Bachelor thesis CKI
Topic: Behavioural targeting
ECTS: 7.5

Ruud van Houwelingen (3347516)

words: 5250

Table of Contents

1. Introduction

1.1 Conventional advertising versus online advertising

As the world is buying smartphones and tablets at an exponential rate, it seems that we are heading towards a completely digital society. Online advertising is therefore taking a prominent role in the advertising industry. Also, online advertising has unique possibilities that are non reachable in traditional advertising methods like television or newspapers. One great advantage is that it has the capability of showing the right advertisement to the right user at the right time. This creates a far more personalized advertising experience than was ever possible in traditional advertising. Another great advantage of online advertising is that advertisers can see results right away. Compared to television ads, in which advertisers can only guess whether the ad was successful or not, they can see how many people clicked their ads and how many of these clicks resulted in people buying products. This automatically leads to another advantage which concerns costs. A lot of online advertisements are paid for by click, so that advertisers only have to pay for clicks that were actually efficient. Online advertising knows many forms, but in this thesis I will focus on behavioural targeting, which concerns gathering user statistics and showing relevant advertisement.

1.2 Objective of this thesis

In this thesis I will look at two very different approaches for behavioural targeting. The first one being one that focusses on single users. However, this approach has some points that require improvement. The second behavioural targeting approach is one that focusses on collective intelligence, and therefore uses the knowledge of a lot of users, and might solve the problems which the former has. The objective therefore is to look whether the approach that uses collective intelligence can overcome these shortcomings.

In order to reach this objective properly I will start by explaining the first approach of behavioural targeting. In the chapter after I will explain the idea of collective intelligence, followed by a chapter that implements this idea of collective intelligence in a different behavioural targeting approach. I also included a chapter concerning legal issues about this topic, that will make clear which aspects happening in the field of behavioural targeting are and which are not allowed, and how this will affect the future of behavioural targeting.

2. Traditional behavioural targeting

Behavioural targeting is a marketing method that enables advertisers to show the right advertisement to the right people at the right time, and therefore leads to increased effectiveness compared to traditional forms of advertising like television or newspaper adds. This is a major difference and very beneficial for advertisers, since they want to spend a small amount of money and still get a great amount of profit. To achieve this, more and more website owners and advertisers record the surfing behaviour of their visitors to figure out their visitors' interests and needs. I can further explain this by means of an abstract example. Imagine a company that sells smart phones. One way they can increase their sales is by placing advertisements on web shops that is specialized in electronical devices like smart phones, since visitors of that store might be interested in buying a new phone. Another way is place advertisements on websites that show information about the latest smart phones, because their visitors might also be interested in buying a new phone. What they could also do, is show their advertisement on the web shop's website only to those visitors that recently visited the website regarding information about the latest smart phones. Now this is the true power of behavioural targeting.

Behavioural targeting techniques can be categorized under three different versions. I will explain them briefly. The first version is on-site behavioural targeting. This method places a permanent cookie on the computer and represents a user profile. In this cookie the website stores clicks, searches, time spend on pages, etc., but only from within the website itself. Based on this data algorithms decide which content to show to the visitor and when, and the visitor's responses are again stored in the profile, that gets smarter by each visit.
The second version is called network behavioural targeting. This method is based not on one website, but on a group of different websites that are part of the same advertising network, which means that the placed cookie can be used by other websites or organisation that are also part of this network. This version gives a representation of the user that is way more accurate than the on-site method since it also contains other interests that might be completely irrelevant to one particular website within the network.
The third version is called ISP behavioural targeting. For this method internet service providers need to grand online advertising organisations permission to their network that shows how their clients use their internet connection. However, a trial period in the United States by a company called NebuAd and a large number of ISPs led to a lot of objections and is therefore no longer maintained. First of all the clients believed that all communication is supposed to be handled with care by their ISP, and should not be for sale. Secondly, there was no real opt-out method, customers were only able to opt-out of the receiving of customized ads, they were unable to opt-out of the actual gathering of behaviour. Thirdly, there was a lack of oversight since it was not clear what third-party companies were doing with the information. Besides that, NebuAd refused to name which ISPs were taking place in the trial, and the whole thing was in conflict with United States wire tap laws, which protects any kind of information communication against interception.

In the next paragraph I will discus recommendation systems, which most of the times can be categorized as on-site behavioural targeting, followed by its underlying techniques.

2.1 Recommendation systems

Recommendation systems are behavioural targeting systems that analyse (un)structured data like item descriptions so that those items which might be interesting to a specific user will be identified and presented to that user. Recommendation systems are mostly used for on-site behavioural targeting, for example by the online retailer Amazon. For a recommendation system to perform properly, the system needs good item representations which are stored in a database, and user profiles that match the user's interests. In the remaining part of this chapter and chapter four, I will use recommendation systems as an example for online advertising, and look at its underlying process. The next paragraph is about content-based recommendation systems, which is a process in which user profiles are matched with item profiles.

2.2 Content-based filtering

Content-based filtering is all about matching item profiles with user profiles, in which candidate items are being compared so that the best matching ones can be recommended. Item descriptions are sometimes unstructured and therefore need to be analysed to create a structured item representation. This process is called stemming, where a specific root form of terms is created, so that for example words like “bike”, “biking” or “bikers” will all count under the same term. Next, these terms are linked to a variable that represents the importance of that term, the value of this variable is called the term-frequency times inverse document frequency weight, or tf*idf. The corresponding formula is as follows:

$$w(t,d)=\left(tf_{t,d}\log\left(N/df_t\right)\right)\div\sqrt{\sum_i\left(tf_{t_i,d}\right)^2\log\left(N/df_{t_i}\right)^2}$$

*w(t,d) = tf*idf weight of term t in document d,*
$tf_{t,d}$ = frequency of t in document d,
df_t = number of documents that contain term t,
N = number of documents in the collection.

The result is that terms which occur more often in one item than another, may have a more central role in the topic of these items and end up having a higher weight value for these term-, item combinations. What is so special about this is that the process does not rely on items that are specifically made analysable for machines, since it is capable of perfectly recommending without the need for proper understanding [1].

Next, a user profile containing the user's interests and its history of interactions with the system has to be generated. User customization may happen manually in the form of an interface that lets users rate items or select their own fields of interests, after which relevant items will be showed to the user. However, this explicit form of manual data collection sometimes lacks in accuracy since, first of all, it requires effort from the user, which is lazy in principle. Secondly, the interests of users may change over time, which results in an out-dated and inaccurate user profile. The interaction history might be used for simple purposes such as displaying a list of recently visited items, but the main reason this information is stored is that it will be used as implicit training data for algorithms that generate the user model which solves the explicit data collection shortcomings as explained above. This data can be stored under multiple categories, so that it will become more usable for the training algorithms. Categories that are frequently used are “things user likes” and “things user dislikes”.

The final step is the matching process, in which traditional machine learning algorithms like Rocchio's algorithm are used to higher or lower weights, and Bayesian classifiers, decision trees, and linear classifiers to estimate the chance that users will like the presented items [2].

So far we have learned the concept of behavioural targeting, and I have shown you a specific behavioural targeting approach, which is content based recommendation systems. This approach focusses on item descriptions and on this basis matches items to users that show interest in these items based on knowledge from user profiles. Content based filtering seems to be a well functioning approach but it lacks efficiency in some way. Imagine a person likes the Harry Potter books, and so far he has only shown interest in reading Harry Potter books, nothing else. A content based recommendation system is then likely to only show Harry Potter books that the user has not read yet, or items that are very closely linked to the Harry Potter genre; items that match the description: magic, school, fantasy, etc. Imagine that this particular user has read all seven Harry Potter books, and is looking for something new, yet slightly matching the Harry Potter genre. A content based recommendation system is therefore too specific in giving recommendations and is not going to be very helpful for this person. Besides, there is the problem of information overload that comes with content based systems. This problem arises when the amount of items are getting bigger and bigger, so that the system can have difficulty in making a proper decision. This may result in some items never been recommended.

In chapter 4, I will show you a recommendation system that solves these specificity problems, and will therefore possibly be a more successful behavioural targeting approach. The system in chapter 4 does not only concern items and one active user, but regards multiple users. To make the possible benefits of such a system more clear, I will first explain the concept of collective intelligence in chapter 3, as well as how this concept can be put into use in online advertising.

3. Collective intelligence

Let me begin by explaining the notion of collective intelligence by means of an example. In 1991, Loren Carpenter invited hundreds of people to a place with a gigantic screen and a small paddle for every person in the audience, where each paddle had two different sides: red and green. It did not take long for the audience to realize that if they held the paddle up in the air, showing either the red or the green side, it was represented on the screen by a dot of the same colour. Next the screen changed into the earlier computer game Pong, which the audience had full control over by showing either the red side-, for down, or green side of the paddle, for up. However, if the bat on the screen required to move a little upwards, some people were actually showing the red side that made the bat go down, if otherwise, the bat on the screen would go all the way to the top and miss the ball. What's fascinating about this, is that it all happened without any instructions whatsoever, but together the audience managed to get positive results. Why is it that a large group of people in which there is no hierarchy of any sort still can be so successful in playing this game of pong? Apparently, without a lot of individual knowledge, when we all work together and share a common goal, we are more efficient.

The first person to actually develop the term collective intelligence was Pierre Lévy [3]. He calls it: "A form of universally distributed intelligence, constantly enhanced, coordinated in real time, and resulting in the effective mobilization of skills". What he means by universally distributed intelligence is fundamentally that no one knows everything, yet we all know something. Properly maintaining this notion of collective intelligence therefore clearly makes us smarter beings. Now how does this view relate to advertising?

3.1 Collective intelligence and online advertising

Along with the internet came many opportunities, among which the access to lots of information, but more importantly, collaboration. This allows us to interact with each other in ways where time and geographical location are no longer relevant. Where advertising used to be knows for a lot of individual agencies that offered their services, there is now only a couple of multinational companies controlling the business. And besides, it is possible to integrate multiple users into the advertising process to potentially create more precise advertising.

In the previous chapter, we have seen a behavioural targeting approach that focussed merely on one user at the time, but what if we change this and focus on a large amount of users? According to the notion of collective intelligence, we become more effective when combining and universally distributing our intelligence. Does this notion still hold when applying it to advertising? In the next chapter, I will discuss a behavioural targeting approach that implements this idea of combining forces among users, and evaluate whether this approach increases efficiency compared to the non-collective approach discussed before.

4. Behavioural targeting based on collective intelligence

So far I discussed a system that focusses on one user and matches items based on the user's interests. After which I explained the notion of collective intelligence. The point of this chapter is to show a system that is closely linked to the one from chapter 2 but implements the idea of collective intelligence. The result is a collaborative filtering approach which I will explain in detail in the next paragraph. At the end of this chapter I will evaluate it's functioning and make a consideration of whether the idea that collective intelligence is more effective also applies in this context.

4.1 Collaborative filtering

In contrast to content based filtering, collaborative filtering puts its focus on multiple users. Compared to the content based filtering approach, in which item recommendations might be too specific, the collective approach discussed in this chapter is more controllable and therefore, can give better recommendations. In this approach the behaviour of specific users is being compared to that of other users to detect similar patterns, so that items will be recommended based on what users with similar interests like. Collaborative filtering therefore be more reliable than the approach discussed in the previous chapter.

For the execution of the collaborative filtering process to take place, it has to go through three stages [5]. In the first stage the user's neighbours have to be assembled, in which each neighbour counts as equal. In the second stage, algorithms are used to create a list of N best item recommendations for the user based on the neighbourhood. And in the final stage, the best N list is being evaluated. In the rest of this paragraph I will dive deeper into each of these stages.

stage 1: assembling the user's k-nearest neighbours

One important factor for the generation of the user's neighbourhood is to define the size of the neighbourhood, the value for k. A low k-value implicates a small neighbourhood and can therefore affect the overall accuracy of item recommendation, since the number of neighbours will not be sufficient. On the other hand, a large k-value that implicates a large amount of neighbours, also affects accuracy, since item recommendation can result in being averaged, so that there won't be enough new recommendations. Therefore, the size of the neighbourhood relies on the kind of data, and should be examined thoroughly for the optimal result. One way to do this is to let the algorithm decide what is best by creating neighbourhoods of varying sizes, and based on effect rates, in this case meaning the actual amount of users that click, make the future neighbourhood size respectively smaller or larger. A/B testing algorithms are designed to do this, and are easy to implement. Another benefit of implementing this way of varying neighbourhood sizes is that when the amount of users or items increases, the algorithm will automatically update k to the perfect new neighbourhood size for optimal results.

For the similarity of two users to be calculated, the Pearson correlation coefficient can be used [6]. In statistics, the Pearson correlation is the most common used method to measure correlation, meaning how well two or more things, in this case users, are related to each other, and it is given by this:

$$w(au,u)=\sum_i\left(\left(v_{au,i}-\bar{v}_{au}\right)\left(v_{u,i}-\bar{v}_u\right)\right)\div\sqrt{\sum_i\left(v_{au,i}-\bar{v}_{au}\right)^2\sum_i\left(v_{u,i}-\bar{v}_u\right)^2}$$

w(au,u) = weight of the active user au with user u,
$v_{u,i}$ = value of user u for interest in item i,
and $\bar{v}_u$= mean interest for user u:

$$\bar{v}_u = \left(1 \div |I_u|\right) \sum_{i \in I_u} v_{(u,i)}$$

I_u = amount of item interests user u has.

In the tables below is an example in which the similarity of the active user (AU) is calculated compared to that of other users (U1 and U2). The first table contains values that represent interest in items stored in the user profile.

	Item 1	Item 2	Item 3
AU	2	1	3
U1	1	3	5
U2	3	1	3

In the following table are the results of applying the Pearson correlation coefficient to the values in the table above.

	U1	U2
AU	0.5	0.87

The user with the highest value (U2) is therefore the active user's nearest neighbour.

stage 2: calculation of the user's N-best item recommendations

One important factor that should be taken into account while generating the list of N-best item recommendations is the size of the recommendation list, the value for N [5]. When N is large, the amount of relevant items will also be large, but the accuracy of some items might not be of sufficient value.

The calculation of the N-best items can happen in an easy manner. Intuitively, one might say that it is simply a matter of counting the item frequency inside the assembled neighbourhood, resulting in the N most frequent items being the N best recommendations. And this idea is true; item frequencies of the active user's neighbourhood are calculated, and items that were already part of the active user's profile are subtracted from this list. Lets look back to the example in the tables above. Assume that the nearest neighbourhood size in this case is set to one, and the size of the N-best item list is set to 2. User 2 would be the entire neighbourhood then. Apart from the items that were also present in the active user's profile, user 2 also has the items 4 (with a value of 3), 5 (with a value of 1), and 6 (with a value of 2) listed in his profile. Then the N-best item list would therefore show items 4 and 6.

stage 3: evaluating the N-best item recommendations

After the item list is presented and a user ignores the item, nothing really happens. But when the user clicks the item, something has to happen. A value needs to be calculated and the item needs to be added to the user profile. The value can be based on a rating from the user, however, the user does not necessarily have to rate in order for a value to be stored. Besides, user ratings are very often inaccurate since it is difficult for users to remain relative to previous ratings, especially when given over time. Therefore, we can calculate the predicted interest value for users, the following function calculates this for the active user a [6].

$$P_{au,i} = \bar{v}_{au} + k \sum_{u=1}^{n} w(au,u)(v_{u,i} - \bar{v}_u)$$

n = the amount of users,
k = normalizer.

The function is based on the mean interest value of the active user. And adds (or subtracts) the amount that the active user's neighbours differs from their respective mean interest value. The reason that we need a normalizer k, is to make sure that the summed absolute values of the weights w(au,u) is equal to one.

4.2 Possible improvements

There is however, a large problem that may arise when implementing collaborative filtering. This is the problem of information overload, because of the countless available options. If the nearest neighbourhood of an active user has to be determined, and the algorithm has to go through all available options meaning all available users, the algorithm will become very inefficient if the amount of users is growing.

One effective solution to this problem is cluster analyses [7]. In cluster analyses, users are being grouped or classified into clusters based on similar attributes. The goal is to create subsets, which each have their own shared characteristics so that for example ten thousand users can be reduced to one hundred subsets or clusters. One important thing to keep in mind is setting the amount of attributes for users to be similar to become part of a cluster. If this value is very low, say for example two, then clusters will become very large so that recommendations may lack in relevance. On the other hand, setting the amount of attributes to a very high value causes clusters to be very small and thus the problem of having too many clusters may arise, so that the algorithm remains inefficient.

Collaborative filtering is having a major breakthrough in social advertising on networks like Facebook, since those networks are fundamentally based on specific types of social relationships like friendships and likes or interests. Based on the behaviour of a specific user on a social network site it is therefore easy to determine its K-nearest neighbours.

In this chapter we have seen collaborative filtering; the behavioural targeting approach that makes use of collective intelligence, and solves the specificity problems that the content based filtering approach is afflicting with. Image the Harry Potter example again. Recall that in the content based filtering approach, the recommendation system will not be able to come up with new recommendations when the user has only shown interests in Harry Potter books. In the collaborative filtering approach, however, the recommendation system will look for other users that also show great interest in Harry Potter, and based on their other interests show recommendations to the user. This problem can thereby be solved by collaborative filtering, because collaborative filtering does not solemnly rely on the information of one user, but on the collective information of multiple users. The next chapter will be about legal issues that arise when implementing behavioural targeting, because unfortunately, there are some.

5. Legal implications

So far we have seen a number of approaches that were all pro behavioural targeting. However, behavioural targeting brings forth the ability to create a very detailed and therefore privacy sensitive profile of internet users [8]. Examples of stored privacy sensitive information include things the user bought at certain stores, medical conditions the user has, etcetera. Advertiser's expectations of behavioural targeting are enormously high but at the same time there is a lot of critique since this privacy sensitive information can apart from being easily used also easily be abused. Therefore it raises the question of whether this form of advertising is in conformity with our privacy regulations. In the next paragraph I will contemplate this issue and give an answer to this question.

5.1 Telecommunications act

To answer this question it is wise to start by looking at the Dutch telecommunications act [9] which states that spam is illegal. It states that information regarding commercial interest may only be send if the user has given permission for this to happen. However, this ban only holds for sent messages that can be stored in the user's devices until they are being retrieved, and therefore does not concern advertisements. Apart from this the act states that the placement of cookies is only allowed if the users are properly informed about its uses beforehand, and a recent change in regulations ads to this that users must give the placement actual permission before it may happen so. There is, however, and exception for services that require the placement of cookies for their proper operation. Most of the times this involves services that require logins like online banking websites or email clients. This exception does not apply to behavioural targeting [10]. Therefore, it is obliged to show the users information regarding the placement of the cookie and its functionality at the first visit, in the form of a pop-up or something similar. And in addition to this, a question must be asked that requires the user's permission as an answer. This is, however, not being applied by most of the online advertising providers.

5.2 Personal data protection act

Legal issues that relate to behavioural targeting discussed above were about the placement or storing of the data, and the following bit of information will relate to the actual usage of this data. The Dutch personal data protection act [11] states that with 'personal data' they actually mean: any information concerning an identified or identifiable individual. Identified meaning that there is an identifier that separates the person from the rest. A name, for instance, is an identifier. People that try to defend behavioural targeting will say that there will be no information stored that can be linked to an actual user, but unfortunately, this is false; the identity of the user can be determined. They have decided that IP addresses do count as personal data because they can be linked to actual people. There are of course some exceptions in the form of public computers, but in most cases IP addresses are a form of personal data and so the personal data protection act applies. What does this mean? It depends. For a decision to be made, a consideration must be made between the concerns of both parties. The advertisers concern is mostly commercial and aims on giving its users a more personalized advertising experience, where the user's concern focusses entirely on the protection of its privacy. In the case of on-site behavioural targeting (paragraph 2.1) violation on the user's privacy is relatively limited, since the information is very often limited to only one field of interest, so the advertiser's concern will most of the time be preferred. In the case of network behavioural targeting (2.1) however, the user's concerns are mostly preferred, since the stored information covers a far more broad view, so that a lot more privacy infringing data can be exposed. It is therefore decided that in the last case, the advertiser needs the user's permission to use its personal data, and this does not happen in most cases.

Taking these legal issues into consideration definitely raises a number of questions. Should advertisers change the implementation of behavioural targeting so that users can easily decide for themselves whether or not to be part of those networks? Or should we stop using behavioural targeting at all? Due to the great success of this advertising strategy it is not likely to be down-shifted into a less effective method. Besides, a vast majority of the population does not seem to care at all about what happens with their personal data, and likes the idea of personalized advertising. This again raises other questions. Should regulations conform privacy on this area change? I do not think it is likely that anything will change in this field in the near future. If we want to maintain any kind of informational privacy, we have to accept the fate that we have to take responsibility for ensuring this for ourselves, in the means of browser regulations like opt-out systems. Should we accept the faith that we are more efficient beings if we work together and combine our intelligence?

6. Conclusion

In conclusion, I have showed you two key approaches in behavioural targeting. A Content-based filtering system which focusses on item descriptions and user profiles, and collaborative filtering, which is the collective approach and focusses on making recommendations based on looking for patterns with other users.

The content-based approach lacks efficiency in some ways since it will only recommend items that are very closely linked to the interest profile of the user, so that no varied recommendations can be made. The question now is whether the collaborative filtering approach solves these issues and can therefore give better recommendations. The answer is yes, collaborative filtering will show more varied items than content-based filtering. For the question of whether this collective intelligence approach improves behavioural targeting in general, a lot of testing is yet to be done. Experiments with actual users has to be done to look at actual results and sale rates.

Until then, both content-based filtering and collaborative filtering have their strengths so that therefore the solution can be to implement both. A hybrid form of behavioural targeting containing both content-based filtering and collaborative filtering might show more effective results than each of the two separately. There are different ways in which this can be achieved, one of these being letting both approaches make separate predictions and combining the results.

Bibliography

1. Pazzani, M., and D. Billsus. “Content-based Recommendation Systems.” *The Adaptive Web* (2007): 325–341.

2. Segaran, Toby. *Programming Collective Intelligence: Building Smart Web 2.0 Applications*. O’Reilly Media, Inc., 2008.

3. Kessler, P. R. “Collective Intelligence and the Paradigm Change of Collaboration.” *P. R. Kessler.* October 26, 2010.

4. Pandey, S., M. Aly, A. Bagherjeiran, A. Hatch, P. Ciccolo, A. Ratnaparkhi, and M. Zinkevich. “Learning to Target: What Works for Behavioral Targeting.” In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, 1805–1814, 2011.

5. Symeonidis, P., A. Nanopoulos, A. N. Papadopoulos, and Y. Manolopoulos. “Collaborative Recommender Systems: Combining Effectiveness and Efficiency.” *Expert Systems with Applications* 34, no. 4 (2008): 2995–3013.

6. Breese, John S., David Heckerman, and Carl Kadie. “Empirical Analysis of Predictive Algorithms for Collaborative Filtering.” In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, 43–52, 1998.

7. Caliński, T., and J Harabasz. “A Dendrite Method for Cluster Analysis.” *Communications in Statistics* 3, no. 1 (1974): 1–27. doi:10.1080/03610927408827101.

8. Greengard, Samuel. “Advertising Gets Personal.” *Communications of the ACM* 55, no. 8 (August 1, 2012): 18. doi:10.1145/2240236.2240243.

9. “Telecommunicatiewet”, 19-10-1998, http://wetten.overheid.nl/BWBR0009950

10. Koëter, Jeroen. “Behavioral targeting en privacy: een juridische verkenning van internet gedragsmarketing.” *Tijdschrift voor internetrecht (2009),* 104-111.

11. “Wet bescherming persoonsgegevens”, 6-7-2000, http://wetten.overheid.nl/BWBR0011468