subsection seen
Nb product pages seen	Nb of same product seen

code" - Boolean value describes the presence of reduction code during the purchase; *"Knew customer"* describes whether the user has been recognized as a client who has already made a purchase on the site; *"Source of navigation"* describes whether the user is entered into our commercial partner's site voluntarily by using for example the search engine, or was pushed to visit the site by the mail company; *"Total time"* gives the length of a session; *"Total universe (1-7)"* represent 7 different features which describe the time that a visitor spend on each universes; *"Nb total pages seen"* describes the number of all the pages visited by the user during a session; *"Nb pages universe (1-7) seen"* represent 7 different features which describe the number of pages visited by a user on each universe; *"Nb universes/section/subsection/sub subsection changes"* - 4 features which describe the number of changes the user make during his navigation, if for example the user switch the universe and then came back to previous one the value of this feature is equal to 2; *"Nb of section/subsection seen"*- 2 features which describe the number of different section or subsection seen during the user session; *"Nb product pages seen"* describes the number of product pages seen in total; *"Nb of same product seen"* describes the sum of product pages that have been seen several times.

4.1 Customer Behavior Analysis

Once the above treatment was performed for users sessions of our learning base (month of April - 1000000 sessions), the customer behavior analysis can be implemented. At the beginning, the work corresponds to analyse the features obtained (data mining techniques) in order to detect the most discriminated features. For customer behaviors detection we decided to perform the clustering. Our approach consist of the study of users comportment in two cases: first comportment can be assign to the cluster the user belongs to (direct rules), second comportment is analyzed in accordance with the users transition from a cluster *A* to a cluster *B* during his navigation (transition rules). The specific rules depends on the characteristic comportments of user navigation and not on the clusters detected, they are build in empiric way and will not be described in this paper.

In order to determine the changes in the user navigation, we perform an analysis on client's session which assigns the customer to one of the classes which were previously detected (cluster). In order to choose the best clustering from commercial point of view, our statistical results were discussed with our commercial partners' experts. In this way clusters for each limitation get their commercial labels. The general rules are based on the labels form and their transition (due to the change of user comportment during navigation).

As it's difficult to calculate the performance of clustering model the choice of final clustering was done by marketing experts based on statistic descriptions of clusters. The data was pre-treated before and normalized. The Ward clustering method was implemented, the number of cluster was calculated in accordance of cubic criterion cut-off parameters ($CCC < 3$), and the cluster criterion chosen was Least Squares.

The interesting results for customer behavior analysis on this stage are the features which are the most discriminated for each session limitation, and the number and parameters of clusters.

Below we will discuss our results of clustering which was chosen depending on statistic and marketing criterions.

`User Session Limitation to 10 Actions`

We obtain 6 clusters. The most discriminated characteristics are: [table 2]

Table 2: Discriminated features for 10-actions sessions.

Source of navigation	Nb pages universe store seen
Nb of subsection seen	Nb pages universe shopping cart seen
Nb subsection changes	Nb pages universe various seen
Purchase	Nb pages universe quick order seen
Hour of begin	Nb of section seen
Nb sub subsecion changes	

The example of description of cluster from statistic and marketing point of view is as follows:

Statistic Features. Cluster Id: 6; Frequency of cluster = 23357 (12.6%); Root-mean-square standard deviation = 0.13; Source of navigation: Mail = 0.88; Nb pages universe store seen = 8.62; Nb of subsection seen = 7.14; Nb pages universe shopping cart seen =

0.03; Nb subsection changes = 1.53; Nb pages universe various seen = 1.33; Purchase: non = 1; Nb pages universe quick order seen = 0.01; Hour of begin = 53511; Nb of section seen = 2.16; Nb sub subsection changes = 1.21

Marketing Label. The cluster refers to sessions in which the gateway is not directly accessible but with an e-mail that pushes the specific store section to a customer. The client is at least a lead which gave his opt-in (level of commitment higher than if entry from search engine). Of the first 10 pages no product has been moved to shopping cart. At this point, a dialog with a virtual seller could be desirable.

`User Session Limitation to 15 Actions`
We obtain 8 clusters. The most discriminated characteristics are [table 3]:

Table 3: Discriminated features for 15-actions sessions

Source of navigation	Nb pages universe store seen
Nb sub subsecion changes	Nb pages universe various seen
Nb of section seen	Nb product pages seen
Nb of subsection seen	Nb universes changes
Purchase	

The example of description of cluster from statistic and marketing point of view is as follows:

Statistic Features. Cluster Id: 8; Frequency of cluster = 1327 (1.2%); Root-mean-square standard deviation = 0.14; Source of navigation: Mail = 0.08; Nb pages universe store seen = 6.03; Nb sub subsection changes = 1.17; Nb pages universe various seen = 6.09; Nb of section seen = 1.60; Nb product pages seen = 2.61; Nb of subsection seen = 2.11; Nb universes changes = 1.1; Purchase: non = 0

Marketing Label. The cluster refers to sessions in which customer already purchase products and validated his command. The other features are not significant and it's the only cluster where client already purchase his product. Despite the presence of purchase the client still navigates on web site or just finishes the validation step. If the customer will stay an a website for few actions there is possibility of seller agent trigger.

`User Session Limitation to 20 Actions`
We obtain 8 clusters. The most discriminated characteristics are: [table 4]

Table 4: Discriminated features for 20-actions sessions.

Nb of subsection seen	Nb pages universe store seen
Source of navigation	Nb subsubsecion changes
Nb pages universe various seen	Nb universes changes
Nb product pages seen	Nb of section seen
Purchase	Nb pages universe shopping cart seen

The example of description of cluster from statistic and marketing point of view is as follows:

Statistic Features. Cluster Id: 2; Frequency of cluster = 25335 (27.6%); Root-mean-square standard deviation = 0.11; Nb of subsection seen = 8.55; Nb pages universe store seen = 18.5; Source of navigation: Mail = 0; Nb sub subsection changes = 0.7; Nb pages universe various seen = 1.36; Nb universes changes = 1.32; Nb product pages seen = 3.79; Nb of section seen = 1.46; Purchase: non = 1; Nb pages universe shopping cart seen = 0.11

Marketing Label. The cluster refers to sessions in which the customer doesn't have a precise idea of product he want to purchase. Client often changes the subsections of web site hierarchy and do not visit the product pages of the same family. The navigation is based on store part of web site but like there are only few products' pages visited and the section is usually changed it seems that the customer checks different family of product and gets the general idea of internet shop offer. The client did not indicate interest of the specific products, or family of products.

4.2 The Set of Triggering Rules

During the step of clustering, we find that in different session's limitations (10, 15, and 20) we have the corresponding cluster (if a customer doesn't change his behavior during navigation). We assign the direct and transition rules with the help of commercial experts which has a strong knowing how in the field of sales. The rules were assigned by analyzing all selling scenarios which are interesting from marketing view. The general form of the set of triggering rules is as follows:

$$(C_{p_{10}} \in c_i \wedge C_{p_{15}} \in c_j \wedge C_{p_{20}} \in c_k) \longrightarrow Ta\{Id, K, R\}$$

where, $C_{p_{10}}$, $C_{p_{15}}$, $C_{p_{20}}$ - customer profile at 10, 15 and 20 actions; c_i, c_j, c_k - existing cluster for each limitation (i=6, j= 8, k=8); Ta - triggering action; Id - Id_User; K - context of triggering rules, R - the set of user navigation features.

The virtual agent triggering can be seen as a condition-action rule: *if* a criterion *then* a triggering action (which answer questions when and why). The antecedent describes the user navigation on session limitation and the consequent describes the presence of triggering rules. The triggering rule parameters are: *Id user* for identify the client for which the virtual seller will appear immediately, and parameters which are transmitted to our project partners to perform the correct dialogue between client and virtual seller. These parameters are represented by the context of triggering (see below) and the set of users' navigation features which contains the value of features

from Table 1 and the description of all section, subsection and sub subsection visited.
We present an example of a direct and a transition rule.

`Case 1.`

Context. A rule concerns the cluster 8 of session limitation to 15 actions. It refers to a customer who already validated an order before the onset of virtual seller. The case was described in section 4.1
Trigger. The customer has already ordered but still browsing on the site. If a customer will not left the site in a few seconds, the virtual agent should appear to push new products based on recommender system and propose to back to shopping cart to add some goods.

`Case 2.`

Context. A rule concerns the transition of cluster 3 to 7 to 2. During two first session limitation (10 and 15) the customer exhibits the behavior of a client who knows his taste, who checks for precise products of particular families, who spends time on the similar products' pages. The last session limitation is different and describes a general navigation on the store without main idea of purchase (section 4.1). The client arrived spontaneously on the site and has a priori a clear idea of the desired product (many of the same family product pages seen). After about 6 minutes, he stops to check product descriptions and seems to disperse in the section of the highest level
Trigger. There is a risk that the customer will leave the site. The virtual agent should appear immediately to struck up a discussion about the consulted products.

5 CONCLUSIONS

The presented work was done for one month without significant sales and marketing campaigns. The first analysis of clusters performed for sale period show strong differences. The future work includes the analysis of number of year period for which the clustering will be implemented in final system.
This paper deals with the design of behaviors for a virtual seller agent. Such agent should mimic the comportment of human vendor in real shop. More precisely, our work focuses on the way to define rules for the triggering of the discussion between the agent and the client. The general steps presented in this paper are: navigation features selections, clustering compatible with marketing assumptions, and design of triggering rules.

We manage the selection of 36 customer navigation features. The adaptation for a new e-commerce site is feasible in a relatively short time because they are based on navigation patterns and not on URL's. We also can assume that the format of e-commerce shops remains in the similar form. We succeed in establish the final clustering model for each user session limitation of our approach. The choice of final clustering model depended on the quality of clusters' commercial labels assigned. At the end the set of triggering rules was performed based most of all on clusters labels. We also implemented the supervised classification on our data and we obtained good and promising results which can evaluate the step of navigation feature selection for clustering part. We didn't find any related work on automatic triggering of agent for e-commerce site. The main contribution of this paper is the methodology used to design the set of triggering rules which mainly depends of customer behavior and his navigation on the web site.

REFERENCES

Cadez, I., Heckerman, D., Meek, C., Smyth, P., and White, S. (2000). Visualization of navigation patterns on a web site using model based clustering. Technical report, Technical Report MSR-TR-00-18. Microsoft Research.

Cao, L., editor (2009). *Data Mining and Multi-agent Integration*. Springer.

Joshi, A. and Joshi, K. (2000). On mining web access logs. In *ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*.

Paliouras, G., Papatheodorou, C., Karkaletsis, V., and Spyropoulos, C. D. (2000). Clustering the users of large web sites into communities. In *Proceedings of International Conference on Machine Learning (ICML), Stanford, California.*

Pierrakos, D., Paliouras, G., Papatheodorou, C., and Spyropoulos, C. D. (2003). Web usage mining as a tool for personalization: A survey. *User Model. User-Adapt. Interact.*, 13:311–372.

Srivastava, J., Cooley, R., Deshpande, M., and Tan, P.-N. (2000). Web usage mining: Discovery and applications of usage patterns from web data. *SIGKDD Explorations*, 1:12–23.

Yan, T. W., Jacobsen, M., Garcia-Molina, H., and Dayal, U. (1997). From user access patterns to dynamic hypertext linking. Technical report.