ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

ΣΤΗ ΔΙΟΙΚΗΣΗ ΕΠΙΧΕΙΡΗΣΕΩΝ

Διπλωματική Εργασία

STOCHASTIC PROCESSES APPLICATIONS IN E-COMMERCE

του

ΣΥΝΑΔΑΚΗ ΑΓΓΕΛΟΥ

Υποβλήθηκε ως απαιτούμενο για την απόκτηση του μεταπτυχιακού

διπλώματος ειδίκευσης στη Διοίκηση Επιχειρήσεων

Ιούνιος 2013

Acknowledgments

For this thesis I would like to sincerely thank the supervisors of this study, Georgiou Andreas and Andronikidis Andreas, Professors at the University of Macedonia, Thessaloniki, Greece. Their contribution made it possible to produce this work.

Abstract

Nowadays, people tend to rely on Internet to do their shopping. Even though shopping online is constantly rising, its channels convert only a small fraction of their visitors into buyers. This might be either because of technical factors (for example, website design or functionality) or because of the initial user's intentions. By using stochastic processes a shop manager can acquire valuable data that can be used to rise the conversion percentage and the successfully completed sales. In this essay, we study the phenomenon of cart abandonment and we propose a hidden Markov model that detects user's characteristics, heterogeneity and intentions on the fly. Then this information is proposed to be used by the store in order to adapt properly and reduce the probabilities of cart abandonment and by the store's manager to gain a better understanding of the way consumers tend to use the website.

Keywords

e-shopping, e-commerce, stochastic processes, Markov chain, hidden Markov models, HMM, website

Contents

Acknowledgments ii

Abstract iii

Contents iv

Index of tables vi

Index of figures vi

1. Preface 1

2. Literature Review 8

3. A mathematical approach 12

    3.1. Markov Chains 12

    3.2. Hidden Markov Model 15

4. The proposed model 20

    4.1. The session stages 20

        4.1.1. Stage 1 – Information seeking 20

        4.1.2. Stage 2 – Consideration stage 21

        4.1.3. Stage 3 – Evaluation stage 21

4.1.4. Stage 4 – Purchase decision 22

4.2. The Markov model 23

4.2.1. Time independent model 23

4.2.2. Time dependent model 25

4.3. The hidden Markov model 26

4.4. Proposed model implementation 31

5. Managerial implications 34

Appendix 1 – Tables 37

Bibliography 39

Papers 39

Websites 42

Index of tables

Table 1: Eurostat – Online shopping data (March 2013)

Index of figures

Figure 1: 2012 online turnover (in billion euros) (Ecommerce Europe, November 2012)

Figure 2: Why consumer prefer online shopping (Invesp Consulting, 2011)

Figure 3: Why do online shoppers leave without paying? (WorldPay, 2012)

Figure 4: A simple Markov chain

Figure 5: A Hidden Markov Model by Rabiner (1989)

Figure 6: Markov model graph designed by Mathematica 9.0

1. Preface

E-commerce is a worldwide emerging way of shopping. According to Eurostat (March 2013), during 2012, 45% of individuals born in European Union (27 countries) have made an online purchase during the last 12 months (appendix 1, table 1). This percentage varies among the European countries from 27% (Cyprus, 2012) to 81% (Norway, 2012). In Greece, 30% of individuals have made an online purchase in 2010, which is the last year of available data. Given the technological advancement of Greece in the last years, it is safe to assume that this percentage has been risen during 2012.

The total turnover of products and services sold online in Europe was expected to reach over 305 billion euro in 2012, compared with 254 billion in 2011 (20% increase). This amount makes Europe the world's largest online market, while United States is the second biggest one with a turnover of 280 billion euro in 2012 and Asia-Pacific the third with a turnover of 216 billion euro (Ecommerce Europe, November 2012)(figure 1).

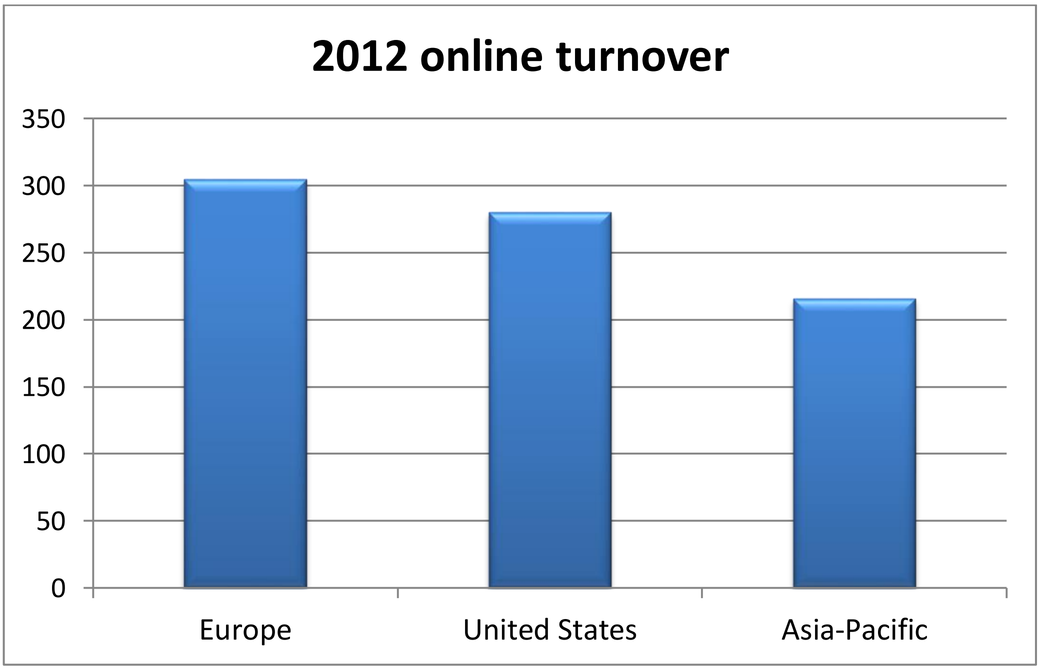

Figure 3 - 2012 online turnover (in billion euros) (Ecommerce Europe, November 2012)

According to Invesp Consulting (2011) online shopping is chosen by consumers at increasing rates for a variety of reasons. Among them, time saving is the most significant factor (73%), together with greater variety of products and services that can be found online (67%). Also, the simplicity of price comparison between similar or competitive products (59%), the lower taxes when store is abroad (e.g. on Asian countries) (30%) and the avoidance of crowds (58%) are quite important factors. Finally, as long as money is concerned, the reduced prices (55%) and the reduction of other costs (such as transportation, parking, etc.) (40%) constitute a considerable advantage (figure 2).

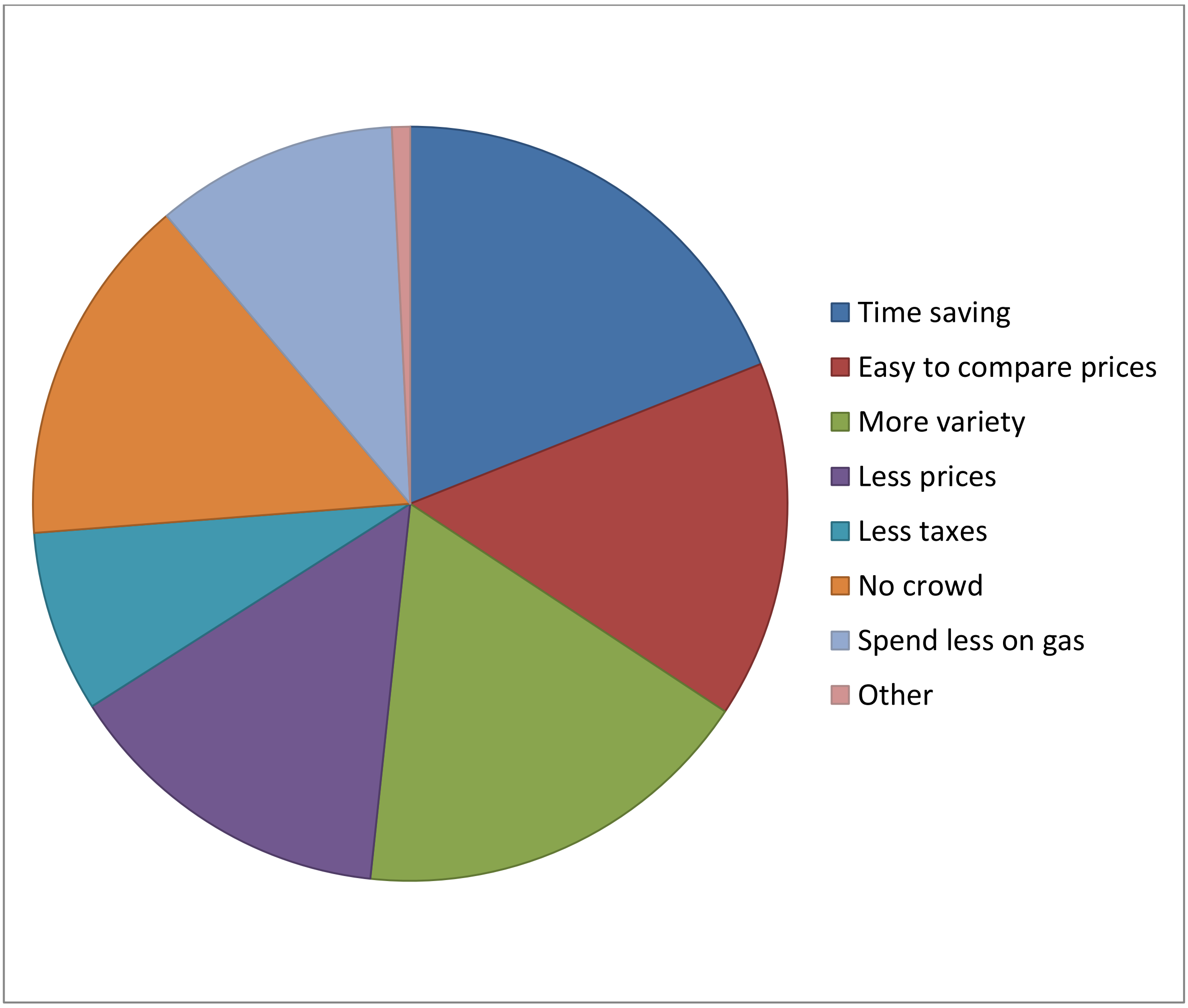

Figure 4 - Why consumer prefer online shopping (Invesp Consulting, 2011)

E-commerce is a vast way of shopping but its channels diachronically convert into buyers on average only 2 to 4 percent (Li and Chatterjee, 2005; Search Marketing Standard, 2007; Smart Insights, March 2013) of their visitors. This percentage may rise up to 8% when there is cart activity, meaning there have been products added to the virtual shopping cart (Smart Insights, March 2013). Consequently, only a small portion of consumers that visit an online store compete a purchase.

There are a lot of online store visitors that do not intent to buy something and they are just browsing through the website seeking for information. Nevertheless there is a significant amount of missed purchases than could have been converted, if the browsing experience or other factors (e.g. psychological ones) were different.

In this essay we are going to focus on methods of increasing conversion rates on online stores by using stochastic processes applications. Specifically, we are going to focus on "cart abandonment", the phenomenon where a consumer leaves an online store without buying anything, even though there were already items placed on the virtual shopping cart. We are going to approach this phenomenon by studying existing hidden Markov models (HMM) that are designed to predict and interact with visitor's intentions and other characteristics.

Moe and Fader (2004) stated that online shopping differs from offline because the latter costs significantly more in terms of time, transportation, etc. Therefore, conversion (visitors to buyers) at offline stores is higher since customers are more likely to buy. Also, information seeking and alternative evaluation has already occurred online, so customers are more aware of whether they want to buy and what, when they visit a brick and mortar store (Kukar - Kinney and Close, 2010). It is concluded that customers visit online and offline stores for a variety of reasons apart from actually buying a product or a service (e.g. information seeking, price comparison, etc.).

In order to focus on online carts and their key role to online buying experience, it is essential to describe the consumer's shopping process at an online store. Such a process consists of 4 stages and follows the information processing theory of consumer choice (Howard and Sheth, 1969):

1. the consumer views a shopping page – a page with product information and, maybe, price (information seeking stage)
2. the consumer adds the product to the shopping cart (consideration stage)
3. the consumer decides to start the checkout process by choosing to visit the shopping cart page (evaluation stage)
4. the consumer decides to buy the items (or some of them) contained in the shopping cart (purchase decision stage)

While the consumer browses through a lot of different products in the online store, the aforementioned stages may occur more than once, except from the decision to buy the items (fourth stage) or abandon the cart which may happen only once. Also, on the first step, the consumer may choose to visit other pages on the online store as well, such as pages with information about the store, the means of payment provided, the shipping costs, store's reviews and testimonials, etc.

Online shopping behavior differs from offline in a great variety of ways and so, shoppers *behave in fundamentally different ways* (Ganesh et al., 2010). Consequently, the usage of online and offline shopping carts should not be considered as the same. While in an offline store, customers tend to add items on their shopping carts when they are almost definite they want them and they are going to buy them. On the contrary, while on online stores, customers may use their shopping cart as a bookmark facility for organizational and research reasons (Kukar - Kinney and Close, 2010) or even for hedonic reasons (Kukar - Kinney and Close, 2010). During browsing through an online store, customers find interesting items or maybe similar items and they place them in their virtual shopping cart so they can review and compare them further when they are done browsing. This item selection does not necessarily mean an intention of buying since it can be only for comparison reason or future reference (Punj and Moore, 2009).

Also, there are different moments during the online shopping process where customers choose to abandon their carts. The majority of cart abandonment occurs before customer enters the checkout process, namely before hitting the “buy” or the “checkout” button or anchor. Another common moment of abandoning carts is at the point of sale, namely when the customer is asked to fill in shipping and payment information.

According to previous studies, approximately 7 out of 10 online virtual shopping carts are abandoned (Oliver and Shor, 2003; Goldwyn, 2002; Smart Insights, March 2013). Online stores may have the opportunity to reduce cart abandonment provided they are aware of why and when it happens. Generally, the reasons of online shopping cart abandonment can be distinguished in two broad categories (Li and Chatterjee, 2005):

- Website design and browsing experience factors (e.g. insufficient product information or shipping costs, limited website functionality, confusing buttons or design, over-detailed registration requirement, unstable interactivity with website and more)
- Consumer’s motivation factors (e.g. high prices, desire of more comparison before purchase, postpone purchase, lack of time, lack of means of payment, shipping issues and more)

According to WorldPay (January and February 2012), the most common reasons consumers are abandoning their online virtual shopping carts are the following (Figure 3):

- Presented with unexpected costs (56%)
- Just browsing (37%)
- Found better price elsewhere (36%)
- Overall price too expensive (32%)
- Decided against buying (26%)
- Website navigation too complicated (25%)

- Website crashed (24%)
- Process was taking too long (21%)
- Excessive payment security checks (18%)
- Concerns about payment security (17%)
- Delivery options were unsuitable (16%)
- Website time out (15%)
- Price presented in a foreign currency (13%)
- Payment was declined (11%)

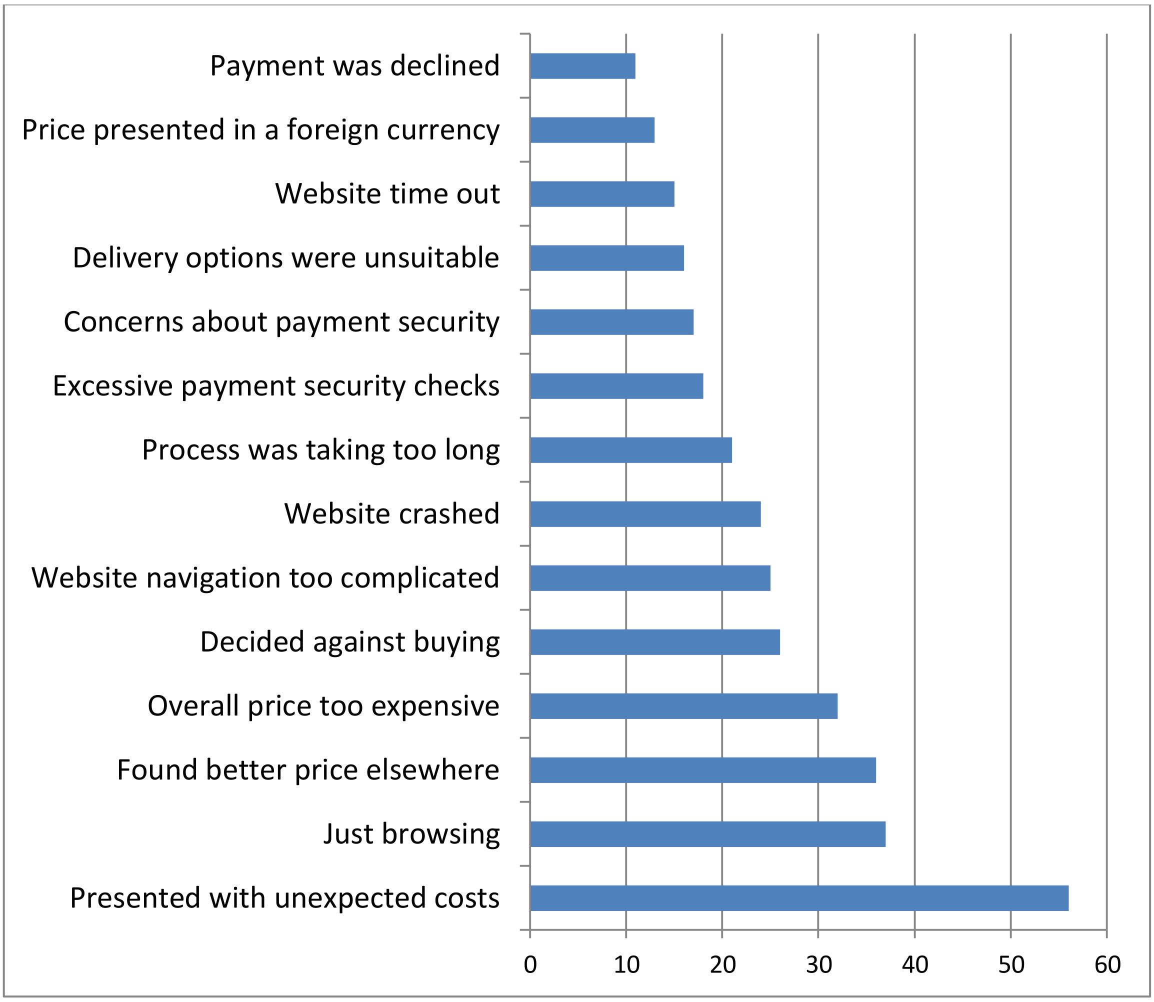

Figure 3 – Why do online shoppers leave without paying? (WorldPay, January and February 2012).

It is, therefore, essential for the shop manager to provide such an experience to its visitors that will lead to more purchases. This experience should take under consideration website design and browsing experience as well as customer's motivation and psychology.

In this essay, we are going to approach the cart abandonment phenomenon by analysing store's data. The aim is to examine the main reasons that customers leave the store prior to buying. Then, by using stochastic processes, we are going to analyze store's consumer heterogeneity. This information can then be used by store owners or managers to design and provide unique online experiences for different customer segments.

2. Literature Review

There are a lot of different definitions as far as "cart abandonment" is concerned. Cho (2004) stated that cart abandonment occurs when a consumer visits an online shop intending to make a purchase but does not complete it and abandons this intention. Moore and Mathews (2006) have a slightly different approach since they define cart abandonment when a consumer puts items in the shopping cart to gather information but decides to abandon the cart before the final purchase stage. A more recent definition by Ouellet (2010) concludes that cart abandonment occurs when a shopper begins the checkout process but does not complete it. The difference between these definitions lies on whether the check-out process has been initiated before the cart abandonment occurred or whether there was only cart activity (selection of products).

As far as the behavioral factors are concerned, Cho (2004) concluded that the intention of buying online depends on the past e-shopping experience of the shopper. Later, Cho et al. (2006) came to the conclusion that delay factors, such as perceived uncertainty, medium/channel innovation and contextual characteristics, combined with consumer characteristics, have a significant impact on online shopping hesitation and, thus, cart abandonment. According to Egeln and Julie (2012), the cart abandonment can be looked at in two different ways: behaviorally and technologically.

There are other behavioral factors that may increase or decrease the buying likelihood of an online shopper. Such factors may be promotional codes or coupons (Oliver and Shor, 2003), hyperlinks, promotions and other advertisements that consumers encounter online (Mandel and Johnson, 2002), etc. People tend to quit an online store when they are prompt to add a coupon code and they do not have one, or they encounter promotions or special offers to products similar to the ones they intended to purchase.

Regarding the technological or web-site factors, the checkout process complexity, the web site design, the graphics load time, the online purchase confirmation time (the time that takes for the user to receive a confirmation that his or her order has been received correctly), the registration process (or the need of a registration for an order to be placed), the login issues, and the credit card security threats are some of the most common factors that affect the buying intention and the cart abandonment (Rajamma et al., 2009).

Furthermore, it is essentially important that there is congruity and consistency between the e-commerce website and the firm and its actual stores (Wang et al., 2009; Close and Kukar – Kinney, 2010). Finally, the website's personality may have a key effect of website's perceived quality and, thus, to consumers' purchase intention (Poddar et al., 2009). People expect to have online a similar experience with the actual stores, namely they want the same branding (colors, mottos and so on), similar level of innovation and technology, same products, etc., in order to maintain the buying desirability.

There is little previous work regarding mathematics and stochastic processes in general, or hidden Markov models in particular. Most related researches analyze clickstream data to produce results. Clickstream is the recording of the parts of the screen a computer user clicks on, while web browsing or using another software application[1].

Wang et al. (2000) built a hidden Markov model in combination with Viterbi algorithm in order to adjust a website on the fly based on user's clickstream data. Montgomery et al. (2004) used clickstream data from a major online bookstore to model the browsing behavior state in relation with user's goals or state of mind. Sismeiro and Bucklin (2004) used a similar methodology on an automobile website in order to model the shopping process. Moe (2003) used clickstream data to reveal shoppers' underlying objectives and categorize them as buyers, browsers, searchers or knowledge-builders. Li and Chatterjee (2005) proposed a four-stage model of online

[1] http://en.wikipedia.org/wiki/Clickstream

shopping process by using hidden Markov process to capture unobserved navigation orientation and account for consumer heterogeneity.

Scott et al. (2006) used nested hidden Markov models to track Internet browsing behavior. In their research they used two hidden Markov chains in order to gather all the necessary data. One hidden Markov model was used to track the sequence of page requests within a session and another hidden Markov model to track consecutive web sessions by the same user. This model was implemented on a small e-commerce website to determine the roles of session-level and page-level chains. Through the collected data, Scott et al (2006) came to the conclusion that session-level model was able to discover the “session type” (for example, if it is a buying session or searching session, etc.), while the page-level model accounted for other data features that would not be possible to be gathered by the session type model.

Ching et al. (2004) applied a hidden Markov model to classify the customers of a company (CRM application). A similar research was conducted by Netzer et al. (2008). Their proposed model captured the dynamics of customer relationships, but also took into account the effect of encountering with the firm and its impact on buying behavior.

In a more recent research, Chang C. (2012) developed a non-homogeneous multiple-segment hidden Markov model to capture dynamic variation and cross-sectional heterogeneity in customers’ channel choice.

Finally, hidden Markov models have been used by researchers in order to estimate the impact of personal characteristics or demographics on the selection of products. Ebbes et al. (2010) used such models to estimate the dynamics of strategic groups. Also, Du and Kamakura (2006) used data from the American market to study the household life cycle in combination with the quality of life. Moon et al. (2007) used hidden Markov models to determine the consumer behavior in combination with the unobserved promotion of the competitive products.

To conclude, previous research has defined cart abandonment, has concluded on the broad categories of factors that affect cart abandonment on online shops and has made a solid connection with the actual shops (if the brand has any).

Regarding mathematics and stochastic processes, there is little previous research and it is mostly on clickstream analysis, browsing behavior, tracking shopping processes, customer relation management applications and customer heterogeneity.

3. A mathematical approach

3.1. Markov Chains

A Markov chain is a mathematical system that undergoes transitions from one state to another (or remain to the same state), between a finite or countable number of possible states (S_1, S_2, …, S_N). A Markov chain is a memoryless process since each transition is only depended on the current state of the system and not on the sequence of previous events.

$$P[q_t=S_j | q_{t-1}=S_i, q_{t-2}=S_k, \ldots] = P[q_t=S_j | q_{t-1}=S_i], \; q_t \text{ is the state on time t.}$$

This mathematical system is named after Andrey Markov, a Russian Mathematician (1856 - 1922) known for his work on stochastic processes[2].

A Markov chain is a random process which in each discrete-time (or step) is in a certain state. Such a system changes randomly, making it generally impossible to predict with certainty the state of the chain at a given point in the future. However, the probability of each state to be the current state on a given step in the future can be calculated.

In order to completely characterize a Markov chain there should be stated all the available states and the transition probabilities. A transition probability is the probability of transition from one specific state to another specific one (e.g. from state A to state B).

$$P_{ij} = P[q_t=S_j | q_{t-1}=S_i], \; 1 \le i,j \le N$$

$$0 \le p_{ij} \le 1$$

$$\sum_{j=1}^{N} p_{ij} = 1$$

[2] http://en.wikipedia.org/wiki/Andrey_Markov

A simple Markov Chain example is described by figure 4.

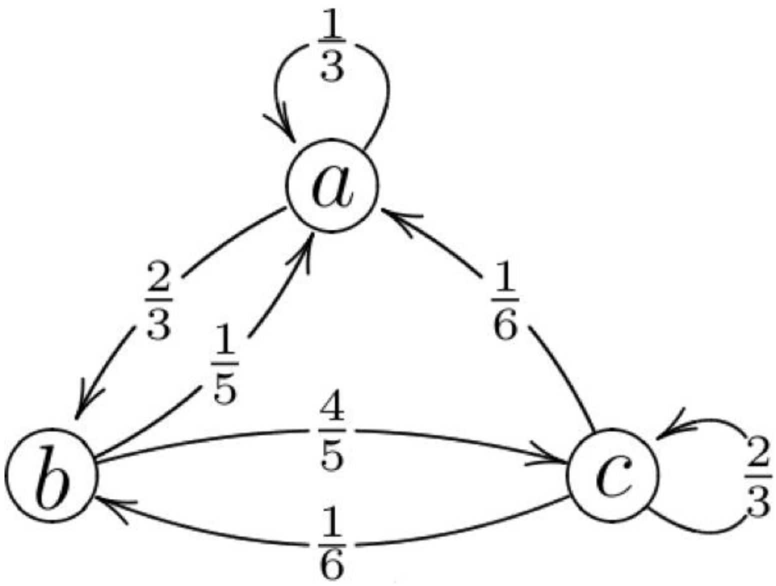

Figure 4 - A simple Markov chain

In figure 4, both the available states and the transition probabilities are mentioned. For instance, if in a given time the system is on state “a”, in the next step it will be again in state “a” with probability 1/3 or in state “b” with probability 2/3. Likewise, if the system is in state “c”, in the next step will be on state “a” with probability 1/6, in state “b” with probability 1/6 or it will remain on state “c” with probability 2/3.

Alternatively, a Markov Chain may be given as a two-dimensional matrix, known as “transition matrix”.

$$\mathbf{P} = \begin{bmatrix} \boldsymbol{p_{aa}} & \boldsymbol{p_{ab}} & \boldsymbol{p_{ac}} \\ \boldsymbol{p_{ba}} & \boldsymbol{p_{bb}} & \boldsymbol{p_{bc}} \\ \boldsymbol{p_{ca}} & \boldsymbol{p_{cb}} & \boldsymbol{p_{cc}} \end{bmatrix} = \begin{bmatrix} \mathbf{1/3} & \mathbf{2/3} & \mathbf{0} \\ \mathbf{1/5} & \mathbf{0} & \mathbf{4/5} \\ \mathbf{1/6} & \mathbf{1/6} & \mathbf{2/3} \end{bmatrix}$$

In this matrix are included all the necessary information (states and transition probabilities) to define a Markov chain.

Given the transition matrix and the current state of the system (x^n), it is possible to calculate the probabilities for each state to be the current state at a later period. For example, if at present (time n) the system is on state "b", then 3 time periods later (at time n+3) the distribution will be:

$$x^{(n+3)} = x^{(n)} P^3$$

$$\mathbf{x}^{(n+3)} = [\mathbf{0} \quad \mathbf{1} \quad \mathbf{0}] \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{1}{5} & 0 & \frac{4}{5} \\ \frac{1}{6} & \frac{1}{6} & \frac{2}{3} \end{bmatrix}^3 = [\mathbf{0} \quad \mathbf{1} \quad \mathbf{0}] \begin{bmatrix} \frac{29}{135} & \frac{34}{135} & \frac{24}{135} \\ \frac{47}{225} & \frac{20}{90} & \frac{128}{225} \\ \frac{61}{270} & \frac{72}{270} & \frac{76}{135} \end{bmatrix}$$
$$= [\mathbf{0.209} \quad \mathbf{0.222} \quad \mathbf{0.569}]$$

Namely, three time periods later, the system will be on state "a" with probability 0.209, on state "b" with probability 0.222 or on state "c" with probability 0.569. By applying the same method, it is possible to calculate the state probabilities for any given future time.

Furthermore, there can be developed second order Markov chains. In these processes the next stage does not depend only on the current stage but it takes into account the last two ones. Likewise, there can be third order Markov processes and so forth. For example:

$$P[q_t=S_j | q_{t-1}=S_i, q_{t-2}=S_k, \ldots, q_{t-n}=S_q, \ldots] = P[q_t=S_j | q_{t-1}=S_i, q_{t-2}=S_k]$$

In general, there can be Markov chains of order m where m is finite. In these Markov chains the next state is depended on the previous m states:

$$P[q_n=S_n | q_{n-1}=S_{n-1}, q_{n-2}=S_{n-2}, \ldots, q_1=S_1] = P[q_n=S_n | q_{n-1}=S_{n-1}, q_{n-2}=S_{n-2}, \ldots, q_{n-m}=S_{n-m}], \text{ for } n>m.$$

Since the state probabilities for future steps are calculable, it is essential to use them in order to make decisions at present time. Such decisions, which are related to the cart abandonment phenomenon (Kim et al., 2010), may be, for example, whether to show some advertisement to the user or make some discount (Oliver et al., 2003) or, probably, provide him with related products (Ghani et al., 2002) or even change the look and feel of the website. The effect of such decisions can be easily calculated by using Markov decision processes.

3.2. Hidden Markov Model

A hidden Markov model (HMM) is a Markov model which contains unobserved (hidden) states. In Markov models, all states are visible and known to the observer, so the only parameter is the transition probabilities. On the contrary, in hidden Markov models the state is not visible by the observer. The observer can only see the output that depends on the state. Hidden Markov Models are widely used in speech recognition, biology, signal processing, neuro-linguistic programming and other scientific fields.

One of the most common hidden Markov model examples was first stated by Rabiner (1989) and is described by Figure 5. In this problem, there is a hidden room (not visible by the observer) which contains three baskets, X1, X2 and X3. Each basket contains a known mix of balls. Each ball is marked y1, y2, y3 or y4. Someone chooses one of the available baskets and draws a random ball from it. Then, he puts the selected ball somewhere were the observer can see it.

The man that selects the balls has some procedure to decide the basket from which he will draw the ball. This procedure is only depended on the previous selected basket. Meaning that the selection of n-th basket is only depended on the (n-1)-th basket.

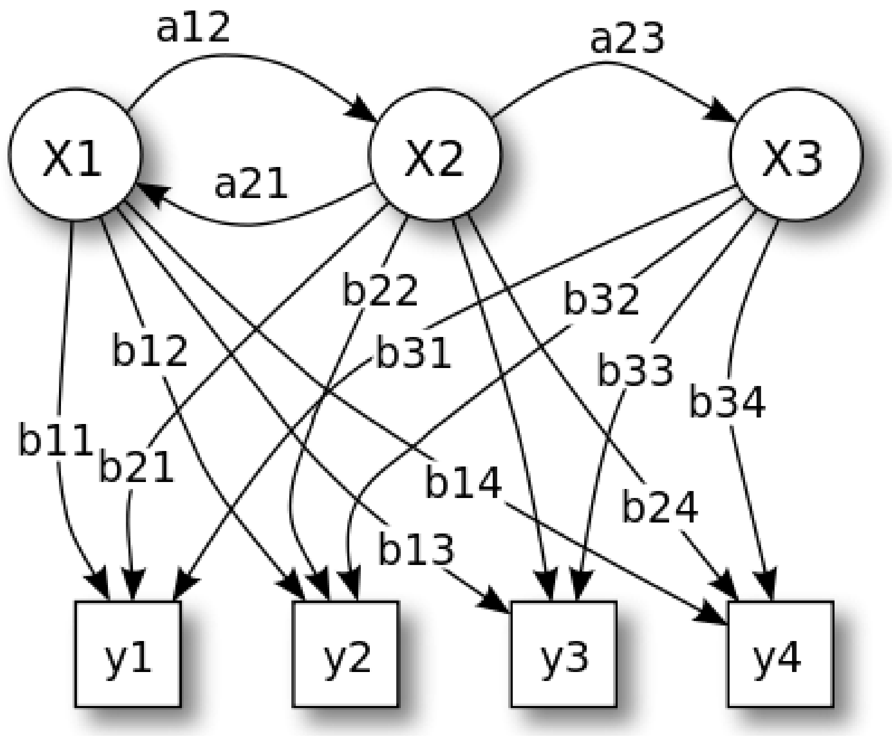

Figure 5 – A Hidden Markov Model by Rabiner (1989)

The actual process cannot be observed and the only available data for the observer is the sequence of labeled balls that he sees. Even if the observer is aware of the composition of the baskets, he still cannot be sure from which basket was the ball drawn. Based on the mathematics developed by Baum et al. (1970), the observer can work out other information, such as the likelihood that the n-th ball was drawn from each of the baskets.

In figure 5, X stands for the states, y stands for the possible observations, a_{ij} state the transition probabilities and b_{jk} the output probabilities.

In general, a hidden Markov model can be completely described by declaring the following:

- the hidden states

$$S = \{S_1, S_2, \dots , S_N\}$$

- the observable states

$$O = \{o_1, o_2, \dots, o_M\}$$

- the transition probabilities (from state i to state j)

$$a_{ij} = P[q_{t+1}=S_j | q_t=S_i]$$

- the transition probabilities for state j of observing the k-th item

$$B = \{b_j(k)\}$$

$$b_j(k) = P[o_k \text{ at } t | q_t=S_j]$$

$$1 \leq j \leq N$$

$$1 \leq k \leq M$$

- the initial probabilities

$$p = \{p_i\}$$

$$p_i = P[q_1 = S_i]$$

$$1 \leq i \leq N$$

Given the aforementioned data, the probability of observing a specific series of events (transitions) is calculated by:

$$P_i(\mathrm{O}_{i1} = \mathrm{o}_{i1}, \ldots, \mathrm{O}_{iT} = \mathrm{o}_{iT}) = \sum_{s_1=1}^{N} \sum_{s_2=1}^{N} \ldots \sum_{s_T=1}^{N} [P(S_{i1} = S_1) \prod_2^T P(S_{iT} = s_T | S_{iT-1} = s_{T-1}) \prod_1^T P(O_{iT} = o_T | S_{iT} = s_T)]$$

The aim of the model is to find the hidden state sequence that was most likely to have produced a given observation sequence (Blunsom, 2004). In this case the Viterbi algorithm can be applied in order to estimate the single best state sequence.

In its most general form, the Viterbi algorithm may be viewed as a solution to the problem of maximum a posteriori probability estimation of the state sequence of a finite-state discrete-time Markov process observed in memoryless noise (Forney, 1973). Nowadays, the Viterbi Algorithm is used in broad variety of applications including keyword spotting, speech synthesis, speech recognition, computational linguistics, bioinformatics and others.

Given a hidden Markov model where S is the state space, p_i the initial probabilities, a_{ij} the transition probabilities from state i to state j and $y_1, y_2, \ldots, y_t$ the observed outputs, the most likely state sequence $(x_1, \ldots, x_t)$ that produced these observations can be calculated by the following relations:

$$V_{1,k} = P(y_1 | k) p_k$$

$$V_{t,k} = P(y_t | k) \max(a_{x,k} V_{t-1,x})$$

In the above relations $V_{t,k}$ is the probability of the most probable state sequence responsible for the first t observations that has k as its final state. The Viterbi path can be retrieved by saving back pointers that remember which states x were used in the second equation. Ptr(k,t) is the function that returns the value of x used to compute $V_{t,k}$ if t > 1 or k if t = 1. Then:

$$x_T = \operatorname{argmax}(V_{T,x})$$

$$x_{t-1} = \mathrm{Ptr}(x_t, t)$$

Furthermore, hidden Markov models have the ability of training in order to be able to estimate the model parameters that best describe the process. Training can be either supervised or unsupervised, depending on the available data. If the input data is only provided, then there should be used the unsupervised training to guess a model that may have produced the provided observations. If both input and output data is given, then there can be performed supervised training by equating inputs to observations and outputs to states.

The supervised training is very useful in the proposed model (chapter 4.3) since it can assist on calculating the required transition probabilities for the initial implementation of the model (chapter 4.4). For this kind of training it can be used the Baum-Welch algorithm (Rabiner, 1989) which only requires a large amount of dataset with both inputs and outputs included.

The Baum–Welch algorithm can compute maximum likelihood estimates and posterior mode estimates for the transition probabilities, when given only inputs as training data. For a given probability in the transition matrix, all paths to that probability are summed and, also, there is a link (transition from S_i to S_j). The joint probability of S_i, the link and S_j can be calculated and normalized by the probability of the entire string (x). Then, there is calculated the probability of all paths with all links beginning from S_i and normalized by the probability of the entire string (t). Finally, the expected transition from S_i to S_j is divided with the expected transitions from S_i. As long as the data provided grows bigger, then some particular transitions will be reinforced and their value will be increased reaching a local maximum (MacKay, 1997).

4. The proposed model

The proposed model is based on a hidden Markov model similar to the one designed by Li and Chatterjee (2005). In order to construct the model, it is essential to describe extensively the different stages that a user walks through while browsing on an online shop. As discussed in chapter 2, the browsing process consists of 4 stages and follows the information processing theory of consumer choice (Howard and Sheth, 1969).

Each visit on a website is called session and lasts until the user quits the site, either having completed a purchase or not. If the user returns to the website later, this accounts for a new browsing session.

4.1. The session stages

4.1.1. Stage 1 – Information seeking

On the first stage the user is engaged to information seeking and processing activities. In this stage, the user is browsing through different pages of the website trying to collect information in order to complete his goal (e.g. make a purchase or compare products). Such information might be product details, characteristics and price, related products, complementary products or information about the store, its physical location, shipping costs or about the means of payment provided.

During the information seeking stage, the user might choose to use a series of features on the store, such as log in to his account (or create an account if he does not have one already), use the in-store search engine to find a specific category, product or brand, bookmark a page for visiting again later and so on. Engaging with such features might indicate the user's intentions. For example, Li and Chatterjee (2005) discovered that users that log in to their account early during their session, in opposition to those that do not log in at all, are purchase oriented users and it is more likely to buy something before leaving the store.

Additionally, the series of visited pages may discriminate “hedonic” from “buying” browsing. For example, hedonic – or information seeking – visitors tend to spent more time on information pages, so they have a longer page view duration. Another important piece of information is the referrer, namely the location from where the user arrived to the online store. A referrer might be a search engine or another website. Sessions without a referrer indicate that the user either typed the store address directly on the address bar or had previously bookmarked the webpage on the browser’s favorites.

4.1.2. Stage 2 – Consideration stage

Consideration stage accounts for adding a product to the virtual shopping cart. This process may vary from one click (no browsing disturbance) to several ones. Adding a product to the cart does not necessarily reflect an intention of buying. Hedonic or information seeking users may add multiple products on their cart only for bookmark or comparison reasons. Li and Chatterjee (2005) discovered that people with a buying intension tend to add fewer items on their carts mainly because they are aware of their needs and actual goals.

Consideration stage is very important because it indicates cart usage and thus accounts for cart abandonment sessions. After stage 2, a visitor may choose either to go back to stage 1 and continue browsing looking for more products, or proceed to stage 3 and initiate the checkout process.

4.1.3. Stage 3 – Evaluation stage

Evaluation stage is triggered when a user has finished selecting items and chooses to start the checkout process. In most cases, visitors that do not intent to buy, drop out of the online shop prior to this stage. Thus, evaluation stage indicates a strong intention of buying.

Usually, in this stage the user is provided with several information about his or her order, such as the selected items, the selected quantity of each item, their price, the total price, shipping costs and so on.

While on evaluation stage, the online shopper can make changes to his order (for example remove items from the cart or change the quantity of the selected items), choose to return to stage 1 (information seeking) in order to find more information on the products he selected or find more desired products or, finally, proceed to stage 4 which is the purchase decision stage.

4.1.4. Stage 4 – Purchase decision

The last session stage is the purchase decision. After stage 3 the customer may initiate the purchase process which consists of filling in all the necessary information for the order to proceed. Necessary information includes shipping address, online payment, etc.

Unlike brick and mortar stores, where people do not usually abandon at the checkout process, the online purchase decision process might be pretty tricky since the customer can abandon his or her cart quickly and easily without second notice and without feeling guilty or embarrassed. This might happen if he or she feels losing control of the process or insecure (for example about the money transfer). Quitting the process at this stage is also referred as "checkout abandonment" since the quitting reasons, usually, are not referred to the visitor's initial intentions but to store variables, such as shipping costs or means of payment (Websitemagazine, May 2013).

During purchase decision stage, the user can still go back to stage 1 to find more products or useful details or to stage 3 in order to make changes to his order before initiating checkout again.

4.2. The Markov model

4.2.1. Time independent model

The aforementioned online shopping process can be formed into a Markov model. This model consists of six stages: stages one to four (S1, S2, S3, S4) plus one stage for quitting the store without buying something (EF) and one stage for quitting the store having made a purchase (EW).

There have already been mentioned the possible transactions between the different stages, but for the model to be complete, it should be provided with the transition probabilities as well. These probabilities are currently unknown, so they will be referred with two indexes, one representing the starting state and the other the ending one. For example, the probability for transitioning from stage one to stage two will be P_{12}, or the probability quitting from stage 2 will be P_{2EF}. Therefore, the transition matrix should be:

$$\mathbf{P} = \begin{bmatrix} \boldsymbol{P_{11}} & \boldsymbol{P_{12}} & \boldsymbol{P_{13}} & \mathbf{0} & \boldsymbol{P_{1EF}} & \mathbf{0} \\ \boldsymbol{P_{21}} & \mathbf{0} & \boldsymbol{P_{23}} & \mathbf{0} & \boldsymbol{P_{2EF}} & \mathbf{0} \\ \boldsymbol{P_{31}} & \mathbf{0} & \mathbf{0} & \boldsymbol{P_{34}} & \boldsymbol{P_{3EF}} & \mathbf{0} \\ \boldsymbol{P_{41}} & \mathbf{0} & \boldsymbol{P_{43}} & \mathbf{0} & \boldsymbol{P_{4EF}} & \boldsymbol{P_{4EW}} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix}$$

Some notes on the transition probabilities:

- P_{11} stands for the probability to visit another informational page consecutively. This is the only case a user to be on the same stage for two consecutive time periods.
- Leave the online store after purchased something can be occurred only through stage 4. Thus, P_{4EW} is the only possibility for “EW”.
- The last two rows are for “EF” and “EW”. Both stages represent the exit from the site. These are absorption stages since the user once reached them, cannot return to any other stage. Going back to the site triggers a new visiting session.

- The probabilities P_{EFEF} and P_{EWEW} are 1 only to make the model complete. They do not represent an actual event or transition.
- After a customer purchases successfully he might choose to continue browsing the site to gather information on other products. This event should trigger a new session.
- Wherever the probability is 0, the corresponding transition is impossible according to the session structure described earlier.

We can create a visual graph of the Markov model using Mathematica 9.0 (figure 6). The transition matrix can be declared with the "DiscreteMarkovProcess" option. The "DiscreteMarkovProcess" option combined with "Graph" can result to the drawing of the model (figure 6).

Input: Graph[{"Stage 1", "Stage 2", "Stage 3", "Stage 4", "EF", "EW"}, DiscreteMarkovProcess [3, {{P11, P12, P13, 0, P1EF, 0}, {P21, 0, P23, 0, P2EF, 0}, {P31, 0, 0, P34, P3EF, 0}, {P41, 0, P43, 0, P4EF, P4EW}, {0, 0, 0, 0, 1, 0}, {0, 0, 0, 0, 0, 1}}]]

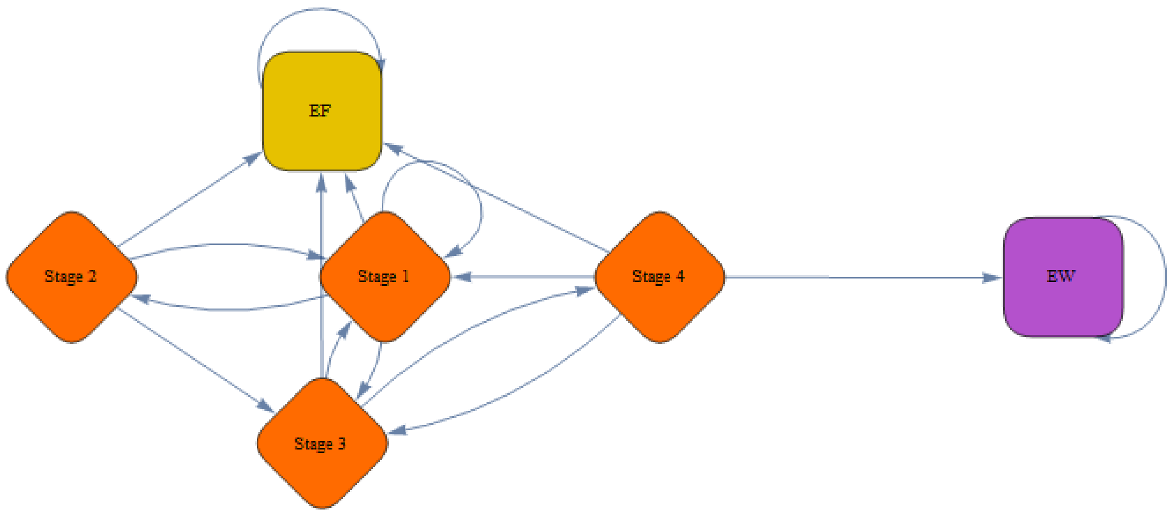

Figure 6 – Markov model graph designed by Mathematica 9.0

The online store manager can use the aforementioned Markov model to produce useful insights about the buying experience his shop offers. By using suitable scripts or applications on the website, the manager can track a series of data and actions or events that occur on the store. The data analysis will calculate all the transition probabilities that are required by the model, which then will give information on the strong and weak parts of the store's buying experience.

The calculation of the probabilities might indicate weaknesses on the system that were not obvious in advance. For example, if the P_{4EF} probability (people that exit the site although they had already begun the checkout process) is high, then probably the checkout process should be improved in order to reduce the cart abandonment phenomenon. Furthermore, the manager can compare this data with the data of the general industry or its competitors (if similar data are available or can be obtained) in order to evaluate further the store's performance.

4.2.2. Time dependent model

In most real life occasions the visitor is affected and influenced by the store and the information provided. In these situations, the transition probabilities are not constant throughout the whole browsing session but change according to the visitors' actions and choices (Mandel and Johnson, 2002; Sismeiro and Bucklin, 2004; Montgomery et al., 2004; Hauser et al, 2009).

Such occasions can be approached with the non-homogeneous Markov chains. The non-homogeneous Markov Chain model does not assume a homogeneous behavior of the stochastic process. In other words, the transition probability matrix is not a constant but a function of time (Fu et al, 2008).

4.3. The hidden Markov model

Although the aforementioned Markov model (either time dependent or independent) produces some interesting results, it still has some important restrictions. The model might indicate spots on the store that need to be improved, but it does not provide any information regarding the general direction of the improvement. For example, if the probability P_{12} (this probability states the addition of a product in the virtual shopping cart) is low then maybe the user is not satisfied by product description pages, but there is no information on what a satisfactory product description page would look like, what information would it provide or in what order, how long should it be and so on.

This model does not take under consideration neither the customer heterogeneity (gender, age, income, family or marital status, race, etc.) nor the interactions that the customer had while on the internet (for instance banner advertisement, promotional messages, usage of website features and more). Such constraints can be overcome by an appropriately designed hidden Markov model, which can provide information on some of the user heterogeneity as far as user characteristics and demographics are concerned.

It is, therefore, important for the model to take under consideration the visitor's unobserved purchase intention. This will be resolved by applying a latent (or hidden) level which will determine the visitor's behavior, thus the probabilities to trigger certain events (such as putting an item in the virtual shopping cart) while browsing. Such intentions do not stand idle during browsing on the website but might change depending on the interactions that the user has with or on the shop (Montgomery et al., 2004).

Moreover, Bloch et al. (2004) concluded that the unobserved purchase intention is influenced by many intrinsic or extrinsic factors and may lead to a variety of different outcomes. The impact of the stimuli that the user absorbs from the store may accelerate or distract from the purchase (Li and Chatterjee, 2005).

On the other hand, one of the most challenging disadvantages or difficulties of the hidden Markov model is that it is expensive in means of data and computer power required. Most of the data gathering can be easily and quickly done by using special scripts or applications that track and record all the events that a user triggers on an online store.

Also, the ability to connect an e-commerce site with the widely used social media can harvest even more user data (for instance gender, age, hobbies and other interests, profession, location, etc.). Thus, the amount of gathered data is very big and not easily analyzed.

The challenge, therefore, lies upon the integration and the analysis of the data which can be done by using statistics and mathematics software packages as well as programming scripts.

Browsing sessions can be divided into four – unobserved – categories (Moe, 2003):

a) Buying
b) Browsing
c) Searching
d) Knowledge building

There might be a relation between these latent categories and the session stages (information seeking, consideration, evaluation, purchase) the user walks through. As mentioned earlier, the visitor's initial intention may change during the browsing session. The tendency of moving from an intention to another is influenced by a series of events, actions or characteristics. These factors are analyzed in the following categories:

1. Demographic variables

Age: The age of the visitor that is browsing the online shop plays a vital role on the purchase intention. Different ages have a different behavior, as far as shopping cart is concerned. For instance, younger customers may add easily a product in their cart but are less likely to complete the purchase (Hoyer and MacInnis 2008).

Gender: The gender of the potential customer in comparison with the shop's target group (Hoyer and MacInnis 2008).

Race: Depending on the country, race might be an important factor (Hoyer and MacInnis 2008).

Marital and family status: Married customers or customers with children have different needs, preferences and habits (Hoyer and MacInnis 2008).

Financial Status: Financial status or income is a factor that influences the consumer's behavior. Income not only affects the purchase intention but the amount that the customer is willing to spend as well. Also, customers with financial issues might add products to their carts without true intention of buying (Hoyer and MacInnis 2008).

2. Page-Specific variables: information or promotions that might be available on a product page (Li and Chatterjee, 2005).

Price: Price is important information for the purchase process. However, when stated it might have a varying effect to the session's intention. For example, customers may be discouraged to add a product in their shopping cart because of the price, although they are willing to pay such a price for that product.

Promotional messages: Promotional messages or campaigns might change the user's orientation. Customers that are on browsing session may purchase a product if the promotion is appealing. On the other hand, promotional messages might have a negative impact. For instance, promotions that refer to buying more than one item for a lower price, might distract customers from their initial intentions and quit store or abandon cart.

Banner advertisement: Banner advertisement can have a positive impact to information seeking visitors and convert sessions from browsing to purchasing. However, banners might distract people from their initial goal by presenting them alternative opportunities. Such a distraction might lead to cart abandonment.

Hypertext links: The presence of hyperlinks (or anchors) to other pages (such as the home page or the category page) may have a different result depending on the orientation of the visitor. Purchase oriented visitors might find the presence of hyperlinks distractive, while browser oriented ones may use them to find more suited information and end up buying something.

3. Session-specific variables (Li and Chatterjee, 2005)

Previous purchase: The familiarity of the website and the online purchase process is a strong intention indicator. Visitors that have bought something from the online store in the past are more likely to purchase something during the browsing session.

Time duration: The time that the visitor stays on a specific page may have an effect on his buying intention. Sessions with longer or shorter duration may indicate intention of buying or information seeking.

Visit depth: The visit depth stands for the amount of pages a visitor views during his or her session. This information may reveal intention against browsing or purchasing.

Use website services: Extended usage of website services, such as internal search engine, log in to account, wishlist and so on, might indicate towards a purchasing intention session.

4. Comparison shopping activities (Li and Chatterjee, 2005)

The last section is about the visitor's behavior not at the online store, but on other sites related to the shop or not. This kind of data is more difficult to acquire since in most cases might require the cooperation of competitive or external companies and stores. Despite the difficulty in collecting the data, this information may reveal a lot of useful insights about the visitors' true shopping intentions.

Comparison shopping sites: if the user has made any searches on comparison shopping sites, such as booking.com for hotels for instance. Extended former research can provide information on the purchase intention but on the available amount of money as well.

Number of page views at other similar online shops: Whether the visitor has browsed other competitive online shops before in order to collect information on product details and prices.

Number of page views at other non-similar online shops: Whether the visitor has the tendency to visit online stores. Spending time on a lot of stores selling different kind of products might indicate a lower purchase intention, since the user tends to browse for hedonic, information seeking or other reasons.

It should be noted that each shop manager should choose which of the aforementioned factors are playing a key role to his or her company and monitor them. Factors that can be calculated programmatically by the website, such as the page-specific and the session-specific variables are the observed variables that are going to be used into the hidden Markov model. The latent variables of the model consist of the demographic variables and the comparison shopping activities, that cannot be calculated or found by the website's algorithms and, thus, have to be estimated by the model.

Furthermore, store managers might conclude into adding more factors to their own list, provided that there is a way to gather the necessary information from the users. It is not necessary to focus on all of the factors. Also, each manager should analyze the data to determine which factors are playing an important role to the cart abandonment phenomenon or to the shop's general function.

4.4. Proposed model implementation

In reality, when a user is browsing throughout an online shop, various online scripts are able to track visitor's actions and collect some of the aforementioned information. So, the store itself holds some information about the current user. Such information might be his or her gender, his or her age, statistics about his or her previous visits (such as time on site, visit depth and other) and so on. All the rest information that is not known to the website comprises the latent (or unobserved) information. This latent information is estimated by the hidden Markov model that was described earlier.

The implementation of the model is divided into two steps. In the first step, there should be some prerequisite actions to make the model applicable to the actual site. During this first, preliminary, step, the store's manager collects data from previous visits on the site, for example from all the sessions during the past months, and runs the model in order to calculate the necessary probabilities. The more the available data, the more accurate the calculated probabilities would be. Though, relatively old data might be proven misleading, since consumer's behavior changes continually over the time.

This step corresponds to the supervised training process described at chapter 3.3. Both inputs and outputs are known to the store's manager and can be used in order to accurately calculate the required probabilities.

Had the probabilities been calculated, the model is ready to be attached on the online store itself. Each visitor's session is divided into two main parts. The first part (the earlier one) is used for estimation reasons while the second part (the later one) is used for predictions. If a visitor has already been in the website in the past, then the first part can be skipped, since for the estimation can be used the data from his previous visits.

During the estimation part, the model follows the user through his browsing session and tracks several pieces of data (mainly clickstream data and any of the demographical or behavioral characteristics possible). Based on the collected data, the model creates a series of probabilities for each one of the latent factors that the store's manager has chosen to track. Namely, following the hidden Markov model description made on chapter 3.3., the model "views" the observable states (estimation part) and tries to predict the hidden ones (prediction part) given the transition probabilities calculated on the preliminary step.

Then, during the prediction part, the store holds some probabilities about the visitor's characteristics. For example, the model can predict that the visitor is male with a probability of p and female with a probability of 1-p. Of course, the same applies to all the unobserved (latent) states, which are the factors chosen from the manager. The longest the visitor stays on the online store, the more accurate the predictions are becoming.

Based on these predictions, the store manager can assign certain actions to be granted. For example, if the model indicates that the user is female with a probability over a defined percentage (i.e. 70%), then the website's design might automatically change as well as the related products, the product's description and so on. Alternatively, if the model indicates that the user is within certain range of age, then the store will take some other initiative in order to fit this characteristic. The target is that the store takes initiatives based on the combined probabilities of all the predicted characteristics, in order to aim for the best result possible (sale).

5. Managerial implications

The insights from the implementation of the aforementioned model provide a prescriptive tool for managers aiming to decrease the shopping cart abandonment rate, among other interesting characteristics. The results can be grouped into two broad categories.

Firstly, store's manager can acquire information on the way people tend to use store's website (e.g. for seeking information, for comparing products, for purchasing, etc.) and secondly he or she can take actions (change or initiate advertising campaigns, redesign the website, change the online buying procedure, etc.), towards the increase of purchase intention and sales.

As long as the website functionality is concerned, there are a lot of parameters that can be changed based on the results of the model. Some proposed examples (Naidu et al., 2007; Kotze et al., 2002) include the name of the shopping cart, the title of the "buy now" or the "add to cart" button, the communication messages that assist the visitor on his or her shopping, the required steps to complete the purchase, the visibility of the price, the signing in process and more.

The store is not a static unity that is designed to serve simultaneously all the different kinds of customers or visitors. Had this model been applied, the store will be dynamic and able to adapt instantly on the customers characteristics and needs, as they have been estimated by the model. For example, by taking into account the model's estimations for the user's gender or age or financial status (latent variables), can change the design of the website (colors, font faces, font sizes, widget positioning, etc.), the suggested related products, the banner advertisements or even the titles or the tags of the elements (buttons, functions and so on). Doing so will encourage users to go through the entire shopping process successfully, leading to more sales.

The website can use the model's predictions in order to decide on the customer's characteristics and adopt the outfit that suits best. The manager is advised to go through the preliminary process of the model's implementation where the prior data is used to calculate the initial transition probabilities. In the same stage, the manager can run a cluster analysis on the same dataset in order to determine the different clusters of customers that visit the website (Fayyad et al, 1996; Adomavicius and Tuzhilin, 2001). When the model will be embodied on the website, it will first categorize the visitor on one of those clusters in order to start adjusting the outlook and its parameters before deciding on the specific characteristics of the user with a higher probability rate.

Also, the data acquired by the hidden Markov model could provide valuable information on the ways customers are finding the store and how effective, in the mean of purchases, these ways are. For example, since the model can target the entrance sources (e-mail campaign, banner advertisement on other website, link on other website, etc.), the manager can use this information in order to design appropriately his or her next promotional campaigns. Also, he or she may take other useful information as well, such as the amount of money someone is willing to spend on the store in combination with the entrance source.

Using this information, the manager can decide the products or the services he or she will promote through each available channel. For example, if e-mail marketing attracts consumers with higher income than advertising on social media or other websites, then promoted products can be rearranged on the different campaigns according to their price or to their targeted audience. If all the latent variables are taken under consideration, the manager is available to build more effective and targeted campaigns.

Finally, regarding the cart abandonment phenomenon, the manager can choose to take under consideration only the data form the sessions where the cart abandonment occurred. By doing so, he or she can decide on actions that should be taken towards satisfying customer's needs that concern the online experience offered by the store.

Such actions should have a positive effect on visitor’s intentions towards buying (against browsing, searching or knowledge building).

To conclude, applying the proposed hidden Markov model has both methodological and practical advantages. Regarding the suggested methodology, the proposed model can produce a series of data whose analysis can indicate the way people come to the online store, the way they use it and the way the leave from it. Thus, apart from any implementation, the manager can acquire a better view of the impact the website has on the consumers and their buying or non-buying intentions. This result is of a great assistance since it does not affect only the functionality of the website, but the company’s management as well.

Regarding the practical advantages, the implementation of the proposed model can convert the online store into a dynamic environment adapted to the personalized needs of each individual customer. In this way, consumers find a store that is suitable to their own desires and needs and by doing so it is maximized the time they spent on the store, the different products they browse through and, finally, the store’s conversion rate.

Appendix 1 – Tables

Table 1: Eurostat – Online shopping data (March 2013)

Country or Geographic region	Internet use: Internet banking	Internet use: selling goods or services	Last online purchase: in the 12 months	Online purchases: computer software	Individuals who ordered goods or services over the Internet from sellers from other EU countries in the last 12 months	Individuals who ordered goods or services over the Internet from sellers from other countries (EU or non EU) in the last 12 months	Individuals who ordered goods or services over the Internet from sellers from the rest of the world (non-EU) in the last 12 months	Individuals who ordered goods or services over the Internet from sellers with unknown country of origin the last 12 months	Individuals who booked travel and holiday accommodation over the Internet in the last 12 months	Online purchases: video games software and upgrades	Online purchases: computer software other than video games and upgrades	Individuals who placed a bet or play gambling or lotto over the Internet in the last 12 months
European Union (27 countries)	42	17	45	12	19	21	8	3	29	7	8	:
European Union (25 countries)	:	:	:	:	:	:	:	:	:	:	:	:
European Union (15 countries)	:	:	:	:	:	:	:	:	:	:	:	:
Belgium	55	14	44	9	31	34	11	:	30	5	5	:
Bulgaria	:	:	:	:	:	:	:	:	:	:	:	:
Czech Republic	:	:	:	:	:	:	:	:	:	:	:	:
Denmark	:	:	:	:	:	:	:	:	:	:	:	:
Germany (including former GDR from 1991)	34	18	46	15	9	10	4	:	25	9	10	:
Estonia	:	:	:	:	:	:	:	:	:	:	:	:
Ireland	50	16	54	15	36	39	16	:	43	9	11	:

Greece	:	:	:	:	:	:	:	:	:	:	:	:
Spain	35	11	39	8	17	22	10	:	27	:	:	:
France	50	21	48	8	23	26	12	:	36	:	:	:
Italy	:	:	:	:	:	:	:	:	:	:	:	:
Cyprus	23	:	27	4	26	26	10	:	18	:	:	:
Latvia	40	:	:	:	:	:	:	:	:	:	:	:
Lithuania	:	:	:	:	:	:	:	:	:	:	:	:
Luxembourg	65	16	64	22	55	60	14	:	52	10	18	:
Hungary	:	:	29	:	:	:	:	:	:	:	:	:
Malta	:	:	:	:	:	:	:	:	:	:	:	:
Netherlands	:	:	:	:	:	:	:	:	:	:	:	:
Austria	53	16	39	13	26	27	:	:	25	6	9	:
Poland	:	:	:	:	:	:	:	:	:	:	:	:
Portugal	44	:	48	:	26	29	:	:	25	:	:	:
Romania	:	:	:	:	:	:	:	:	:	:	:	:
Slovenia	:	:	:	:	:	:	:	:	:	:	:	:
Slovakia	:	:	:	:	:	:	:	:	:	:	:	:
Finland	94	:	77	:	:	:	:	:	67	:	:	:
Sweden	86	:	71	:	:	:	:	:	61	:	:	:
United Kingdom	63	:	74	:	:	:	:	:	52	:	:	:
Iceland	85	:	52	30	41	46	:	:	35	:	:	:
Norway	90	:	81	:	:	:	:	:	62	:	:	:
Croatia	:	:	:	:	:	:	:	:	:	:	:	:
FYROM	:	:	:	:	:	:	:	:	:	:	:	:
Serbia	:	:	:	:	:	:	:	:	:	:	:	:
Turkey	:	:	:	:	:	:	:	:	:	:	:	:

Bibliography

Papers

- Adomavicius, Gediminas, and Alexander Tuzhilin. "Using data mining methods to build customer profiles." *Computer* 34.2 (2001): 74-82.
- Baum, Leonard E., et al. "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains." *The annals of mathematical statistics* 41.1 (1970): 164-171.
- Bellman, Richard. *A Markovian decision process*. No. P-1066. RAND CORP SANTA MONICA CA, 1957.
- Blunsom, Phil. "Hidden markov models." *Lecture notes, August* (2004).
- Bloch, Peter H., Daniel L. Sherrell, and Nancy M. Ridgway. "Consumer search: an extended framework." *Journal of consumer research* (1986): 119-126.
- Chang, Chun-Wei. *Multichannel Marketing and Hidden Markov Models*. Diss. University of Washington, 2012.
- Ching, Wai-Ki, Michael K. Ng, and Ka-Kuen Wong. "Hidden Markov models and their applications to customer relationship management." *IMA Journal of Management Mathematics* 15.1 (2004): 13-24.
- Cho, Jinsook. "Likelihood to abort an online transaction: influences from cognitive evaluations, attitudes, and behavioral variables." *Information & Management* 41.7 (2004): 827-838.
- Cho, Chang-Hoan, Jaewon Kang, and Hongsik John Cheon. "Online shopping hesitation." *CyberPsychology & Behavior* 9.3 (2006): 261-274.
- Close, Angeline G., and Monika Kukar-Kinney. "Beyond buying: Motivations behind consumers' online shopping cart use." *Journal of Business Research* 63.9 (2010): 986-992.
- Du, Rex Y., and Wagner A. Kamakura. "Household life cycles and lifestyles in the United States." *Journal of Marketing Research* (2006): 121-132.
- Ebbes, Peter, Rajdeep Grewal, and Wayne S. DeSarbo. "Modeling strategic group dynamics: A hidden Markov approach." *QME* 8.2 (2010): 241-274.
- Egeln, Laura S., and Julie A. Joseph. "Shopping Cart Abandonment in Online Shopping." *Atlantic Marketing Journal* 1.1 (2012): 1.
- Fayyad, Usama, Gregory Piatetsky-Shapiro, and Padhraic Smyth. "From data mining to knowledge discovery in databases." *AI magazine* 17.3 (1996): 37.
- Forney Jr, G. David. "The viterbi algorithm." *Proceedings of the IEEE* 61.3 (1973): 268-278.

- Fu, Gongkang, and Dinesh Devaraj. *Methodology of Homogeneous and Non-homogeneous Markov Chains for Modelling Bridge Element Deterioration*. Michigan Department of Transportation, 2008.
- Ganesh, Jaishankar, et al. "Online shopper motivations, and e-store attributes: An examination of online patronage behavior and shopper typologies." *Journal of Retailing* 86.1 (2010): 106-115.
- Ghani, Rayid, and Andrew Fano. "Building recommender systems using a knowledge base of product semantics." *Proceedings of the Workshop on Recommendation and Personalization in ECommerce at the 2nd International Conference on Adaptive Hypermedia and Adaptive Web based Systems*. 2002.
- Hauser, John R., et al. "Website morphing." *Marketing Science* 28.2 (2009): 202-223.
- Howard, John A., and Jagdish N. Sheth. "The theory of buyer behavior." (1969).
- Howard, Ronald A. "Dynamic programming and Markov Processes." (1960).
- Kim, Jong Uk, Woong Jin Kim, and Sang Cheol Park. "Consumer perceptions on web advertisements and motivation factors to purchase in the online shopping." *Computers in human behavior* 26.5 (2010): 1208-1222.
- Kotzé, Paula, Karen Renaud, and Tobias Van Dyk. "Feedback and Task Analysis for E-Commerce Sites."
- Kukar-Kinney, Monika, and Angeline G. Close. "The determinants of consumers' online shopping cart abandonment." *Journal of the Academy of Marketing Science* 38.2 (2010): 240-250.
- Li, Shibo, and Patrali Chatterjee. "Shopping Cart Abandonment at Retail Websites-A Multi-stage Model of Online Shopping Behavior." *UCR Sloan Center for Internet Retailing* (2005).
- Littman, Michael L., Thomas L. Dean, and Leslie Pack Kaelbling. "On the complexity of solving Markov decision problems." *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 1995.
- MacKay, David JC. *Ensemble learning for hidden Markov models*. Technical report, Cavendish Laboratory, University of Cambridge, 1997.
- Mandel, Naomi, and Eric J. Johnson. "When web pages influence choice: Effects of visual primes on experts and novices." *Journal of Consumer Research* 29.2 (2002): 235-245.
- Moe, Wendy W. "Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream." *Journal of Consumer Psychology* 13.1 (2003): 29-39.
- Moe, Wendy W., and Peter S. Fader. "Dynamic conversion behavior at e-commerce sites." *Management Science* 50.3 (2004): 326-335.
- Montgomery, Alan L., et al. "Modeling online browsing and path analysis using clickstream data." *Marketing Science* 23.4 (2004): 579-595.

- Moon, Sangkil, Wagner A. Kamakura, and Johannes Ledolter. "Estimating promotion response when competitive promotions are unobservable." *Journal of Marketing Research* (2007): 503-515.
- Moore, Simon S., and Shane W. Mathews. "An exploration of Online Shopping Cart Abandonment Syndrome–a matter of risk and reputation." *Journal of Website Promotion* 2.1/2 (2006): 71-88.
- Naidu, Shivashankar, and Barbara S. Chaparro. "Top Ten Mistakes of Shopping Cart Design Revisited: A Survey of 500 Top E-Commerce Websites." *Usability news* 9.2 (2007).
- Netzer, Oded, James M. Lattin, and V. Srinivasan. "A hidden Markov model of customer relationship dynamics." *Marketing Science* 27.2 (2008): 185-204.
- Oliver, Richard L., and Mikhael Shor. "Digital redemption of coupons: satisfying and dissatisfying effects of promotion codes." *Journal of Product & Brand Management* 12.2 (2003): 121-134.
- Poddar, Amit, Naveen Donthu, and Yujie Wei. "Web site customer orientations, Web site quality, and purchase intentions: The role of Web site personality." *Journal of Business Research* 62.4 (2009): 441-450.
- Punj, Girish, and Robert Moore. "Information search and consideration set formation in a web-based store environment." *Journal of Business Research* 62.6 (2009): 644-650.
- Rabiner, Lawrence R. "A tutorial on hidden Markov models and selected applications in speech recognition." *Proceedings of the IEEE* 77.2 (1989): 257-286.
- Rajamma, Rajasree K., Audhesh K. Paswan, and Muhammad M. Hossain. "Why do shoppers abandon shopping cart? Perceived waiting time, risk, and transaction inconvenience." *Journal of Product & Brand Management* 18.3 (2009): 188-197.
- Scott, Steven L., and Il-Horn Hann. "A nested hidden markov model for Internet browsing behavior." (2006).
- Sismeiro, Catarina, and Randolph E. Bucklin. "Modeling purchase behavior at an e-commerce web site: a task-completion approach." *Journal of Marketing Research* (2004): 306-323.
- Wang, Shi, et al. "Adaptive online retail web site based on hidden markov model." *Web-Age Information Management*. Springer Berlin Heidelberg, 2000. 177-188.
- Wang, Sijun, Sharon E. Beatty, and David L. Mothersbaugh. "Congruity's role in website attitude formation." *Journal of Business Research* 62.6 (2009): 609-615.
- Hoyer W., and MacInnis D. "Consumer Behavior (5rd)." (2008).

Websites

- Websitemagazine (May 2013), Checkout vs. Cart Abandonment. Available online at: http://www.websitemagazine.com/content/blogs/posts/archive/2013/05/06/checkout-vs-cart-abandonment.aspx (June 16th, 2013)
- Eurostat (March 2013), E-Banking and e-Commerce. Available online at: http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/dataset?p_product_code=ISOC_BDE15CBC (March 1st, 2013)
- Ecommerce Europe (November 2012), European E-commerce to reach over € 300 billion in 2012. Available online at: http://www.ecommerce-europe.eu/press/2012/12/european-e-commerce-to-reach-over-300-billion-in-2012 (March 14th, 2013)
- Invesp consulting (2011), How big is ecommerce? Available online at: http://www.invesp.com/ecommerce.jpeg (March 2nd, 2013)
- Search Marketing Standard (July 2007), What is the average conversion rate? Available online at: http://www.searchmarketingstandard.com/what-is-the-average-conversion-rate (April 3rd, 2013)
- Smart Insights (March 2013), Ecommerce conversion rates. Available online at: http://www.smartinsights.com/ecommerce/ecommerce-analytics/ecommerce-conversion-rates/ (April 3rd, 2013)
- Ouellet M. (2010), Recovering lost sales through an automated shopping cart abandonment strategy. Available online at listrak.com website at: http://www.listrak.com/Whitepaper/Recovering-Lost-Sales-Through-Shopping-Cart-Abandonment.pdf
- Goldwyn C. (2002), The art of the cart: Why people abandon shopping carts. Available at visibility.tv: http://visibility.tv/tips/shopping_cart_abandonment
- WorldPay, Reasons for consumers to drop out of an online purchase in 2012, 2012. Available at http://www.statista.com/statistics/232285/reasons-for-online-shopping-cart-abandonment/