Utrecht University

User Profiling Techniques:

A comparative study in the context of e-commerce websites

Author: Maaike Fleuren (3471128)

Supervisor: dr. ir. J.M. Broersen

July 25, 2012

Abstract

Today's society is filled with a lot of information. It has become difficult to find the information we are looking for on the World Wide Web. When using e-commerce websites this is believed to lead to less sales. In order to facilitate this problem many websites can adjust the information that is presented to users based on the users' needs. To achieve this several intelligent techniques can be used. In this study a selection of these techniques are explained and compared in order to give the audience a sense of the different possibilities accompanied by the strong aspects and pitfalls of these different techniques. The selected techniques are: Bayesian networks, decision trees, case-based reasoning, association rules and neural networks.

Bachelor's Thesis Artificial Intelligence (7,5 ECTS)

Table of Contents

Introduction

In today's society, where people receive an overload of information, it becomes more and more difficult for users on the internet to find the information they need. Too much information can lead to cognitive overload, thus creating an unpleasant environment for customers of an e-commerce website. Studies quoted by Jansen (2012) show that online the conversion rate is somewhere in the region of 1 – 3.5% compared to about 20 – 25% in offline stores. Although a lower conversion in online stores is not surprising, a difference of a factor of 25 is quite stunning. Clearly there is great potential in improving the way information is presented to visitors of an online shop.

To enable the operator of a website to offer visitors tailored information, it is important to gather and process information about the users. Often websites already show suggestions for products that other people also bought when they bought a product that is being viewed. This is a rather simple and not always reliable way of making suggestions to a user. What would be more interesting is to gather more information about a user rather than just the item he or she is viewing at that moment, to sketch a profile of this person. The internet offers techniques to gather a lot of data about users. Processing this data into useful information is something that needs to be done with care. Different techniques are discussed to aid this process of profiling users from raw data. The focus lies on techniques that will help create a profile of a user from the data available rather than techniques for improving search results and presented information. The techniques described will be used to create a base from which tailored information can be derived.

First a broad introduction to the domain of persuasion and personalization is given to give the reader an understanding of the context. This is followed by a brief discussion on the concerns involving personalization and user profiling. Several techniques for user profiling are then discussed including their positive and negative aspects. The paper is concluded by discussing these aspects and a comparison of the earlier mentioned techniques.

Online Persuasion

Today, the internet is becoming increasingly important for commercial organizations. Just having a physical store is not enough anymore. As a company you need to have at least a web site, preferably with a web shop, to survive and stand a chance against your competitors. But how can you convince (potential) customers online to buy your products? In a physical store, you can do several things to achieve this. For example, you can present your products in an attractive way, or depending on the type of product, you can let your customers feel, taste, smell or try out your product. Your salesman can give a customer advice based on his or her personal situation or specific preferences, etc. But how do you take this process of *persuading* to the web?

Techniques

Before we proceed, let us first take a step back and talk about what is meant by *persuasion*. Persuasion is an attempt to 'change people's attitudes or behaviors' (Fogg 2003). In the case of online stores it would mean changing a visitors' attitude or behavior to stimulate sales of the products. Convincing visitors, who should be seen as potential customers, to buy a product can be aided by different concepts such as reciprocation, commitment and consistency, social proof, liking, authority and scarcity (Cialdini 2006). By creating desire using these concepts, buyers are more likely to buy certain products. Such concepts however contribute to a sales strategy rather than a method of persuasion. Fogg (2003: 31-54) describes a number of tools or techniques to aid persuasion:

- **Reduction**: The less work that has to be done to achieve something, the lower the threshold of doing it.
- **Tunneling**: Using persuasion while guiding the user through a one way process.
- **Tailoring**: Persuasion through customization according to individual needs, interests, personality etc.
- **Suggestion**: Suggesting a certain behavior in the right place at the right time.
- **Self-monitoring**: Achieve a predetermined goal by monitoring yourself.
- **Surveillance**: People tend to co-operate when they have a sense that their actions are being watched or followed.
- **Conditioning**: Using rewards, or positive reinforcement, to persuade.

Although all afore mentioned techniques are valuable for storeowners to increase sales, not all techniques are directly applicable to the online sales environment. Giving a person the sense of being watched will probably motivate him or her to comply to certain rules, but it will not convince someone into buying (more) products. Self-monitoring is more related to persuading users to fulfill tasks they find themselves less motivated to do. This is hardly related to sales and websites.

While the principle of attempting to persuade customers on the internet is similar to offline persuasion in many ways, studies show that the current conversion rates are significantly different: 20 - 25% on average for offline stores compared to 1 - 3.5% on average for online stores (Jansen 2012). Needless to say, there is a lot of room for improvement here. As previously illustrated in this chapter there are several techniques that vendors can exploit to increase their sales online. One of the most promising techniques is *tailoring* (personalization), which will be the subject through the rest of this paper.

Personalization

As previously explained, personalization technology is a persuasion technique that ‘provides relevant information to individuals to change their attitudes or behavior or both’ (Fogg 2003: 37). Applied to the web, it includes ’any action that tailors the web experience to a particular user, or a set of users’ (Mobasher et al. 2000). Personalization can take place in two ways: through making recommendations to the user and through reorganizing the web site to the user’s needs (Koutri et al. 2002).

Recommendation

When a user is browsing, rather than searching, he or she is open for suggestions. This is where recommendation comes in. Iskold (2007) suggests the following categories of approaches to recommendation:

Item recommendation*: recommend things based on the thing itself*
This approach makes use of content-based filtering. Items with a similar content are shown. One common way to do this is through the use of tags. Another way is through machine learning techniques, such as artificial neural networks.

Social recommendation*: recommend things based on the past behavior of similar users*
This approach makes use of collaborative filtering. Items that other people liked, that also liked the current item, are shown. The idea behind this is that the user gets recommendations from people with similar taste. This is usually done through rating systems. Combining and weighting the preferences of the different users forms a key problem.

Personalized recommendation*: recommend things based on the individual's past behavior*
This approach combines content-based filtering and collaborative filtering. The system can give recommendations based on items the user previously viewed. Related items and items other users also liked are shown. This goes further than just item and social recommendation: the recommended items are not only based on one current viewing item, but based on a whole history of viewed (or purchased) items.

Reorganization

There are several sites that offer the option to *manually customize* their site according to the preferences of the user. For example, the user can change the background color, redesign layout or choose content modules (Koutri et al. 2002). Although this is a form of personalization, it provides a minimal degree of automation and thus will not be elaborated on.

An adaptive website on the other hand, is a web site that automatically improves its organization and presentation by learning from user access patterns (Perkowitz and Etzioni, 1997). The reorganization of a web site can vary from as little as highlighting a link to as big as completely restructuring the layout of the web site. Adaptive websites are part of the next generation of the internet, which makes use of *web intelligence* (the combination of artificial intelligence and information technology on the web).

You would probably recognize the familiar ‘*it’s got to be here somewhere...*’, searching through all menus and submenus of a web site to find that piece of information you are looking for. Especially when a web site is information rich, it can be a challenge to present the right information to the right user at the right time. What makes this so difficult is that different users have different goals, and the same user can have different needs at different times. This is a typical A.I. challenge.

Showing the right information at the right time to a particular user, can be viewed as *path prediction* (i.e. predicting the path a user is going to take on the website, and thus be a step ahead). In order to create a system that can do this, several questions should be answered (Perkowitz and Etzioni, 1997):

- **What are we predicting?** Are we trying to predict the user's next step or the user's eventual goal?
- **On what basis do we make predictions?** On basis of one particular user or a generalization of multiple users?
- **What kinds of modifications do we make on basis of our predictions?** Highlighting links or change the layout?

It is also possible for new users, of whom no information is gathered yet, to benefit from such a system. When information of individual users is accumulated, generalizations could be made on what mainstream users would prefer. This can then be applied to new visitors, which makes the web site in general easier to use.

Concerns

Although there are a lot of very interesting aspects to personalization and despite the fact that personalization can create a better experience for many users of the online world, personalization technology should only be applied with caution. There are several concerns related to personalization that should be kept in mind when implementing such techniques. As the focus of this study is exploring the possibilities of personalization rather than the constraints or implications that it brings these concerns will not be elaborated on extensively. These concerns are not unimportant, but will only be discussed briefly to shed light on their existence.

Ethical concerns

Not all users may appreciate that personal information such as their likings or behaviors are stored and used, especially not for commercial purposes. Some may feel 'manipulated' into buying something because with this information companies know exactly what buttons to push. It may be perceived as a huge privacy invasion, which could result in a lawsuit if a company is not careful (Treiblmaier et al., 2004). Because this information is very sensitive, it should be stored securely. Also, a company should take into account that they may lose the customers that value information transparency, because they are less willing to be profiled (Awad and Krishnan, 2006).

Filter Bubble

Another concern is that 'personalized filters limit what we are exposed to and therefore affect the way we think and learn' (Pariser 2011). This filter bubble exposes us merely to those things we already know, like or are interested in and prevents us from encountering other viewpoints than of our own. This could have a great influence on our personal development, especially when personalization technology is applied by more and more web sites.

Law constraints

The fact that certain technologies are available for general use this does not mean that they are lawfully allowed by all jurisdictions. For example, as of the 9th of May 2012 so-called tracking cookies, small files that are stored on a user's computer that make it possible to follow the user's surf behavior, are not allowed in the Netherlands without explicit permission from the user (Zeldin 2009). Although there is an opt-in possibility for website operators to use these tracking cookies you obviously do not want to disturb your users with permission notifications. It will also be more difficult to apply personalization techniques for users that decide to opt-out. Due to the fact that the afore mentioned law restrictions are a result of the new European Telecom Law, it is uncertain how other European countries will enforce these laws. It is far from certain whether other countries will allow such an opt-in solution.

The above merely being an example of restrictive legislation, one does not need to argue that there could be more (privacy related) laws that restrict website operators in effectively applying personalization and persuasive technologies.

User Profiling

A *user profile* is a description of an individual user and contains the most important or interesting facts about him or her. User profiles are built because each user differs in preferences, interests, backgrounds and goals. These differences form the basis of personalization (Schiaffino and Amandi, 2009).

Gathering Information

There are several types of information that can be stored in a user profile in the context of e-commerce (Schiaffino and Amandi, 2009):

Personal information is the most obvious type of information to gather. This includes age, gender, city, country etc. When creating an account at a web shop it is not unusual that the user is asked to fill in this type of information.
Interests of a user are a key part of personalization. Interests can represent hobbies-related topics, work-related topics, news topics and more. These topics can be derived from various sources, such as purchasing history or browse history.
Behavior is a type of information that is gathered implicitly. It is important that the behavior is repetitive in order to detect patterns. For example, ‘when purchasing product X, user Y usually also buys product Z’.
Goals of a user are important to detect when you want to optimize the convenience for the user by showing the wanted results sooner than usual. This type of information is crucial to learn from for goal prediction.

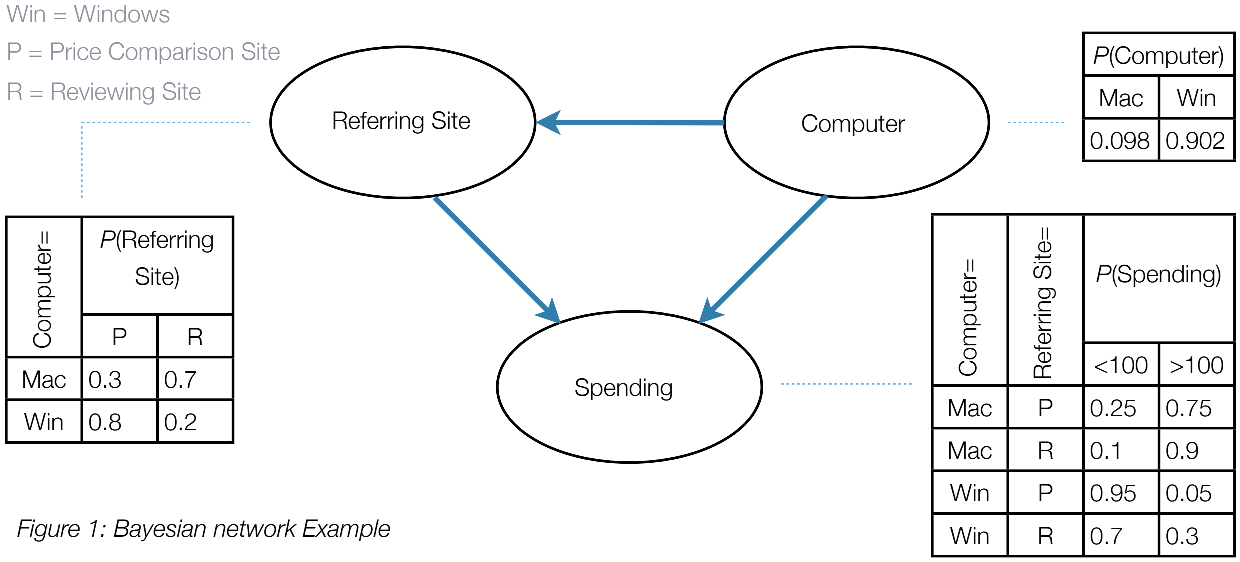
Data can be gathered *implicitly* as well as *explicitly*. Explicit data refers to information that is knowingly provided by the user, for example through a survey or through ranking items. Implicit data is gathered through surfing behavior, geographic location etc. which is tracked through logging or recording actions. Patterns can be discovered using techniques such as machine learning or data mining, making this gathered information useful.

Intelligent Techniques

The difficult part of user profiling is how to get useful information from raw data. Here we will discuss some intelligent techniques for automatically creating user profiles coming from areas such as machine learning, data mining or information retrieval, using the gathered information discussed in the previous section (Schiaffino and Amandi, 2009).

Bayesian Networks

A Bayesian network is a data structure that represents knowledge in an uncertain domain. It is a directed acyclic graph in which the nodes represent random variables with continuous or discrete values. When there is an arrow from node X to node Y, node X is said to be a parent of node Y. Each node X_i has a conditional probability distribution
$\mathbf{P} (X_i \mid \text{Parents} (X_i))$ (Russell and Norvig, 2010).



P(Computer)	
Mac	Win
0.098	0.902

Computer=	*P*(Referring Site)	
	P	R
Mac	0.3	0.7
Win	0.8	0.2

Computer=	Referring Site=	*P*(Spending)	
		<100	>100
Mac	P	0.25	0.75
Mac	R	0.1	0.9
Win	P	0.95	0.05
Win	R	0.7	0.3

Figure 1: Bayesian network Example

Let us take a look at an example for a better understanding of this concept in the context of this research, see also figure 1. According to *The Wall Street Journal* (Mattioli 2012) Mac users spend 30% more a night on hotels than Windows users. An average nightly hotel booking is around $100. Additionally, Mac users make up 9.8% of the U.S. personal computer market. This means that on average, Mac users spend more than $100 and Windows users less than $100.

Suppose there is another variable, Referring Site, which also has an influence on the chances that the user will spend more or less than $100. Referring Site is the type of web site the user is coming from when visiting the hotel booking site. In this example this can be either an a price comparison web site or a reviewing web site. Suppose that the type of computer that the user is using has an influence on the chances whether the user visits a price comparison site or a reviewing site. Note that this would not be an unrealistic assumption: given that Mac users are 40% more likely to book a four- or five-star hotel than PC users (Mattioli 2012), one could imagine that Mac users value quality over quantity. For the sake of simplification of this example we will ignore the fact that there might be other types of computer users than Mac or Windows, that there might be other types of referring sites than a price comparison or reviewing site and that the Spending amount could be exactly $100.

Figure 1 illustrates a Bayesian network of this information. The type of computer the user is using is logged when the user visits the web site. From this type of computer, assumptions can be made. For example, when statistics say that Mac users are usually big spenders, then one can assume that when the user uses Mac, it is more likely that this person is a big spender. According to this information, the website should show the user the more luxurious products. Of course one can imagine that such assumptions could be inaccurate. When more information is gathered, such as the referring site the user comes from, the assumptions that are made can be fine-tuned. Eventually, when a lot of information is used, these assumptions have a greater chance of being correct. This is how a user is profiled using Bayesian networks.

Naïve Bayes

An approach that is commonly used for user profiling is Naïve Bayes (Schiaffino and Amandi, 2009: 211). Naïve Bayes is actually the simplest form of a Bayesian network, in which all random variables have the same parent, and are independent given the value of this parent (Zhang 2004). This approach is called naïve because it is rarely the case that these variables are actually (conditionally) independent. Although it is therefore a simplification of the real world, it can work surprisingly well.

Pros & Cons

Bayesian networks are a powerful tool for classifying users. They allow the system to put users in different categories based on some key features that are characteristic for these different users. Statistics from many sources can be used to set up a meaningful Bayesian network. Previously an example was given about Mac users being offered more expensive hotels. The knowledge used by the website in this example came from several sources. Bayesian networks allow for the use of self-gathered data but also data from third parties.

Bayesian networks do however also have a weakness. If non-reliable third party information is used, or if there is a lack of relevant information, it is very difficult to make a Bayesian network meaningful often leading to less relevant classifications. Also domain specific knowledge is needed to gather meaningful variables that will actually classify users into meaningful classes.

Creating a Bayesian network is something that needs to be done manually, due to the necessity of domain specific knowledge. Especially when variables are dynamic and can not be computed from a database or dynamic dataset a Bayesian network could become outdated quite quickly, possibly and probably leading to less accurate results in the course of time.

A great strength of Bayesian networks is that when a suitable network has been devised and applied, users can often be classified based on just a few variables that could be known from the first arrival of a new user. The previously explained Bayesian network used information that is easily gathered. It is not difficult to track where users came from as this information is often passed on by referring websites and respectively what operating system they are using as this information is sent to the server together with the website page request to a server. Because users are classified based on generalization, it could be advisable to put less weight on information computed from Bayesian networks when other techniques have gathered far more detailed information. Bayesian networks can be rough in their classification.

Decision Trees

A decision tree is a tool that can support decision making such as how to classify a new user. This can be done based on decision trees that are automatically learned from data of existing users. Table 1 shows a small data set of existing users with data of whether they did or did not respond to the marketing campaign of company X. From this, the a decision tree can be derived, see figure 2.

Sex	Country	Responded
male	domestic	yes
female	international	no
female	domestic	yes
male	international	yes
female	international	no

Table 1: An example of a small set of data

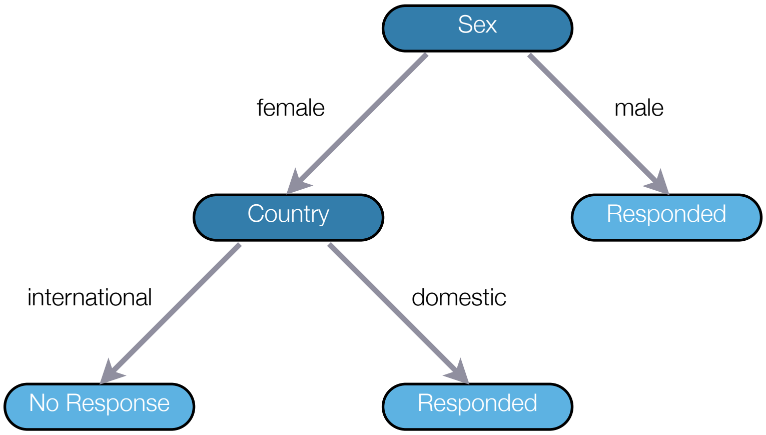


Figure 2: An example of a decision tree

Based on a decision tree like this, a new customer can be classified. For example, when the new customer is a male, it is likely that the new customer will respond to a similar marketing campaign. Of course this example is very simplistic, but imagine a decision tree of a very large data set. Then, it would not be far fetched to classify a customer in such a way.

Pros & Cons

When provided with the right algorithm, a tree can automatically be derived from a dataset. Using this tree a user can be easily classified by comparing the profile variables to those in the tree. A great aspect of decision trees is that they, unlike Bayesian networks, can contain a diverse set of variables that are in no way related. A side-effect of this aspect is that when information about a variable is missing, the user will be classified by the majority of that variable. This is also known as a greedy algorithm, meaning that the best path down the tree will not necessarily be found due to a wrong turn based on a lack of information about a certain variable.

A factor that needs to be taken into account is that it is possible to divide up the tree to such small variables that the classification based on the end of the tree become useless. This is also referred to as overfitting. There are solutions to this problem such as pruning. Selecting a threshold for pruning can be difficult as you might be too cautious resulting in maintaining the overfit or you could be too ruthless by chopping away valuable pieces of data.

Case-Based Reasoning

Kolodner defines Case-Based Reasoning as ‘a technique that solves new problems by remembering previous similar experiences’ (Schiaffino and Amandi, 2009). It does not solely rely on general knowledge of a problem domain, whereas many other techniques, such as Bayesian networks, do (Aamodt and Plaza, 1994).

Case-Based Reasoning (CBR) is used extensively in day-to-day common-sense reasoning (Kolodner 1992). It is also used by various professions. For example, when a doctor has a patient with a certain combination of symptoms, he might remember another patient in the past that had the same kind of symptoms, and propose the same diagnosis. This type of reasoning can be applied for building user profiles: when a new customer has a certain combination of interests, the CBR could look up what products customers with similar interests bought and propose these to the new customer.

Figure 3 on the next page illustrates how Case-Based Reasoning works (Aamodt and Plaza, 1994).

When new information is available about a certain customer, this can be seen as a problem: how can we optimize the personalization for this customer, given the old as well as the new information? This forms a *New Case*. The first step in the CBR cycle is then to RETRIEVE a similar previous case, which would be another existing customer with similar information. Then, the *Retrieved Case* and the *New Case* are combined through REUSE into a *Solved Case* that suggests a solution to the problem, such as which products to show to the customer. The next step in the cycle is to apply this solution and to test its success through REVISE. The solution may be adjusted and repaired when the solution was proved not to be successful. This brings us to a *Tested and Repaired Case* and this gives a confirmed solution. The last step in the cycle is to RETAIN the case and save it in the database of previous cases as a *Learned Case.*

General Knowledge of the problem domain can be used throughout the cycle for support. For example, when a doctor proposes a diagnosis to a patient, this proposal is not only based on previous experience, but also on what he has learned at medical school. In the case of user profiling general knowledge can be any kind of research relevant for personalization.

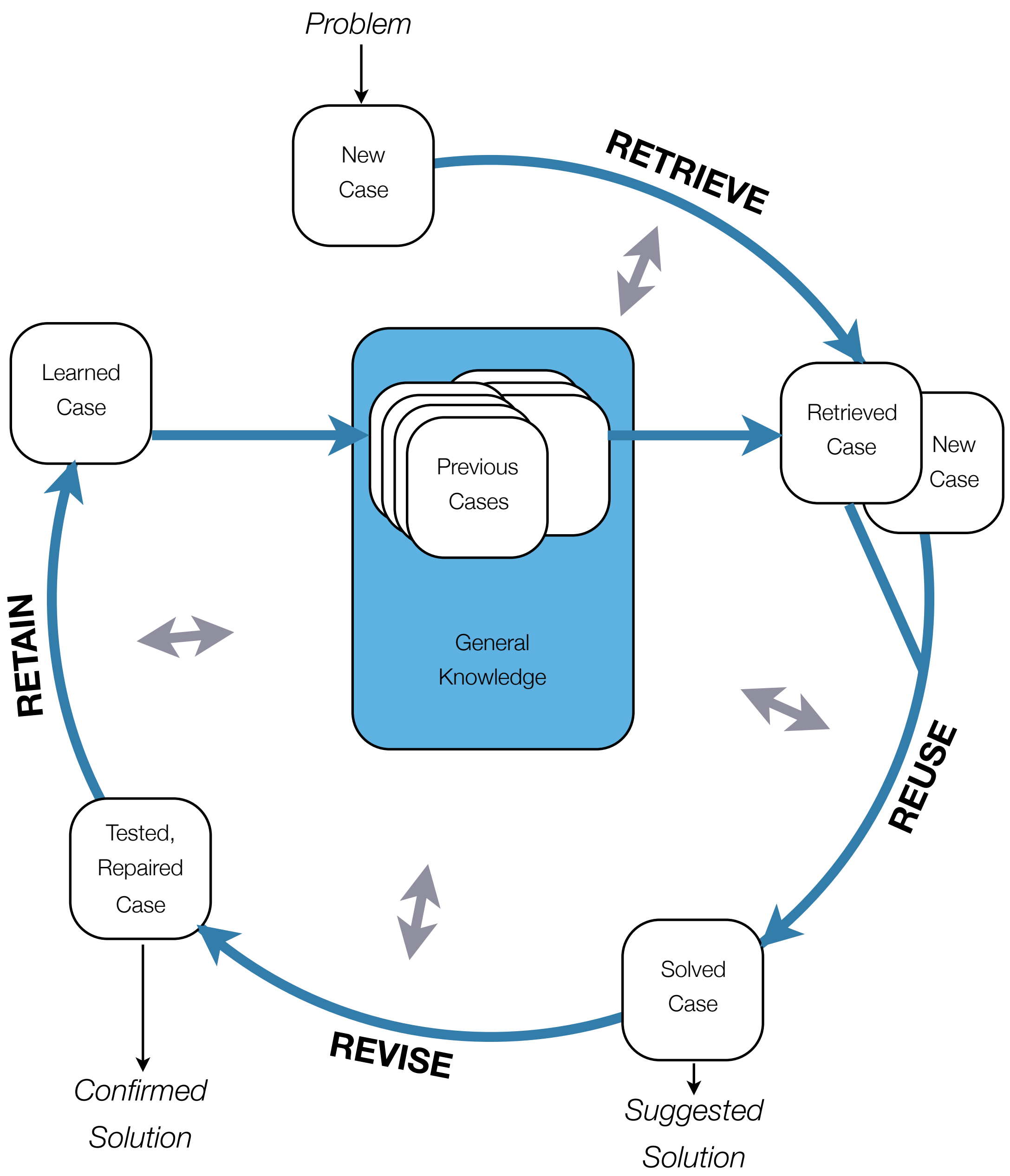


Figure 3: The Case-Based Reasoning Cycle

Pros & Cons

Bayesian networks and decision trees are great tools when it comes to quickly placing users in a certain category to present them with certain website content. One of their weaknesses is that they use generalization to approach a new user. When applying case based reasoning, the system looks at each case separately, compares it to previous cases and applies a best fit solution. It then evaluates its decisions, updating the knowledge base of cases. Due to the incremental learning the system is directly able to apply newly learned cases to new cases. Also solutions will adjust according to changing variables, keeping cases and solutions up-to-date.

The difficulty that has troubled researchers is that there is no standard for storing and retrieving cases in a proper way (Aamodt and Plaza, 1994). It is very difficult to create a model for storing and retrieving such information. Also these models can differ a lot between different domains.

Association Rules

Association rule learning is a data mining method for discovering interesting relations in large data sets (Tan, Steinbach and Kumar, 2005: 327). This large data set may contain all transactions that are made by all customers. Table 2 illustrates an example of these *market basket transactions.*

Transaction ID	Items
1	{ muesli, yoghurt }
2	{ bread, butter, cheese, milk }
3	{ bread, muesli, milk }
4	{ cheese, muesli, yoghurt }
5	{ bread, butter, muesli, yoghurt}

Table 2: An example of market basket transactions.

From this data set, the following association rule can be extracted: `{muesli}` → `{yoghurt}` with a *confidence factor* of 75% (Agrawal, Imielinkski and Swami, 1993). This means that in 75% of the transactions, when muesli is purchased, yoghurt is also purchased. An association rule consists of a *antecedent*, here `{muesli}`, and a *consequent,* `{yoghurt}`. An antecedent or consequent is a set, in this case a set of products. In our example the set contains only one product, but it is also possible for it to contain more than one product.

Interesting relations may be found by asking queries such as ‘find all rules that have *eggs* as a consequent’ or ‘find all rules that have *coffee* as an antecedent’. From the results of these queries, one could derive for example what to do to boost the sale of a certain product or what product sales would be affected when a certain product would be discontinued (Agrawal, Imielinkski and Swami, 1993).

For user profiling, this technique can be applied in various ways. Instead of looking at single transactions, one could look at transactions over a period of time. For example, when a customer buys new ink cartridges for his printer every month, the ink cartridges can be shown on the home page of the web site around a month later of his last purchase to remind him of his need for new ones and make it easy for him to repurchase this item.

This technique can also be applied for user profiling in a dynamic way. For example, when the customer is still shopping and adding items to his shopping cart, the system may notice that there are certain products that are frequently bought together with the items already in the cart and suggest this to the customer. It can even look at a history of purchases of the customer in combination with the products already in the shopping cart to make more accurate suggestions.

Pros & Cons

Association rules are quite simple as their task is to connect recurring pairs of sets in the data for example pairs of products that often recur in the same shopping basket or certain products that often occur in the same time span. One of the great advantages of association rules is that in the case of recurring pairs of products, the system is far more likely to suggest complementary products rather than similar products (e.g. ink cartridges when looking at a printer rather than a different brand of printers).

Although association rules are difficult to apply to a new website due to a lack of historic data, it is possible to apply such rules to information that was gathered far before these rules are implemented. Also the general data can be applied to new users directly. When a new user is browsing laptops the system might suggest anti-virus software if this is a trend observed in previous transactions of earlier customers. When the system is able to gather information about a single user over a longer period of time, the concept of association rules can be applied with much more detail and a higher level of personalization.

A disadvantage of this method is that the system may find association rules between products that are only related by coincidence. For example, when the rule `{toilet paper} → {coke}` is found, theoretically one could explain this by saying ‘when people go to the toilet, they get thirsty and drink coke’, but this would not make much sense. It would be more logical to say that although the system found this rule, there is no actual correlation between these products. It is difficult to filter out these rules that seem to be ‘nonsense’.

Neural Networks

An artificial neural network is a machine learning technique that is inspired by the biological brain. Roughly speaking, the brain gets input, processes it, and gives output. A biological neural network consists of neurons that communicate with each other by firing small electrical signals. In an artificial neural network these are represented by nodes and connections between these nodes, see figure 4 on the next page. Each connection has a certain weight. A neural network learns to improve its results by adjusting these weights, based on what the network thinks the outcome should be and on what the outcome actually is.

A neural network is particularly good in recognizing patterns. For example, it can learn to distinguish between the letters A and B, by putting more weight on the horizontal bottom row pixels and the vertical left column pixels, and less weight on the middle horizontal row pixels. These are the pixels where the letters A and B are more or less distinguishable.

For user profiling neural networks can be used to classify the user into stereotypes by using the gathered information of the user as input (Chen, Norcio and Wang, 2000). When the characteristics of each stereotype are known, it is a matter of putting the pieces of information of a particular user together and look at what stereotype matches best. In this case, the pieces of information are ‘to be recognized’, the stereotypes are the ‘patterns’ and the matching is the ‘recognizing’, thus pattern recognition. By classifying users into stereotypes, certain assumptions are made. For example, suppose there is a user of whom it is known that he has an expensive car and lives in an expensive neighborhood and that this person is classified in the stereotype ‘rich’. From this, one may infer that this person may also like to play golf, since this is also part of that stereotype. Of course, these assumptions are not always accurate.

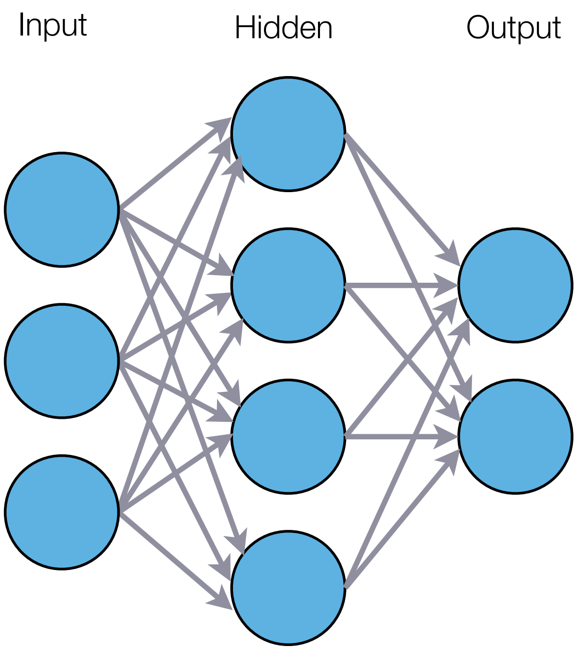


Figure 4: An example of an artificial neural network

Pros & Cons

Neural networks are quite useful due to the fact that they can guess missing information in a user profile with quite a high level of detail. By applying stereotypes to a user, assumptions about a user will be made until the contrary has been proven. In this way the system can start adjusting the profile it has of a user from the moment he or she enters the website.

The best way of implementing this technique is when the website is able to identify the user with some certainty. This technique relies on the ability of building a profile over a longer period of time. There is the risk of users that share accounts or a single internet connection (and thus a single IP), for example within a corporate environment, leading to mixed, inaccurate results.

Another point of concern is that a neural network relies on feedback. By receiving feedback a neural network is able to adjust a node's weight. The system however only receives positive feedback when a user spends a lot of time on a product's page or when he decides to buy a certain product and negative feedback when a user repeatedly ignores a suggestion or spends very little time on a product's page. It is highly imaginable that this information is far from adequate to produce a reliable neural network. Especially since the neural network gives no insight into how it has reached its results over a period of time.

Conclusion

In the previous chapter we have seen different techniques for analyzing user data and producing information about a user from this data. Some techniques rely a lot on data collected from previous users of a website and others rely more on the collection of information about a certain user. Eventually the common goal of using these techniques is creating a way to present information to a user that will lead to users buying more products, buying more expensive products or become loyal customers.

Comparison

In order to compare these techniques a model has been devised in which different aspects of the different techniques are compared, see table 3.

	Bayesian Networks	Decision Trees	Case-Based Reasoning	Association Rules	Neural Networks
User Classification	✓	✓	✓	✓	
User Profile Building					✓
Based on Learning			✓		✓
Based on Statistics	✓	✓		✓	
Learn from Single Users	-	-		-	✓
Learn from All Users	-	-	✓	-	

Table 3: A comparison of the properties of the different user profiling techniques

User Classification & User Profile Building

In this section of the model we look at the nature of a technique. Most techniques are well suited for classifying a user into a certain group. An example would be a Bayesian network classifying Mac users into 'high spenders'. Other techniques are focused more on building a user profile around a certain user. In this study neural networks is the only technique that actually builds a profile around a certain user. By making educated guesses a neural network tries to learn as much about a certain user as possible improving the profile as it goes along. A reason why building a profile around a certain user is difficult is that it can be difficult to confirm a website user's identity. Therefore profile building techniques are less suitable for website environments.

Based on Learning & Based on Statistics

In the model we make a distinction between techniques that learn and techniques that simply apply a statistical calculation to the presented data. A learning technique is a technique that uses past experience and applies it to new data. Although a system can create and revise a decision tree from the data it is not considered learning. The system does not learn to make a better tree by choices it made in the past. A neural network however, depends on previous decisions. By adjusting weights within the network it fine-tunes itself, eventually leading to an accurate profile of a user.

Association rules are clearly based on statistics. By counting the times that certain products are bought together in the same transaction or in regular time cycles suggestions can be made towards a user. Decision trees and Bayesian networks are techniques that have been devised to calculate the statistical chance that a certain user is in a certain class.

Learn from Single Users & Learn from All Users

In this study only two techniques are based on learning. Case-based reasoning however, relies on information learned from the mass. The system learns from previous cases from all users to create a databank of cases that it can refer to in new cases. These can generically be applied when the system sees fit. A neural network is based on a single user and is filled with information about this single user only. Although the system relies on stereotypes defined by other users, the network does not define these stereotypes itself. These are more likely to have been extracted from data by creating decision trees or similar techniques.

Bayesian networks, decision trees & association rules

What might strike you is that in the previously explained model there are three techniques with the exact same characteristics. Bayesian networks, decision trees and association rules are all classifying statistic techniques. The difference between these techniques lies in the way that they are implemented in the system and the type of results that are meant to be produced. Bayesian networks are quite suitable for classifying users based on just a few variables. New visitors to a website can easily be classified using a Bayesian network. Decision trees are better at classifying users based on larger datasets with more and more diverse variables. Also decision trees can make stereotypes visible, allowing the technique to be suitable as a complement to neural networks. Association rules do utilize classification but to a different extent. Association rules are useful for finding products that the user is likely to respond to based on the products he is buying or viewing. Although these techniques are similar in the model, their use is quite different. There might be some overlap between Bayesian networks and decision trees, but the suitability of these techniques is defined by the problem they are meant to solve.

Applicability to new and existing systems

The above mentioned techniques might or might not be usable straight away. There are two factors that define the amount of time needed until a technique is usable. These are the learning curve of an algorithm and the amount of available data. All techniques need a certain amount of collected data. When the operator of a web shop first launches, he will have little or no data to apply. An existing shop might have a great amount of historical data. If relevant data has been collected and stored in a suitable manner the techniques that are strictly statistical can be applied directly. Otherwise the operator needs to wait until he has acquired a minimum amount of information. The other factor, the learning curve, is applicable to learning algorithms. A neural network will only say something useful about its user after it has been adapting to him for a longer period of time. With case-based reasoning a collection of cases need to be built up to have a suitable existing case when new cases occur.

Most suitable technique

It is not possible to point at a single technique as being the best or most efficient or suitable. As described earlier in this chapter and as made visible to a great extent in the model described above each technique focuses on a different aspect of user profiling. The technique applied to a system depends a lot on the problems that need to be solved. It could even be possible to create a website that involves each of the techniques mentioned in this article.

Imagine a web shop for media, such as CDs and DVDs that uses Bayesian networks to classify users whether they are more likely to like music or movies. It also uses decision trees to find trends in the genres that the user enjoys. Case-based reasoning allows for the website to choose the most effective way of leading the user through the check-out encouraging extra purchases on the way out. Association rules might suggest soundtrack CDs to go with a DVD film and a neural network might keep track of the user's profile in order to create a collection of items to present to a user in the weekly newsletter.

Further Reading & Research

In this study several techniques have been discussed. All these techniques are useful for converting raw user data into usable information about the user. There are however a few algorithms that are very useful for selecting information most suited for a certain user. The two methods listed below are interesting to investigate when further implementing user profiles and personalization techniques as they can be seen as an extension on the previously discussed methods. As they are suitable for enhancing search results, they did not fit in the scope of this study.

K-Nearest Neighbor algorithm

The K-Nearest Neighbor, or K-NN, algorithm is a method well suited for generating personalized query results. Search results are personalized by comparing the current user's profile to other user profiles and selecting the most similar one. Gemmell et al. (2009) describe how they use K-NN to suggest tags for a music piece a user wants to classify, based on the profile of the user. By looking at tags that similar users used, tags that are likely to be used can be suggested.

Rocchio-based method

The Rocchio-based method is used for enhancing search results based on the content of a text. Degemmis et al. (2003) used a Rocchio-based method to classify texts. By defining whether the text gets a positive classification or not based on a user's profile, products that are suitable for the user can be displayed more prominently or at the top of the search results.

Furthermore there was a mention of the fact that there is no uniform way of modeling cases when applying case-based reasoning. This is a problem that still needs to be addressed in order to apply this technique successfully. Further investigation in this subject will hopefully allow for operators of e-commerce websites to apply case-based reasoning effectively.

References

Aamodt, A. and Plaza, E. (1994) *Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches.* Available from URL: http://www.iiia.csic.es/People/enric/AICom.html [Accessed 9 July 2012]

Agrawal, R., Imielinkski, T. and Swami, A. (1993) *Mining Association Rules between Sets of Items in Large Databases.* Available from URL: http://www.almaden.ibm.com/cs/projects/iis/hdb/Publications/papers/sigmod93.pdf [Accessed 10 July 2012]

Awad, N.F. and Krishnan, M.S. (2006) *The Personalization Privacy Paradox: An Empirical Evaluation of Information Transparency and the Willingness to be Profiled Online for Personalization.* Available from URL: http://www.jstor.org/stable/pdfplus/25148715.pdf?acceptTC=true [Accessed 8 June 2012]

Chen, Q., Norcio, A. F. and Wang, J. (2000) *Neural Network Based Stereotyping for User Profiles.* Available from URL: http://userpages.umbc.edu/~norcio/papers/2000/Chen-NNSterUProfiles-NCA.pdf [Accessed 11 July 2012]

Cialdini, R.B. (2006) *Influence: The Psychology of Persuasion.* New York: HarperCollins Publishers. ISBN: 978-0061241895

Degemmis, M., Lops, P., Ferilli, S., Di Mauro, N., Basile, T.M.A. and Semeraro, G. (2003) *A Relevance Feedback Method for Discovering User Profiles from Text.* Available from URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.60.4140&rep=rep1&type=pdf [Accessed 24 July 2012]

Fogg, B.J. (2003) *Persuasive Technology: Using Computers to Change What We Think and Do.* San Francisco: Morgan Kaufmann Publishers. ISBN: 978-1558606432

Gemmell, J., Schimoler, T., Ramezani, M. and Mobasher, B. (2009) *Adapting K-Nearest Neighbor for Tag Recommendation in Folksonomies.* Available from URL: http://www.dcs.warwick.ac.uk/~ssanand/itwp09/papers/gemmell.pdf [Accessed 24 July 2012]

Iskold, A. (2007) *The Art, Science and Business of Recommendation Engines.* Available from URL: http://www.readwriteweb.com/archives/recommendation_engines.php [Accessed June 1 2012]

Jansen, G. (2012) *Online Persuasion.* Available from URL: http://prezi.com/ef1pzizucd92/online-persuasion/ [Accessed 21 May 2012]

Kolodner, J. (1994) *An introduction to Case-Based Reasoning.* Available from URL: http://web.media.mit.edu/~jorkin/generals/papers/Kolodner_case_based_reasoning.pdf [Accessed 9 July 2012)

Koutri, M., Daskalaki, S. and Avouris, N. (2002) *Adaptive Interaction with Web Sites: an Overview of Methods and Techniques.* Available from URL: http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=C7A38D2557A0E7E1DE05CF6F09E2D370?doi=10.1.1.13.8052&rep=rep1&type=pdf [Accessed 30 May 2012]

Mattioli, D. (2012) *On Orbitz, Mac Users Steered to Pricier Hotels*. The Wall Street Journal. Available from URL: http://online.wsj.com/article/SB10001424052702304458604577488822667325882.html [Accessed 26 June 2012]

Mobasher, B., Dai, H., Luo, T., Sun, Y. and Zhu, J. (2000) *Integrating Web Usage and Content Mining for More Effective Personalization*. Available from URL: http://maya.cs.depaul.edu/~classes/cs589/papers/ecweb2000.pdf [Accessed 31 May 2012]

Pariser, E. (2011) *The Filter Bubble: What the Internet Is Hiding from You.* New York: The Penguin Press. ISBN: 978-1594203008

Perkowitz, M. and Etzioni, O. (1997) *Adaptive Web Sites: An AI Challenge.* Available from URL: http://www.cs.washington.edu/homes/etzioni/papers/ijcai97.pdf [Accessed 1 June 2012]

Russell, S. and Norvig, P. (2010) *Artificial Intelligence: A Modern Approach.* Upper Saddle River: Pearson Education, Inc. ISBN: 978-0136042594

Schiaffino, S. and Amandi, A. (2009) *Intelligent User Profiling.* Available from URL: http://www.exa.unicen.edu.ar/catedras/knowmanage/apuntes/56400193.pdf [Accessed 25 June 2012]

Tan, P., Steinbach, M. and Kumar, V. (2005) *Introduction to Data Mining.* Boston: Addison Wesley. ISBN: 978-0321321367

Treiblmaier, H., Madlberger, M., Knotzer, N. and Pollach, I. (2004) *Evaluating Personalization and Customization from an Ethical Point of View: An Empirical Study*. Available from URL: https://springerlink3.metapress.com/content/h116661574707141/resource-secured/?target=fulltext.pdf&sid=f0gtf5bbpfjitjqrli30awjd&sh=www.springerlink.com [Accessed 8 June 2012]

Zeldin, W. (2012) *Netherlands: Amended Telecommunications Act Prescribes Net Neutrality, Stricter Cookie Provisions.* Available from URL: http://www.loc.gov/lawweb/servlet/lloc_news?disp3_l205403143_text [Accessed 25 June 2012]

Zhang, H. (2004) *The Optimality of Naive Bayes.* Available from URL: http://www.cs.unb.ca/profs/hzhang/publications/FLAIRS04ZhangH.pdf [Accessed 17 July 2012]