Mobile Contextual Advertising

Klaas Kox
HAIT Master Thesis series nr. 08-005

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ARTS IN COMMUNICATION AND INFORMATION SCIENCES,
MASTER TRACK HUMAN ASPECTS OF INFORMATION TECHNOLOGY,
AT THE FACULTY OF HUMANITIES
OF TILBURG UNIVERSITY

Thesis committee:

Prof. Dr. A.P.J. van den Bosch
Drs. A.M. Bogers

Tilburg University
Faculty of Humanities
Department of Communication and Information Sciences
Tilburg, The Netherlands
December 2008

Index

Abstract

Mobile marketing strategies commonly use SMS messages. While one has to pay for each sent message, these messages should be matched to the right receiver in order to have an optimal result and low costs. In this research we investigate the possibilities of using machine learning techniques to optimize the matching of advertisements to receivers within an SMS-marketing system. We compare four machine learning algorithms on their classification task. We also experiment with the use of profile and content features and a combination of both and find that the use of both feature sets lead to a better classification than the use of only one set.

1 Introduction

1.1 Mobile Marketing

Marketing finds its way into all forms of communication. From the moment that e-mail was introduced to the masses, companies started using the service for marketing purposes. The Short Message Service (SMS) allows mobile phone users to send short text messages between mobile phones. Although this service of course was not invented to serve the needs of advertisers, the emergence of a mobile commerce market is a logical effect. Fogg (2003) mentions the *Principle of Convenience*, which highlights an advantage of mobile commerce: "Interactive experiences that are easy to access (ideally, just a click away on a mobile device) have greater opportunity to persuade" (p. 189).

Despite the fact that e-mail and SMS share a lot of characteristics, some important differences should be noted. First, as the name reveals, SMS messages are short; they can only contain a maximum of 160 characters. Second, SMS messages cannot contain any mark-up, in contrast to e-mail messages. The last and maybe most important difference is the fact that users have to pay for sending SMS messages, while e-mails can be sent for free. Targeting e-mail advertisements does not immediately implicate extra costs. Spam mail, or unsolicited bulk e-mail, has become the omnipresent result. In contrast, in order to create a beneficial SMS marketing campaign, one has to make sure that only relevant receivers are reached, so no costs are wasted.

1.2 Recommender systems

Targeting advertisement can be seen as a process of matching receivers with advertisements. Here, a comparison can be made with the field of recommender systems (Resnick & Varian, 1997; Oard, 1998; Ben Schafer et al. 1999). These systems, often web applications, recommend items to users according to some matching function. These functions may be based on user characteristics, or the user's browsing and searching behaviour, and may involve the comparison of the user's characteristics compared to those of other users. Recommendation techniques are commonly used in commercial settings (Sarwar et al., 2000; Linden et al., 2003). In such commercial settings, these recommendations are typically phrased as "Customers who bought this item also bought …". To compute these matching functions, big databases with user profiles need to be kept.

In this study we investigate the possibilities of using recommendation techniques in order to select appropriate advertisements in a mobile marketing model. Our idea is to base this recommendation on the matching between a new receiver and existing receivers, while taking into account profile information, the content of the advertisement, or a combination of both. We will investigate the possibilities of using different machine learning techniques to predict whether advertisements will be relevant to a receiver or not. Our expectations are that the use of both user profile data and advertisement content information will lead to a better classification, as we expect the combination of profile and content data to be more valuable than profile and content data separately.

1.3 Research Question

In this research we investigate the efficacy of using machine learning techniques to optimize the matching of advertisements to receivers within an SMS-marketing system. We compare the use of profile and content features and a combination of both as the base of the machine learning classification. We focus on the question which of the two feature sets (profile and content) in isolation yields the best classification results and whether the combination of profile and content features outperform the use of only one of the two features. During our machine learning experiment we will test four machine learning algorithms in terms of their generalization performance, in order to determine which performs best in the classification task.

Chapter 2 introduces the field of mobile advertising: what are the possibilities of using mobile phones for marketing purposes. Chapter 3 provides a brief overview of machine learning and the opportunities it offers for mobile advertising models. After that, Chapters 4 and 5 provide details on our experimental setup and our results. In the last chapter we will discuss our results in order to answer our research questions and recommend points of future research.

2 Mobile Advertising

Mobile phones have become an important new class of media. In the modern society large parts of the population can be reached directly by mobile phone. In this chapter we will describe the features of this new medium and the marketing possibilities it offers.

2.1 Mobile Marketing

There has been a tremendous growth in mobile phone usage in the last 15 years. While in 1995 there was an average of three mobile phone connections per 100 people in European countries, now there is an average of more than 100 connections (Statistics Netherlands, cbs.nl[1]). This does not immediately implicate that almost everyone has a mobile phone nowadays, but it does offer an ideal instrument to use for marketing purposes. Because of the mobility aspect of the device, people can be reached almost any time of the day. This makes the mobile phone arguably a good medium for marketing purposes. However, not everyone will agree with that. Bauer et al. (2005) found that the user's acceptance of mobile marketing depends to a high degree on the user's attitute towards the information value and entertainment value of mobile communication.

[1] http://www.cbs.nl/nl-NL/menu/themas/bedrijven/publicaties/digitale-economie/artikelen/2007-2214-wm.htm

The Short Message Service, a particularly popular form of mobile communication, enables mobile telephone users to send and receive short text messages with their telephone. It is widely used by most of the mobile phone users. Around 80% of all European mobile phone users use the service. A key characteristic of the SMS technology is the limited length of the messages. Each message can only contain 160 (7-bit) characters.

2.2 Spam

Spam is the collective term for unsolicited electronic messages. E-mail spam is the most widely recognised form of spam, but there are many other forms of spam, such as instant messaging spam, search engine spam, and wiki spam[2]. With the tremendous growth in mobile phone usage, mobile phone spam or m-spam has also arisen (Yamakami, 2003; Camponovo, 2004). M-spam messages are small textual ads that are superficially similar to normal SMS messages, but are unsolicited. Most mobile phones are not capable of spam filtering. When using the mobile phone platform for marketing purposes, there is a big risk of being perceived as spam, even if the receiver did sign up for receiving these messages. As people in general are not willing to read messages any further as soon as they recognize them as spam, marketers need to find a way to prevent ads from being experienced as such. An important concept in this process is personalisation, which we will describe in the next section.

2.3 Personalisation

Personalisation generates significant potential for mobile marketing (Bauer et al., 2005). There are two different approaches to personalisation. The first is to mention the receiver's name in the message in order to show that the message is directed to the receiver. The other approach is closely related to the recommendation techniques we mentioned before, and tries to select ads that are likely to be relevant to the receiver. In order to do this a database with user profiles is needed. The idea here is that ads that are more relevant to a receiver are less likely to be experienced as being spam.

---

[2] http://en.wikipedia.org/wiki/Spam_(electronic)

Although personalisation can improve the acceptance of advertisements, many messages that only consist of an advertisement will be deleted right away. A possible way to overcome this problem is to add advertisements to normal messages that are relevant to the receiver and that the receiver will want to receive and read, for example, by adding small ads at the end of a regular message. To do this, one has to set up a service in such a way that its users will accept the fact that their messages are accompanied with ads. For instance, one might think of a free web e-mail service that adds small commercial messages at the bottom of every e-mail, or a free web-based SMS service that reserves a number of characters at the bottom of every text message for commercial utterances. In exchange for such a service the user accepts the presence of ads. Because of the precedence of information that is interesting to the user, the commercial texts are more likely to be read; at least they will not be deleted.

Due to the fact that there are many free e-mail services, there are not very many possibilities for such advertising models in the SMS market. There has to be some kind of (often financial) advantage for users to make them accept the presence of ads. The market of the Short Message Service (SMS) may be appropriate for this, because a user pays for all messages he or she sends by default; free SMS sending is a much-wanted commodity. Since the early 2000's several web-based commercial initiatives have appeared that offer a free SMS sending service, in exchange for taking about 40 characters of every message for commercial utterances. Some examples of these services are hotSMS.com and text4free.net.

2.4 SMS Marketing

As mentioned above, text messaging enables advertisers to communicate directly to a specific target group. A big advantage of text messaging is the fact that, just like e-mail, receivers do not necessarily have to actively acknowledge the receipt of a message, unlike for instance voice communication via the telephone. However, there are some drawbacks of SMS one has to be aware of.

First, the limited length can cause problems when messages get more complex. Messages that contain more characters than the maximum length due to the complexity have to be split up in more messages. Another disadvantage of text messaging can be that it is not possible to add any kind of layout. SMS messages can only contain plain text. Unlike e-mail, no font colours, font sizes, images, tables, etc. can be added.

These limitations however can easily be viewed as a benefit (Haig, 2002). "The fact that SMS messages are so limited in format means that marketing promotions look similar to texts sent from a mobile user's friend. Consequently, the divide between commercial and personal messages is narrowed, and so, providing the messages are of value and are permission-based, they are usually well received."

Negative image

Bauer et al. (2005) mention the 'risk'-factor of mobile marketing, which is caused by people's negative experiences with mobile services. For example, companies that sell wallpapers or ringtones for mobile phones tempt mobile phone users to subscribe to their services via television or Internet campaigns. Consumer's associations warn people for the high costs of these services. In many cases users are unable to unsubscribe or receive extremely high telephone bills. This negative image of these mobile phone services extends easily to seemingly similar services. This tends to result in the situation that although a service is totally free to the user, people are reluctant in accepting the service. Informing users about the way the service operates and the potential costs is crucial to possible success of the organisation.

3 Machine Learning

Mitchell defines Machine Learning as follows: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E".

The goal of this research is to explore the possibilities of using machine learning to determine whether an advertisement will be relevant to a receiver or not.

3.1 Intelligent Machines

Most computer programs are designed to facilitate tasks performed by humans. In general, computer programs are not intelligent; they can only perform simple calculations and they only do what the programmer told them to do, in contrast to humans, who are able to learn with experience. Most software programs do not learn with experience. It does not matter whether a program is in use for only one day or more than ten years; its functioning does not change.

Practice shows, however, that certain software could be more user friendly when some kind of learning capability would be added. For example, when a user keeps using the same font in its text processor, the program might decide to store it in its default font list, just like a human probably would do. Although this "intelligence" is evident to users, designing such a system for computers is a complex issue.

Machine learning deals with the question how to construct computer programs that automatically improve with experience (Mitchell, 1997). The idea is that computer programs should have a certain *intelligence* that is fed by the input it gets from its user. Future use of the program will be influenced by this intelligence. A simple example is the machine learning application that is used in modern elevators. These elevators have the capability of learning on which floor they should wait for people in order to work as efficient as possible. The elevator remembers at which floor most people get in at what time of the day. It tries to find certain patterns in this elevator use and adapts its routing to that. The more experience such an elevator has, the more efficient it will work.

3.2 Knowledge Based or Machine Learning?

As described above we want to perform a classification task. There are two main approaches to do this: the knowledge-based approach and the machine learning approach. The knowledge-based approach applies previously programmed knowledge (usually taking the form of rules) to new cases in order to classify them. In the present case, an example rule may be to always send an advertisement related to fashion to all female users. The problem with such approaches is that before this method can be applied, one has to define rules by hand (e.g., if-then constructions) that determine the class of new cases. This introduces a risk of being influenced by lack of knowledge, or even prejudice, as one might not know the users of the system as good as one thinks. The alternative approach is to employ machine learning. Machine learning techniques do not need predefined rules in order to classify new items, but derive rules themselves by analyzing a set of already classified test data. Machine learning offers some advantages in comparison to a knowledge-based approach. First, machine learning proper does not suffer from any prejudices, as it has a "blank" intelligence before learning. Second, a system based on machine learning can easily be developed for more than a single domain, without having to define new rules for each new domain. This means that, on condition that a set of already classified data is provided, it is not necessary to perfectly know the users of a system. In essence, machine learning does not assume the presence of knowledge beforehand, but as a trade-off, does demand that a sufficient amount of examples of classified data is available. Of course, knowledge-based and machine learning approaches can be combined as well.

3.3 Matching Advertisements with their Context

Advertising on the Internet offers many possibilities for advertisers to target their ads. Advertisements can be targeted to specific people, for example via e-mail, but another option is to target an advertisement to a specific category of websites. Lots of websites offer space to advertisers for placing commercial ads. The basic idea is that an ad is selected to be displayed based on the context of the website.

Broder et al. (2007) propose as system for contextual ad matching based on a combination of semantic and syntactic features. They classified 105 webpages and 2680 unique ads into a 6000 nodes commercial advertising taxonomy to determine their topical distance. They found that the effectiveness of their approach depends on the quality of the webpages used.

Anagnostopoulos et al. (2007) mention that a problem arises when advertisements have to be matched to dynamic pages. While matching advertisements to static pages can be based on prior analysis of their entire content, dynamic pages need to be analysed on the fly because of their dynamic content. An analysis of full webpages each time they are loaded would take too much time. Therefore, Anagnostopoulos et al. propose a method that uses text summarization to classify pages. They found that when a 5% fraction of the page is carefully selected through summarization, the relevance of automatically assigned ads drops only 1%-3%. This text-summarisation approach allows matching ads with pages in real time, without a prior analysis of the page content.

3.4 Machine-learning-based Recommendation

Recommendation is an increasingly used method of presenting information that is likely interesting for a user, mostly for commercial purposes. A recommender system learns from a user and recommends items to the user (Ben Schafer et al., 1999). This recommendation can be done by matching a user's profile to previously gathered profiles of other users and their characteristics. When looking for recommendations for a specific user, items that are related to matching profiles can be recommended.

These recommendation techniques can only function properly when a large database of user profiles, items and the relation between those two is available. In order to collect these user profiles explicit and implicit data collection forms can be used. An explicit form is to ask a user directly to rate, rank or choose items, and store these features in a database. Implicit forms of collecting these profiles monitor the user's behaviour (purchases, click paths) in order to deduct their preferences. These methods have the advantage that the user is not bothered with providing information for the data collection. However, because there is no quality control, the system can deduce strange and unwanted profile-item relations.

In this research we investigate the possibilities of using recommendation techniques, based on machine learning, in order to create a system that selects appropriate advertisements for users of a mobile marketing platform.

4 Experimental setup

In this chapter we discuss the pre-test and the main experiment we performed in order to answer our research questions.

The goal of this research is to investigate the possibilities of using machine learning in order to define rules that can be used to improve systems that match advertisements to recipients, content-based and profile-based.

4.1.1 iThumb

iThumb was a starting company at Tilburg University that provided a mobile advertising application. The iThumb platform enables organisations to easily send SMS messages to (groups of) people for free. In exchange for this service, small advertisements (40 characters) are added to each message. The idea is that the advertisers pay for all costs of sending out text messages (Figure 4.1).

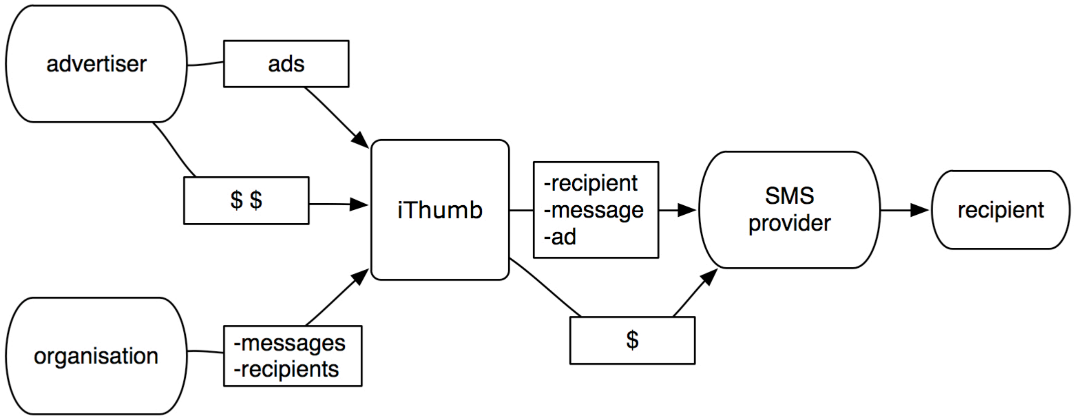

Figure 4.1 - The iThumb Platform

This research is based on the iThumb system, and explores possibilities of improvement.

4.2 Pilot experiment

In order to prepare our main experiment, we ran a pretest. We asked the communication department of the Humanities faculty at Tilburg University to participate in this experiment. They agreed with the idea of supporting their standard communications to the students (via e-mail) by sending out short SMS messages. Twenty students were asked to fill in a short questionnaire to gather information about their background and interests. In the following weeks they received four SMS messages sent by the Humanities Faculty. Each message consisted of around 120 characters of body text and 40 characters of a short advertisement. After that they filled in a form about whether the messages and the accompanying advertisements were relevant to them. This pre-test was mainly used to test the matching procedures and data-collection methods. Based on our experiences we prepared our main experiment, which we will describe in the next section.

4.3 Main experiment

The main experiment was held during the introduction week at Tilburg University. The introduction's organisation used the iThumb system to inform the students about the activities during this week. Around 75% (1269) of all participating students subscribed to this service. During one of the activities that week we invited students to participate in our experiment. We asked them to fill out a short questionnaire to gather information about their age, gender, place of residence. The interests where categorised into 4 groups:

Sports
- Soccer
- Tennis
- Hockey
- Fitness
- Dancing
- Korfball

Hobby
- Computers
- Music
- Cars
- Shopping
- Fashion
- Cooking

Horeca
- Cinema
- Restaurants
- Pubs
- Concerts
- Student's union
- Disco

Other
- Internet
- Newspapers
- Magazines
- Travelling
- Books
- Investing

All students received one message to confirm their subscription and eight subsequent messages that where sent by the introduction week organisation. These eight messages where all accompanied with a short advertisement that was randomly assigned to each message by the iThumb system.

4.4 Data collection

After the introduction week was finished and all messages had been sent, the participants were asked by e-mail to fill out an online questionnaire about how they experienced the messages they received (see Appendix). The form displayed all received messages (body-text) and the accompanying advertisements. Participants could select their degree of interest in the body-text and in which degree the advertisement was relevant to them, in 7-point Likert scales. Because of the fact that only a limited set of branches of advertisers participated in the introduction week, we also asked for the participants' opinion about three dummy advertisements. Finally, we asked whether they experienced the added advertisements as annoying or not.

The advertisements were categorised manually into 2 classes. First we classified the form of each ad:

Form of the message

Branding ad	The ad contains the name of the advertising company and is aimed to increase the brand awareness
SMS-Reply	The ad triggers people to respond to the message in order to receive

	more detailed information in a following message
Informative	The ad purely informs people about a certain topic, event or company
Invitation	The ad triggers people to undertake action (for example to subscribe to a newsletter

Secondly, we determined in which branch the ad belonged. We determined six different branches: Internet Services, Mobile Services, Employment Agency, Pubs / Disco, Concerts, Sport Shops.

4.5 Statistics

The collected data contained 17 unique advertisements and 57 profiles of people. 143 persons filled in their profile previously to the experiment, but only 57 of them judged the messages they received at the end of the experiment. They received an average of 8.7 messages each. An advertisement was added to every message. This resulted in 497 unique judged profile-advertisement combinations.

We converted the 7-point Likert scale judgements that described the relevance of the advertisements into discrete values (0: not relevant, 1: relevant). Because of the goal of our experiment, to predict whether an ad would be relevant to someone, only positive judgements (>4) were classified as relevant (1), while the other judgements (<=4) were classified as irrelevant (0). We assume that a neutral judgement (4) does not indicate any relevance and therefore should be classified as irrelevant in our experiment. This resulted in 111 of all 497 cases being classified as being relevant, so 77,7% could be predicted correctly by classifying all cases as being irrelevant.

4.6 Measures

In this section we describe the measures we used to evaluate the machine learning experiments we performed.

4.6.1 Precision and Recall

Our goal implies that, rather than accuracy, we optimize the precision and recall of our systems. Precision and recall are the measurements of choice to describe the quality of the retrieved results of our binary recommendation classification task

Recall is the ratio between the number of correctly classified relevant matches (true positives) and all cases of relevant matches. A high recall system is needed when retrieving all relevant items is more important than leaving out the items that are not relevant (false positives). For example, a lawyer looking for jurisprudence does not want to miss any of the true positives, and will accept it when a system retrieves some irrelevant items. In our experiment the true positives are the advertisements that were matched correctly to a receiver, and the false positives are the falsely matched advertisements.

Recall does not take false positives into account. When a system is used to identify relevant receivers of commercial messages, false positives (people falsely classified as relevant receivers) may well experience the received message as being spam, because the message is not relevant to them. Obviously, the number of false positives is an important measure for our system's performance. Precision takes false positives into account; it is the ratio between the number of items classified positively by the system and the total number of positively classified items (true and false positives). When all positively classified items are relevant (true positives), the precision will be at its maximum (precision = 1). In general one should carefully determine whether the main goal of the system demands a high recall or a high precision, or a trade-off.

A commonly used measure that combines precision and recall is the F-score, which expresses the harmonic mean of precision and recall.

In our experiment the main goal is to select a relevant advertisement for each recipient. The setting of SMS marketing makes that it is rather important to have a high precision classifier in which the quality of the retrieved items is high, i.e. preferably all of the retrieved advertisements should be relevant to the recipient.

4.6.2 ROC Curve

When dealing with skewed class distribution, a commonly used evaluation method is to create Receiver Operating Characteristics (ROC) graphs in order to visualize a system's performance. In order to do this, first, given a classifier and a test set, a two-by-two confusion matrix can be constructed (Figure 4.2).

<table>
<tr><th></th><th></th><th colspan="2">True class</th></tr>
<tr><th></th><th></th><th>Positive</th><th>Negative</th></tr>
<tr><td rowspan="2">Hypothesized class</td><td>Positive</td><td>True Positives</td><td>False Positives</td></tr>
<tr><td>Negative</td><td>False Negatives</td><td>True Negatives</td></tr>
</table>

Figure 4.2 - Confusion Matrix – Fawcett, 2004

The True Positive Rate (TPR) is estimated as the number of true positives divided by all positive instances (i.e. the part of all instances with the positive class that was classified correctly), while the False Positive Rate (FPR) is the number of false positives divided by the total number of negative instances. By making a scatter plot with both the TPR and FPR values, as measured by different parameter settings, the system's settings for optimal performance can be determined. The calculation of the area under the ROC curve (AUC) is commonly used as a summary statistic about a system's performance (Figure 4.3). When designing a system, one should aim for a low FPR and a high TPR.

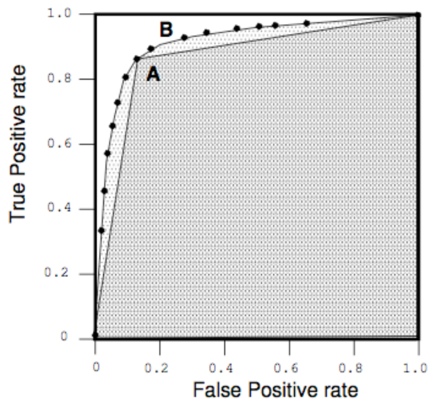

Figure 4.3 - Area under the curve (AUC) -
Fawcett, 2004

4.7 Machine learning algorithms

In order to evaluate the data, we used the Weka software from the University of Waikato[3]. This software includes a collection of machine learning algorithms for data mining tasks. We experimented with a range of algorithms and our classification task in which the algorithm needs to predict whether an advertisement would be relevant to a specific person represented by a profile. In every experiment, we used a smoothing parameter to evaluate the algorithm ranging from unsmoothed to extreme smoothing settings. In the next sections we will briefly describe the algorithms we used (Mitchell, 1997).

4.7.1 Rule Induction

A rule induction algorithm defines rules that represent the observations, given a set of observations. The algorithm prefers small rules over long rules, and tries to

[3] http://www.cs.waikato.ac.nz/ml/weka/

construct as few rules as possible. In order to smoothen the results we systematically tuned the parameter that defines the minimal number of cases a rule is allowed to cover.

4.7.2 Decision Tree

Decision tree learning is a classic machine learning method that tries to represent the classification function by a decision tree. Each node in the tree representation corresponds to an attribute, and its leaves represent the possible values of that attribute. Leaves are labelled with the most likely classification at that point in the tree. Tree representations can be rewritten to if-then rules with the use of conjunctions. The algorithm prefers small trees over large trees. The smoothing parameter we systematically varied is similar to the parameter we tune with the rule induction algorithm, namely the minimal number of cases a leaf is allowed to cover.

4.7.3 K-nearest neighbor algorithm

The k-nearest neighbor algorithm is an instance-based (lazy) learning method. These methods, in contrast to the rule induction and decision tree algorithms, do not construct a model each time training examples are provided (Mitchell, 1997). The k-nearest neighbor algorithm assumes all instances correspond to points in a multi-dimensional space in which each dimension represents one of the task attributes. Every time a new instance has to be classified, the algorithm chooses the most common class of the new instance's *k* nearest training examples in the multi-dimensional space. To smoothen the results we systematically tuned the parameter k.

4.7.4 Support Vector Machines

Like k-nearest neighbor classifiers, support vector machine learning methods represent training cases as points in a multi-dimensional space (Mangasarian, 2000). In order to be able to classify new given observations, the algorithm tries to separate the points in the n-dimensional space by constructing a separating hyperplane that maximizes the margin between itself and the two separated sets of data points. By altering the parameter that defined the severity of the separation (from disallowing any incorrectly separated point to allowing a fair amount of badly separated points) we smoothened the model.

4.8 Feature Selection

Feature selection is a technique that can be used to choose an optimal subset of predictive features by eliminating other features with less predictive information. We used WEKA to test whether feature selection could improve the results we obtained from the machine learning experiments that were based on all features. As the working of the rule induction and decision tree algorithms already has implements an internal feature selection, we applied the feature selection only to the k-nearest neighbor and the support vector machine models.

4.9 Data coding

All profile and content information was stored in a database and later converted to an arff-file, which is readable in WEKA. The profile features were age (numeric), gender (male, female) and the features (0: not interesting – 1: interesting) as described in Section 4.3. In order to code the content information we generated a bag of words with all unique words of all advertisements. After that each advertisement was coded for each word with a '0' if the advertisement did not contain the word and a '1' if it did. We also stored the form and the branch of the advertisement (Section 4.4).

5 Results

5.1 Comparing JRIP, J48, IBk and SMO

We tested four machine learning algorithms on their classification task performance. When comparing the four machine learning algorithms to each other (Figure 5.1), the support vector machine algorithm (SMO) clearly performs better overall than the other three algorithms (JRIP, J48, IBK).

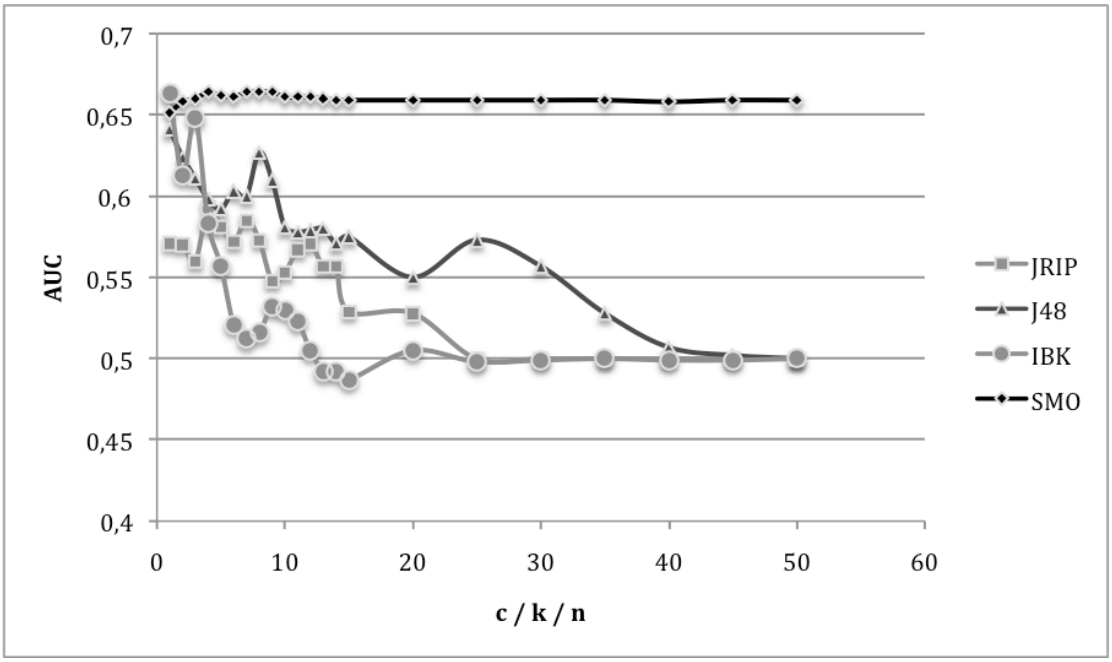

Figure 5.1 - Comparing four algorithms (profile + content)

We performed a 10-fold cross validation experiment with the four algorithms while letting their respective smoothing parameters range from 0 (no smoothing) and 50. When analyzing the peak performances of each algorithm, we observe that only the rule induction approach (JRIP) performs significantly worse ($p < .05$) than the other three approaches. There are no significant differences between the peak performances of the decision tree, k-nearest neighbor and the support vector machine approach.

As Figure 5.1 shows, it is remarkable that the support vector machine approach keeps a stable AUC-score when the coverage increases, while the AUC-scores of the other three algorithms clearly decreases when the smoothing factor increases.

In our rule induction, decision tree and support vector machine approaches we see an overall improvement of the TPR, FPR, F and AUC scores when content information is added to the profile information.

5.2 Feature Selection

We tested whether the use of feature selection would improve the classification task. Therefore we compared the results of both the k-nearest neighbor algorithm and the support vector machine algorithm with feature selection to the algorithms without characteristic selection. The support vector machine experiment however, did not benefit from feature selection at all.

k	TPR		FPR		F		AUC	
	without FS	with FS	without FS	with FS	without FS	with FS	without FS	with FS
1	0,135	**0,279**	0,041	**0,093**	0,211	**0,348**	0,547	**0,593**
2	0,135	**0,243**	0,041	**0,091**	0,211	**0,312**	0,547	**0,576**
3	0,135	**0,234**	0,039	**0,096**	0,213	**0,299**	0,548	**0,569**
4	0,135	**0,225**	0,034	**0,083**	0,216	**0,298**	0,551	**0,571**
5	0,135	**0,216**	0,031	**0,083**	0,217	**0,287**	0,552	**0,567**
6	0,126	**0,216**	0,026	**0,078**	0,207	**0,291**	0,550	**0,569**
7	0,117	**0,252**	0,026	**0,080**	0,194	**0,329**	0,546	**0,586**
8	0,099	**0,252**	0,021	**0,070**	0,169	**0,337**	0,539	**0,591**
9	0,117	**0,216**	0,034	**0,070**	0,190	**0,296**	0,542	**0,573**
10	0,054	**0,198**	0,016	**0,070**	0,098	**0,275**	0,519	**0,564**

Table 5.1 - K-Nearest Neighbor approach & Feature Selection

We increased k from 1 to 100 in our experiment, but the TPR, FPR, F and AUC scores from k=1 to k=10 were all higher than the scores at k > 10. The addition of feature selection to the k-nearest neighbor approach results in clear improvements of all scores from k=1 to k=100.

c/k/n	Rule Induction		Decision Tree		k-Nearest-Neighbor		Support Vector Machine	
	profile	profile + content	profile	profile + content	profile	profile + content	profile	profile + content
1	**0,558**	0,571	0,547	**0,641**	0,616	**0,663**	0,496	**0,651**
2	**0,558**	0,570	0,547	**0,624**	**0,614**	0,613	0,496	**0,658**
3	0,545	**0,560**	0,548	**0,611**	0,614	**0,648**	0,503	**0,660**
4	0,566	**0,593**	0,551	**0,598**	**0,614**	0,583	0,499	**0,664**
5	0,569	**0,581**	0,552	**0,592**	**0,614**	0,557	0,499	**0,662**
6	0,560	**0,572**	0,550	**0,603**	**0,611**	0,521	0,502	**0,661**
7	0,572	**0,585**	0,546	**0,600**	**0,590**	0,512	0,501	**0,664**
8	0,573	0,573	0,539	**0,627**	**0,571**	0,516	0,501	**0,664**
9	**0,574**	0,548	0,542	**0,610**	**0,569**	0,532	0,501	**0,664**
10	**0,580**	0,553	0,519	**0,581**	**0,569**	0,530	0,501	**0,661**

Table 5.2 - Comparing AUC scores for profile and profile+content information

Table 5.2 shows the peak performances of the four algorithms, comparing the use of only profile features to the use of profile and content features. We also tested the use of only content features but that approach did not lead to any appropriate classification at al. All algorithms show that the use of both feature sets results in a better classification than when using only the profile set.

5.2.1 Important features

We analysed the features that were frequently used in the models created by the rule induction and decision tree algorithms. Figure 5.2 shows an example of a set of rules generated by our rule induction approach (JRIP) using both profile and content data.

```
(concerts = 1) and (computers = 1) and (the word "GRATIS" = 1) => relevant=1

(the word "IN" = 1) and (newspaper = 1) => relevant=1

(the word "WIN" = 1) and (computers = 1) and (soccer = 0) => relevant=1

 => relevant=0
```

Figure 5.2 - Rules generated by the JRIP algorithm, using both profile and content data

Figure 5.2 shows 3 conditions, all leading to a relevant classification. All cases that do not obey these conditions will be classified as being irrelevant. This illustrates the fact that the receivers classified more than 77% of the messages as being not relevant. Both profile and content information appear in the rules. This indicates that combining both feature sets is probably more effective than only using a single feature set. Figure 5.3 displays a decision tree that was generated by J48, using only profile data.

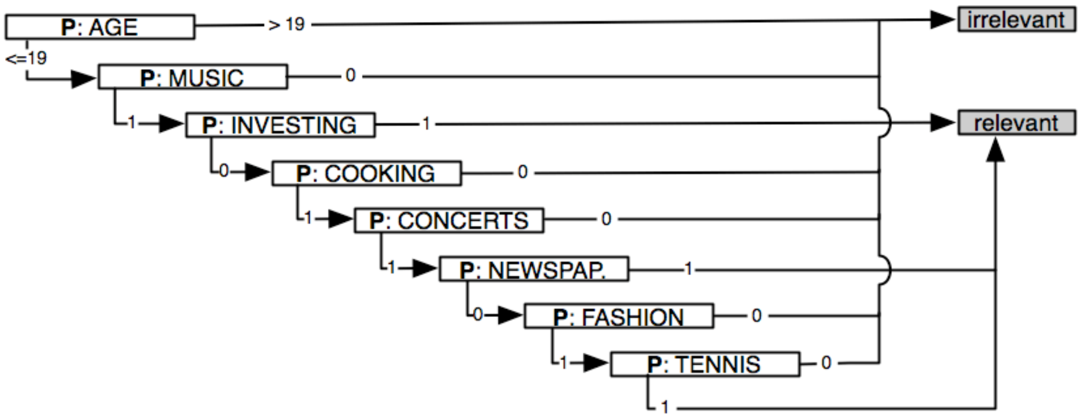

Figure 5.3 - Decision Tree generated by J48 on profile data.

On average, the paths leading to a relevant classification in this decision tree are longer than the paths leading to an irrelevant classification, which also illustrates the skew distribution in our data. This decision tree is quite clarifying, but when we add the content data the decision tree model (Figure 5.4), becomes a much more complex model. The fact that this large decision tree also uses a mix of both feature sets indicates that the combination of profile and content information is valuable.

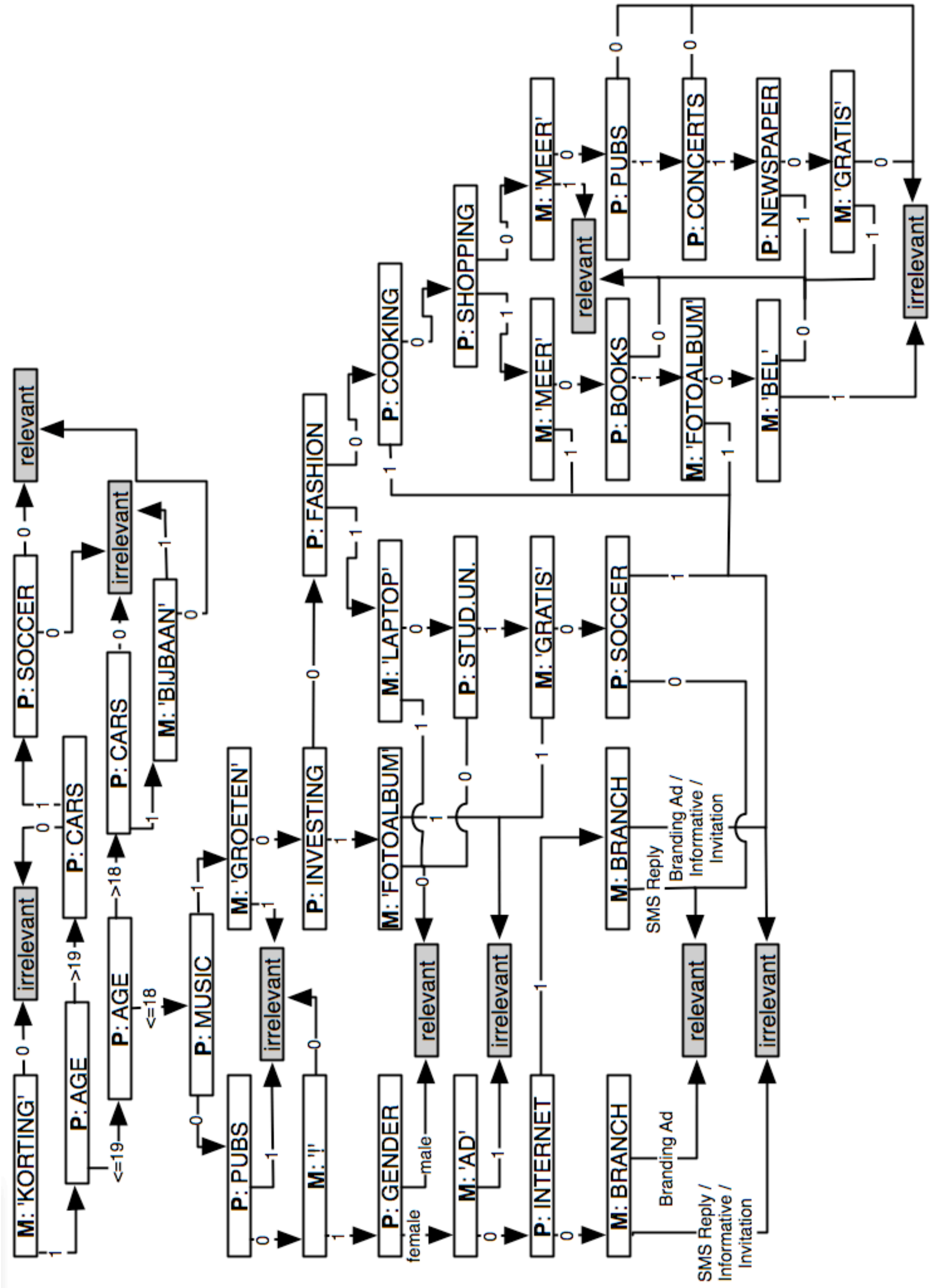

Figure 5.4 - Decision Tree generated by J48 on profile and content information. (P = profile information, M = message data)

To retrieve yet another perspective on the importance of features we looked at the feature selection approaches in which the algorithm selects a group of features with high predictive information itself. The algorithm selected the following seven features:

Profile	Age
	Music
	Concerts
	Disco
Content	Form of the message (Section 4.4)
	Use of the word 'KORTING' (discount)
	Use of the exclamation mark ('!')

In sum, a small set of features tends to be strong indicators of a relevant classification. For example, participants who indicate to be interested in computers and/or concerts tend to experience advertisements as being relevant more often. Also, advertisements that contained the words "WIN" and "KORTING" (discount) were also classified as being relevant in most cases. When combining profile and content features, one could conclude that advertisements containing words like "WIN" and "KORTING" might be more effective when they are targeted to people interested in music, as this is of course the binding factor between music, concerts and disco.

6 Discussion and conclusions

In this chapter we discuss the results we presented in the previous chapter. Based on the discussion of the results we answer the research questions we posed in the first chapter.

6.1 Data sparseness issues

As mentioned in Chapter 2, a major problem in SMS marketing by adding short commercial texts to messages is the limited length (160 characters) of those messages. Only a small part of those characters can be reserved for the advertisement, as there must be enough space left for the user's message. In the model iThumb uses, 40 characters are available for commercial texts. Although textual creativity and the use of SMS language can significantly decrease the number of characters needed, advertisers often need more space to communicate their message.

The machine learning experiments we performed also encountered some difficulty caused by the limited length of the messages, as this produces only sparsely available positive examples. The same experiment performed in another domain, for example e-mail messages (which have an unlimited length) instead of SMS messages, would probably produce less sparse data.

The data we collected contained 497 unique judged profile-advertisement combinations. While this is sufficient for machine learning experiments, extending this collection would probably lead to results that can be generalised more easily.

6.2 Importance of profile information

Our results clearly show that using the profile feature set leads to better results than using the content feature set; the usage of the content feature set does not lead to any appropriate classification at al. However, when joining the two feature sets, the results show a small improvement, so combining both profile and content information does lead to a better classification. The availability of more features makes it possible to build better models describing the test data. Better models will lead to a better classification.

Our results confirm the importance of profile information. Gathering this profile information is an expensive and time-consuming process. Therefore, in commercial models, one should carefully consider whether this approach will be profitable eventually. When the mobile marketing system is part of a larger network, forms of implicit profile data collection should be considered.

6.3 Recommendations

Our results do not show one algorithm performing significantly better than the other algorithms. However, the rule induction approach performed significantly worse than the other three approaches (decision tree, k-nearest neighbor and support vector machine). While these three do not significantly perform differently, we recommend using the more understandable decision tree algorithm. The support vector machine and k-nearest neighbor approach operate as a “black box”, not showing a clear model on which the classification tasks are based. This makes it hard to analyse the working of the system in order to get a better understanding of the way different features correlate with the receiver's acceptance of advertisements. When the decision tree is implemented in a specific system, it creates "understandable" rules on which it bases its classification. The system's developers can interpret these rules and can reimplement them in any programming language.

6.4 Further research

As mentioned before, it would be interesting to perform a follow-up study with a larger data collection. Extending this data collection can be reached in two ways: more participants and more advertisements. Especially a larger number of different advertisements would be interesting to use as this could lead to more detailed information about which words or textual forms correlate with the perceived relevance.

7 References

Anagnostopoulos, A., Broder, A., Gabrilovich, E., Kosifovski, V., Riedel, L. Just-in-Time Contextual Advertising. CIKM`07, November 6-8, Lisboa, 2007.

Bauer, H., Barnes, S., Reichardt, T. and Neumann. M. Driving Consumer Acceptance of Mobile Marketing: A Theoretical Framework and Empirical Study. Journal of Electronic Commerce Research, vol. 6, no. 3, 2005.

Ben Schafer, J., Konstan, J. and Riedl, J. Recommender Systems in E-Commerce. E-Commerce 99, Denver, Colorado, 1999.

Broder, A., Fontoura, M., Josifovski, V. and Riedel, L. A Semantic Approach to Contextual Advertising. SIGIR`07, July 23-27, Amsterdam, 2007.

Camponovo, G. and Cerutti, D. The Spam Issue in Mobile Business - A comparative regulatory overview, Proceedings of the Third International Conference on Mobile Business, 2004.

Fawcett, T. ROC Graphs: Notes and Practical Considerations for Researchers. Tech Report HPL-2003-4, HP Laboratories, 2004.

Fogg, B. J. Persuasive technology: Using computers to change what we think and do. Morgan Kaufmann Publishers, San Francisco, 2003.

Haig, M. Mobile Marketing. The Message Revolution. Kogan Page, London, 2002.

Linden, G., Smith, B. & York, J. Amazon.com Recommendations: Item-to-Item Collaborative Filtering. IEEE Internet Computing, vol. 7, p. 76-80, 2003.

Mangasarian, O. L. Generalized support vector machines. Advances in Large Margin Classifiers, p. 135–146, MIT Press, Cambridge, 2000.

Mitchell, T. Machine Learning. The McGraw-Hill Companies, 1997.

Oard, W. D. Implicit Feedback for Recommender Systems. Proceedings of the American Association for Artificial Intelligence Workshop of Collaborative Systems (AAAI98), p. 80-83, 1998.

Resnick, P. & Varian, H. Recommender Systems. Communications of the ACM, no. 3, vol. 40, p. 56-58, 1997.

Sarwar, B., Karypis, G., Konstan, J. & Riedl, J. Analysis of Recommendation Algorithms for E-Commerce. EC'00: Proceedings of the 2nd ACM Conference on Electronic Commerce, p.158-167, ACM, 2000.

Yamakami, T. Impact from Mobile SPAM Mail on Mobile Internet Services. Parallel and Distributed Processing and Applications, vol. 2745/2003, Springer, Berlin, 2003.

Appendix


UNIVERSITEIT VAN TILBURG
thumb
Do you ?
Onderzoek SMS notificatie
De Universiteit van Tilburg en iThumb onderzoeken samen de effectiviteit vam reclame uitingen in SMS berichten.
Vul deze enquete (1 min) in en maak kans op een van de waardebonnen van € 50,-
Telefoonnummer
06
E-mailadres
Aanmelden


UNIVERSITEIT VAN TILBURG
thumb
Do you ?
Onderzoek SMS notificatie
Voornaam
Klaas
Achternaam
Kox
Telefoonnummer
0641225253
E-mailadres
k.kox@uvt.nl
Woonplaats
Tilburg
Leeftijd
25
jaar
Geslacht
man
vrouw
Verder


UNIVERSITEIT VAN TILBURG
thumb
Do you ?
Onderzoek SMS notificatie
Selecteer uit elke categorie minstens 1 interesse die bij je past
Sport
voetbal
tennis
hockey
fitness
dansen
korfbal
Hobby
computers
muziek
auto
koken
winkelen
mode
Uitgaan
bioscoop
restaurant
kroeg
concerten
studentenvereniging
discotheek
Overig
internet
beleggen
kranten
tijdschriften
reizen
boeken
Verder


UNIVERSITEIT VAN TILBURG
thumb
Do you ?
Onderzoek SMS notificatie
Advertentie:
Bel/SMS gratis! Studentenbellengratis.nl
In welke mate vond je deze advertentie relevant?
1 2 3 4 5 6 7
relevant
niet relevant
Opslaan