Learning to Classify Users in the Buyer Modalities Framework to Improve CTR*

Laurent Mertens[1,2], Peter Coopmans[3], and Joost Vennekens[1,2]

[1] KU Leuven, De Nayer Campus, Dept. of Computer Science
J.-P. De Nayerlaan 5, 2860 Sint-Katelijne-Waver, Belgium
firstname.lastname@kuleuven.be
[2] Leuven.AI - KU Leuven Institute for AI, B-3000 Leuven, Belgium
[3] Addrelevance
Grevensmolenweg 28, 3800 Sint-Truiden, Belgium
peter.coopmans@addrelevance.be

Abstract. The Buyer Modalities framework divides buyers into 4 profiles, where each profile has its own specifics as to how it makes its purchasing decisions. We built an online prediction system that categorizes website visitors based on this framework. According to this categorization, a specific banner ad variant tailored to that profile was shown to the visitor, rather than a default “neutral” variant, resulting in a significantly improved CTR.

Keywords: Data mining · Predictive modeling · Ensemble methods · Online advertising.

1 Introduction

The Buyer Modalities framework [3] is a model that defines four distinct psychological profiles of consumers according to how they make their purchasing decisions. These four types – competitive, spontaneous, methodical and humanistic – are illustrated in Fig. 1, and are based on two main axes: decision speed (impulsive vs. deliberate) and rationale (emotional vs logical). It states that each profile reacts to different types of information. If we consider, e.g., the purchase of a new car, people with a methodical profile will be more interested in a detailed list of features of the car as can be found in the brochure, whilst the humanistic profile will be more served with testimonials from people who already own the car.

The implication of this model for advertising is that in order to have an effective campaign, ideally each profile is targeted with an ad tailored toward its information needs. The issue with this of course is that one needs to know the profile of the user, which one typically does not. In order to remedy this issue, we propose a framework that uses historical user-website interaction data

* Supported by the Flanders Innovation & Entrepreneurship TETRA project “Start To Deep Learn”.

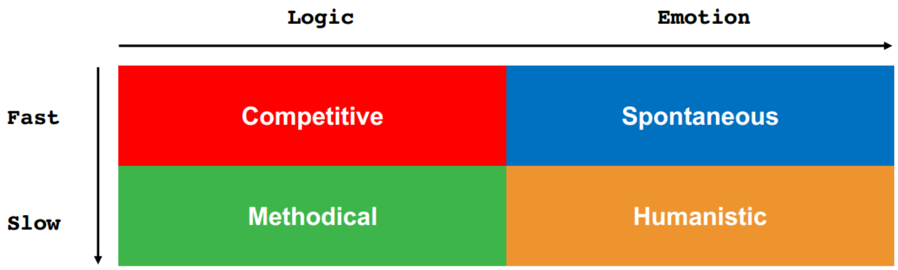

Fig. 1. A schematic overview of the four psychological profiles described by the Buyer Modality framework.

to predict the user profile when needed, and hence allows the ad server to display the appropriate variant for this particular user in a dynamic way. We show that this approach results in a significant increase in the click-through rate (CTR), i.e., the percentage of users viewing a web page who click on an ad displayed on that page.

Ads come in many forms, e.g., pop-ups or "sponsored content", each with its own specifics. Our work involves so-called "banner ads", banner-like graphical ads often displayed at the top or in the margins of a page. Ever since the advent of online advertising, the CTR has served as a key measure for the succes of an ad campaign. Given the multi-billion industry that is online advertising, the question of what makes an ad effective has been given quite some attention.

Specifically for banner ads, [8] look at the effect of the banner ad size, style and orientation on its success. In [1], the authors attempt to predict the CTR, rank according to CTR and categorize into "high" and "low" CTR a set of ±10K banners by using a custom defined set of 43 different visual features. In all three tasks, they manage to consistently outperform the baseline. In contrast, [7] performed two eye tracking studies to investigate the relation between visual design and relatedness to the page content, and visual attention devoted to the ad. In a first study, they show a professionally designed graphical ad to one group of participants and a text-only banner (with the same text as the graphical ad) to another. Besides this, half of the ads (graphical and text-only combined) were related content-wise to the page, whilst the other half were not. They found that none of these parameters had a statistically significant effect on the dwell time. This prompted a second study, in which they showed that dwell time does increase significantly if an ad is relevant to a user's intent or task, rather than to a page's content. Somewhat closer ideologically to our work is [4], who studied the effect of demographically targeting banner ads on users' visual attention and brand evaluation. This kind of targeting focuses on demographic properties of the users such as gender, age and location, and follows the assumption that "similar people act in a similar way". Hence, by tailoring ads to these properties, it should be possible to increase user attention. They found that targeting ads this way does indeed increase users' visual attention, but not necessarily their brand evaluation.

2 Conceptual setup

In this section we will first provide a conceptual description of our work, followed by a detailed description of the concrete implementation for our use case in the following section.

Conceptually, our setup is the following. Given a particular website that displays ads, we track certain user-website interaction data and use this historical data to extract features to be fed to a predictive model. Ideally, an expert should identify salient elements to be tracked, e.g., specific hyperlinks or other so-called Calls to Action (CTA), that are likely to appeal more to one profile rather than the others.

We distinguish two phases. During a first phase data is being collected to be used to train a predictive model. During this phase, different variations of the same ad targeted to each profile will be displayed at random. This means that a single user can get served different variants of the same ad. When a user clicks on one of the variants, we assign the associated profile to the user to obtain the training targets.

During the second phase, we continue to collect user-website interaction data, and use this data to query the predictive model we obtained in phase one in real time to obtain a user profile. This prediction then determines which ad variant will be shown to this user.

3 Case specifics

Our experiment was performed in collaboration with a commercial partner, Produpress [6], a company that owns amongst others a number of automotive magazines and corresponding websites. We worked with two websites, www.autogids.be (Autogids) and www.moniteurautomobile.be (Moniteur), which are essentially the Dutch and French language versions of the same content. Both sites target a Belgian audience. A large part of the content are extensive car reviews. There were some differences in data collection during training and deployment phases, which we will discuss in the following sections. These differences are due in large part to the fact that data collection was performed by a third party for the first phase, whilst being performed by ourselves for the second phase.

3.1 First phase

In this phase, four different variations of an ad banner for a specific car ad where designed, one for each buyer modality, which were shown on a random basis. The main difference between the variants was the CTA used (i.e., text), rather than the graphics. If a user clicked on a banner, the profile targeted by the variant becomes the profile of the user. E.g., if a user clicked on the "competitive" variant, when training the model later on we take this user as being a target for the "competitive" profile.

The number of positive samples we collected per profile can be seen in Table 1. Note that some users clicked on more than one ad variant; there were 353 unique users who clicked on a variant for a total of 370 clicks. These users where treated as targets for all the variants they clicked on. The numbers used in this and following tables correspond to the different types as follows: 1 = competitive, 2 = spontaneous, 3 = methodical and 4 = humanistic.

Table 1. Number of collected positive samples per profile

	Autogids	Moniteur	All
1	89	79	168
2	46	30	76
3	38	26	64
4	34	28	62
Total	207	163	370

Besides target labels, also aggregated and custom features for each distinct user were collected. The 7 custom features basically correspond to (the URLs leading to) the main sections of the car reviews, here translated from Dutch: "Read our test report", "View the gallery", "Robotportrait and conclusion", "Tested version", "Users reviews", "Compare this car" and "Find a dealer". With these, the set of features for this specific experiment consists of, per user and over the data collection period (abbreviations correspond to Fig. 2):

- The number of pageviews. (PgV.)
- The average time spent per pageview. (AtP.)
- The number of sessions. (#Ses.)
- The average time spent per session. (ASD.)
- Per custom feature: the number of sessions the user saw this particular content type, i.e., clicked on the corresponding URL. (CT1–CT7)
- Per ad variant: the number of sessions the user was shown this particular ad variant. (AV1–AV4)

Note that at this stage, we did not have any other information besides these aggregated features. This means that we were unable to determine when exactly a user clicked on an ad, which in turn means that all aggregates were determined by also taking into account data from *after* when a user clicked an ad. Ideally, these statistics would have only been determined by using data prior to a click. Fig. 2 shows the average feature values between clicking and non-clicking users for each ad variant separately. This graph clearly illustrates that indeed there appears to be a behavioral difference between both groups of users, as indicated by the fact that for "click" samples the average values are consistently higher than for "no click" samples.

To further analyze the data, we first checked whether or not we could distinguish between "click" and "no click" samples in general, regardless of profile type. For the remainder, all models were trained using Python's Scikit-learn package [5]. Table 2 contains the average accuracy over 200 Random Forest classifiers, each consisting of 200 trees with `max_depth = 3`. For each iteration, a random selection of negative samples was chosen to complement the positive ones, and 20% of the data was held out as test data. As the data shows, performance is far better than random, although also far from perfect, with Autogids and Moniteur performing very similarly, suggesting that it is indeed possible to predict what users are more inclined to click on an ad, regardless of profile type.

Table 2. Click vs. No Click classification performance over all ads by means of a Random Forest classifier. Average performance over 200 forests of depth 3.

Dataset	Train	Test
Autogids	0.743 ± 0.017	0.679 ± 0.054
Moniteur	0.738 ± 0.018	0.655 ± 0.062

In a next step, we checked to what extend it was possible to distinguish between each pair of profiles. The assumption is that if there is no correlation between user profiles and ad variants, users will randomly click an ad variant and hence it will not be possible to discriminate between ad pairs. To test this hypothesis, we again looked at the average accuracy over 200 Random Forest classifiers with 200 trees each and `max_depth = 3`, with an 80/20 train/test data split. The results are shown in Table 3, and except for the last pair (3 vs 4) show performance that is in line with the "click" vs "no click" scenario. This indicates that it is possible, up to a point, to discriminate users based on the ad variant they clicked. In other words: different people do have a different ad variant preference.

Table 3. Full dataset: Random Forest accuracy per ad pair.

Ad Pair	Train	Test
1 vs 2	0.767	0.650
1 vs 3	0.836	0.689
1 vs 4	0.852	0.685
2 vs 3	0.854	0.576
2 vs 4	0.857	0.612
3 vs 4	0.806	0.501

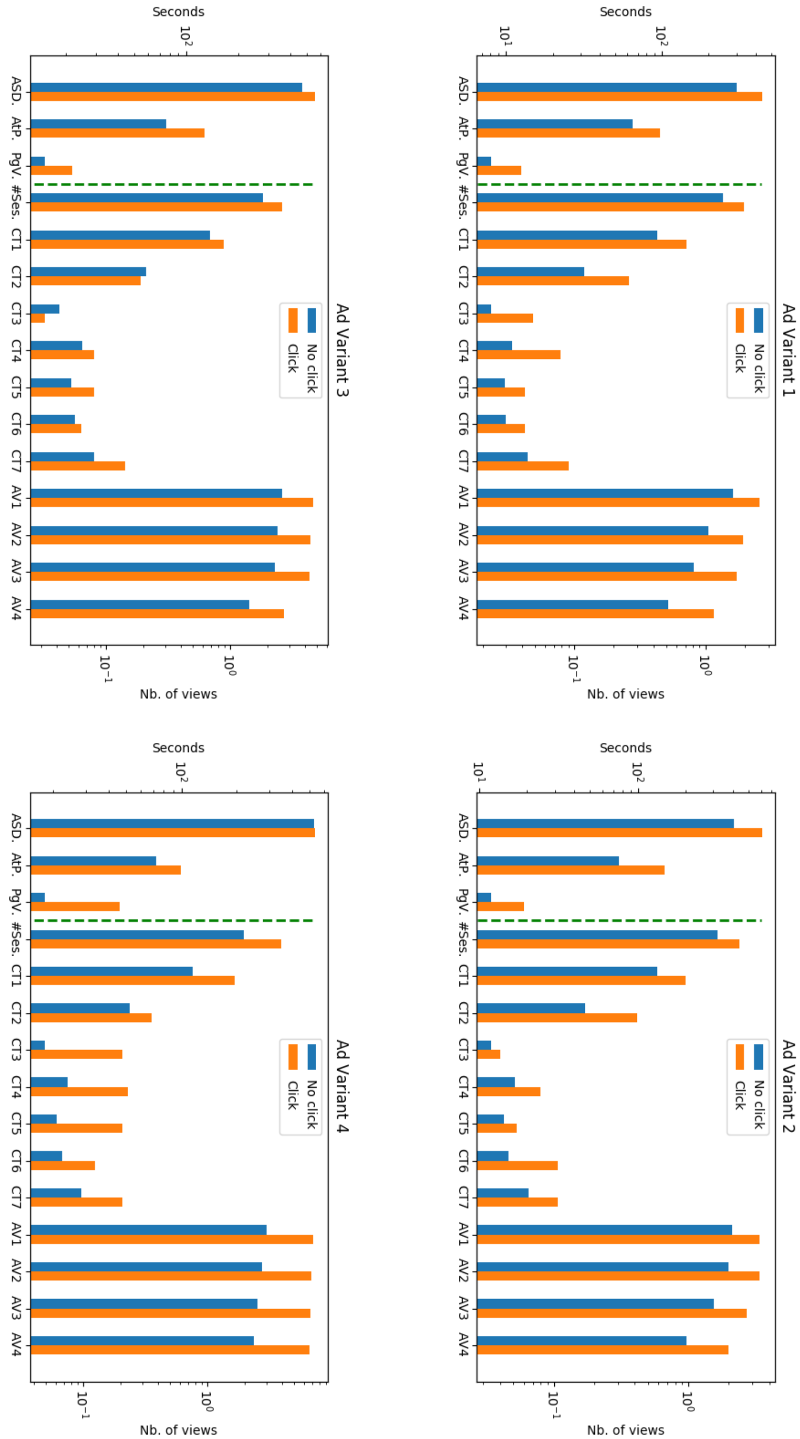

Fig. 2. Averages per feature between “click” and “no click” samples for each ad variant.

Since our assumption is that each user, regardless of whether they click an ad or not, can be described by one of the four buyer modalities, we wish to train a model that always predicts one of these profiles, and does not make a "neutral" prediction. As training data we used all positive samples of all four variants combined (i.e., no negatives), and again opted for a Random Forest model with 200 trees, albeit multiclass this time, to predict one of the four profiles for each user. Train and test sets were stratified so as to have equal ratios of samples per class. Table 4 shows results with `max_depth = 3` and `max_depth = 10` settings by again taking the average over 200 iterations, with an 80/20 data split at each iteration.

Table 4. Full dataset: multiclass Random Forest accuracy

`max_depth`	Train	Test
3	0.515 ± 0.013	0.470 ± 0.022
10	0.946 ± 0.011	0.405 ± 0.043

Performance is considerably lower than with previous experiments, even taking into account the fact that the baseline is 0.25 this time. As can be expected, the experiment with `max_depth = 10` results in overfitting, as apparent by the large discrepancy between train and test accuracies. Nevertheless, we chose to go with this model for phase 2, as our philosophy was that given the low number of samples at our disposal, we preferred the model to overfit on these so that they can serve as stringent prototypes, rather than making a more "diffuse" model.

3.2 Second phase

The second phase was ran in light of a specific advertisement campaign for a new car. The compaign ran for four weeks total; two on Autogids and two on Moniteur. Similar to phase one, five variants of the ad banner were made: one for each profile, plus a "neutral" variant in case the profile of the user could not adequately be predicted. A major difference with the first phase is that we collected the user-website interaction data ourselves. This allowed us to compute the features in an online way. Data was collected by means of a custom JavaScript script, that would send the data to a PHP service to be stored in a MySQL database. Information stored included, a.o., a unique user ID, page visits and clicks.

To allow the website to request a profile for a visiting user, we developed a Python API using CherryPy [2]. Whenever a profile was requested, the known data for this user would be retrieved from the SQL database, and the same features as used during phase one computed on the fly. A prediction was only made if the user had visited the site during at least 3 sessions (including the one at prediction time). If this was the case, we would then verify that the highest profile score returned by the model > 0.35. If so, the corresponding profile would be returned. In all other cases (also if the user ID was unknown), a default "neutral" profile would be returned. Recall that we used the `max_depth = 10` model described at the end of §3.1, whose performance is shown in Table 4.

We are not allowed to report specific CTR numbers because of contractual obligations to our commercial partners. Hence, we can only report changes w.r.t. the baseline. For both Autogids and Moniteur, a baseline CTR was determined over 14909 and 12502 impressions of the "neutral" banner respectively. This means that the users that belonged to this control group did net get to see a banner based on our predictive system. The CTR for our system was determined over 63284 and 56751 impressions respectively. This does not mean that all users belonging to the test group saw a customized banner, simply that for these users, we attempted to make a prediction. The CTR on banners displayed using our system were 31% and 35% higher than the baseline CTR for Autogids and Moniteur respectively, for an average increase of 33%.

Given this result, it was decided to run a second campaign, for a different car by the same brand as the first campaign, using our system over a period of four weeks, but without further involvement from our part. CTR for this campaign were 129% and 94% higher that the baseline determined in the previous campaign. Unfortunately, a new baseline was not determined and hence these numbers are only reported by way of indication.

4 Conclusion

In this work, we described how the Buyer Modalities framework can be used to improve the CTR on online ads. We built a Random Forest model based on features extracted from aggregated web analytics, and used this model in a system that allows to predict the Buyer Modality profile of a website visitor. Using this predicted profile to dynamically adapt the ads shown to the user resulted in a 33% improvement in CTR compared to the reference user group.

We would like to point out that our method theoretically does not require a new training phase for each new ad, since although the ads change, the user modality profiles do not. This implies that, given careful design of the ad variants, once a model has been trained it should be applicable to any ad campaign. Consequently, by collecting data over several campaigns, the model can also continually be further improved by incrementally retraining the model.

Moreover, for this particular experiment the raw data consisted of aggregated features. We expect that having data available at a more granular level, as collected by ourselves in phase 2, should allow the development of more and better features to further improve the accuracy of the predictive model.

References

1. Azimi, J., Zhang, R., Zhou, Y., Navalpakkam, V., Mao, J., Fern, X.: Visual appearance of display ads and its effect on click through rate. In: Proceedings of

the 21st ACM International Conference on Information and Knowledge Management. p. 495504. CIKM '12, Association for Computing Machinery, New York, NY, USA (2012). https://doi.org/10.1145/2396761.2396826, https://doi.org/10.1145/2396761.2396826
2. CherryPy: https://cherrypy.org/
3. Eisenberg, B., Eisenberg, J.: Waiting for Your Cat to Bark. Thomas Nelson (2006)
4. Kaspar, K., Weber, S.L., Wilbers, A.K.: Personally relevant online advertisements: Effects of demographic targeting on visual attention and brand evaluation. PloS one **14**(2), e0212419–e0212419 (2019)
5. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
6. Produpress: https://www.produpress.be/
7. Resnick, M.L., Albert, W.: The influences of design esthetic, site relevancy and task relevancy on attention to banner advertising. Interacting with Computers **28**(5), 680–694 (2016)
8. Sigel, A., Braun, G., Sena, M.: The impact of banner ad styles on interaction and click-through rates. Issues in information systems **9**(2), 337–342 (2008)