Utilising The EM Conceptual Framework in The Study of Targeted Advertising

Judy Palimonka
1266015

Abstract

Targeted Advertising (TA) facilitates the process of adapting online adverts to users interests based on their past online activity. The process of creating targeted advertisements can be easily accounted for by defining a few of the core concepts and indicating the relationships between them. The process itself is however very implicit, therefore despite its simplicity, the public often seems to misunderstand the core ideas in TA or makes incorrect assumptions which may rise unnecessary concerns, create misconceptions and cause the misinformation in the Internet and public awareness. Users have limited understanding of web technologies and there are also limitations to what browsers can reveal with regards to the TA activities performed in the background. This paper will discuss the suitability of the EM framework for presenting this type of problems with the EM model developed specifically for this project in JS-EDEN. The model will allow for developing a construal of TA's by showing how the concepts are involved in processes being part of ad targeting. This EM artefact will also help the modeller visualise what happens when we decide to block trackers and indicate the impact of blocking on the information collected by the firms following user as well as the information being displayed by the website. We hope that the knowledge gained during the interaction will allow develop the understanding of TA which will eventually let the modeller take a stand on the privacy related issues.

1 Introduction

This section provides background information about Targeted Advertising (TA), the motivation for choosing this particular topic and discusses the suitability of EM framework for presenting the ideas in TA.

1.1 Background Information

Targeted advertising (TA) allows for adapting advertising content to a particular user or an audience of users. The fundamental technique to TA is web tracking. It is accomplished by first, tagging user with a unique ID and then associating user's web usage with the ID. The examples of gathered data can be the history of visited websites, a list of accessed links, click stream, purchases and many more. The categories of data collection depends on who collects the data and where the data is sent to. The categories include first-party, third-party and fourth-party tracking. The former technique is utilised by the web publishers who use the collected data in order to advertise their own products and, since they do not share it with others, the process is considered to be relatively safe. Websites employing third-party tracking, on the other hand, allow other websites (trackers) to collect data about each visitor. Fourth-party tracking is usually a result of third-party tracking and takes place when the trackers pass the received data further onto other companies.[3]

1.2 Privacy Issues and Motivation

The third-party and fourth-party tracking are hidden from the user and there is no effort made from the web publishers side to obtain users consent in an explicit manner. The lack of transparency in this area of TA raises many privacy issues among Internet users and privacy enthusiasts. Current research [1, 2] suggests that those aware of web tracking, are concerned about some specific issues such as the amount of collected data, its sensitivity (with respect to users), the stage or time of data collection, the entities collecting data and their intentions. Furthermore, the targeted adverts are not annotated (or vaguely annotated) which makes it more difficult for the user to see the connection between web tracking and advertising. However, this only applies to the people who are aware of the practice. There is still a significant number of users who are not aware of or are vaguely familiar with the web tracking for advertising purposes. The lack of visibility in TA is quite problem-

atic. Firstly, it leaves the public in ignorance and secondly, it does not provide the choice to decide what happens to their data. On the other hand, in order to make an informed choice, the public would need to be provided with the practical means of learning about TA. Looking the implications of putting new cookie policy from May 2012 [4] in practice, asking user for consent on a visit to each website may not necessarily be an answer to privacy concerns and user's protection. The reason being that it does not guarantee that sufficient information is provided for the user to actually understand tracking and continuous prompts have a negative impact on browsing experience (may even cause users to ok the message without giving it any consideration). A better solution perhaps, would be to engage users with an interactive model that explains TA. In such a way, they could make a decision which would apply to all the websites they visit. The secondary purpose of this project is to address the above privacy concerns by providing the means to learn about tracking employed in TA. The model will help develop a construal of how web tracking is used in TA and introduce the modeller to the notions of first-party website, tracker, third-party tracking, fourth-party tracking (or sharing), targeted adverts and the relationship between these concepts. Apart from providing the artefacts for the interaction, the model will provide the visualisation of the blocking of web tracking. We hope such understanding will enable the modeller to make an informed choice about web-tracking.

1.3 Related Work

There is a number of applications that support the detection of trackers on entry to each website. *Ghostery* [7] and *AVG Do not track* [8] records the number of the trackers, research companies and advertisers present on each visited web page and allow for blocking certain addresses. *AdBlock* [9] allows for blocking non-targeted and targeted adverts. Google Chrome's *Incognito Mode* and Firefox *Private Browsing* allow users for clearing browsing session data after browser window has been closed, preventing the information collected by trackers from being preserved. These tools however only provide the measures for disabling web tracking. The authors provide some or little information with regards to the relationship of web tracking with TA and they still do not solve the problem of making user understand how this process is being used as a part of advert targeting. A program developed specifically for the purpose of explaining TA practices was a part of an unpublished BSc project called "Detective - Increasing Visibility of Targeted Advertising".[6] Apart from providing the list of companies following the user, it also had the functionality of indicating the exact stages of data collection, the type of data collected and it detected targeted adverts. Also, the relationship between the data collected and the adverts that were adapted to users have been indicated by providing the sources of information the tracker has used in order to produce a particular advert. In spite of the program providing the answer to the most of privacy issues mentioned in previous section, the impact of program on user's understanding did not seem to be too significant. The study with 100 users suggested that the users saw the process as very implicit and showed that they paid little attention to program's notifications. Our conclusion was that the actual act of browsing pages distracted the users from the notifications given by the program. In the end, the users noticed the number of trackers only and rather than investigating what type of data has been used to produce a targeted advert and they felt uncomfortable with someone following them on the web.

2 The Model

2.1 EM As a Methodology

In the light of the methods that failed to fully explain the operation of TA in an extensive and a simple to comprehend way, the benefits of utilising the EM approach is clear. Let's start with the drawbacks of the measures described in previous section. Presenting information textually have not proved very effective because it requires users to posses an understanding of web technology (to certain extent) and may fail to produce enough information to construct a clear comprehension of the concept, or in EM terms, form a construal. The latter technique (programs producing notification during web browsing) requires users to divide the attention between browsing and reading (and understanding) the notifications which also allows little time for memorising each action resulting in user's confusion. EM allows us for presenting the problem visually and letting the users interact with observables making them realise how the interaction immediately changes the state of observables and how they are connected by observables. Also, since there is more emphasis put on the state of observables rather than the actual act of browsing, therefore we can capture the necessary portion of user's focus.

2.2 EM Conceptual Framework

This section provides a high-level overview of the EM concepts identified in the model, a very general one that will only list the most explicit concepts. Description has only been added to the relatively less straightforward concepts.

Modeller Function

Modeller is a super-agent and has an indirect effect on all agents described below. In other words by one action, the super-agent triggers a chain of events where the final state depends on the action performed by agents, the state of observables and the dependencies. While the agents defined below cause changes to a number of observables, the model can't run independently of the modeller. The agents won't act unless changes are initiated by the super-agent.

Agents

The following agents are a source of random changes from the perspective of modeller:
Advert Generator
Party Detector

Key Observables

website, tracker, segment, advert, advertiser, blocking enabled

Subordinate Observables

virtual browser, advertising network, total trackers on current page, page viewed previously, data flow log

Main Dependencies

siteTrackerLinks : a website is associated with one or more trackers
websiteSegmentLinks : the subject of website is classified into one or more customer categories (or segments)
trackerAdLinks : a tracker is associated with one or more advertisers that are the source of adverts for the tracker
segmentAdvertiserLinks : an advertiser sells different categories (segments) of products

Dependency Implementation

Dependencies have been expressed in JavaScript arrays. The position in the array corresponds to an observable. In the example below, the position of the sub-array is a collection of trackers associated with a website ID equal to the position of sub-array in the main array.

```
siteTrackerLinks=[[0],[0,1],[2,3]...
```

Constants

The content of websites, addresses, the content for advertisers and others are all constants.

```
trackers=['GoogleAdSense','DoubleClick'...]
advertisers=['ebay.co.uk","amazon.com"...]
```

2.3 The Obstacles Encountered During Model Development and Changes

The initial plan was to make most part of the model consist of definitive scripts. Unfortunately at the situations when it was necessary for JavaScript variables to interact with EDEN observables, a few issues have been encountered. Firstly, the "execute" procedure used to execute JS script in EDEN context worked intermittently. It often did not produce the expected results especially when place inside of the scope of a method. In result, it was necessary to translate most of observables into JavaScript variables. The ability to translate them to observables turned out to be extremely useful and reliable under every circumstances. The dependencies between observables corresponding to JavaScript variables have been expressed using two dimensional arrays. For example, an observable was given an ID corresponding to the position of the observable/variable in the original array. The initial plan for the model considered the "fourth-party tracking" which would need to show the relationships between trackers. It was implemented at some point however the model seemed to have looked less clear and very confusing.

3 Conclusions and Recommendations

3.1 Model Evaluation

The objective of this model was to help modeller understand the key facts about Targeted Advertising, i.e. how the users are tracked, what information is collected and then used in order to target an advert. The key idea to convey is that TA is not 'evil'. Despite not asking user for permission to collect the data, the details gathered only includes non-PII data and trackers or advertisers do not attempt to identify user as a real person. The idea of *Tracker Profiles* section was to make user realise that the the collected information is used to make assumptions about user interests.

The *Advertising Network* part aimed at user's imagination and raised the richness of experience. The user should be able to understand why the websites show adverts related to our interests even though we have not visited them before (also indicated by the boxes *Visited Before* and *User Recognised* in *Visibility Status* section). In fact, the term network suggests that, once online, users become a part of integrated community of people automatically sharing the details with each other. The *Data Flow* gives a detailed overview of where the data goes and whom it is received by during the process of targeted ad creation. The ability to see the impact of blocking may make user feel safe and convince that tracking, even anonymous, can be prevented. We believe that the model would serve as an excellent tool for learning about the steps being part of TA process. A further study with modellers would provide more insight into how effective the model would be in conveying the main message to the users. Author is convinced however that such cognitive study would bring excellent results.

3.2 Future Improvements

Possible model improvements could include making it more real e.g. by incorporating the links to adverts; and also changing the algorithm for choosing targeted adverts. The current algorithm utilises "keyword-per-interest" approach which, with an adequate understanding of adaptive personalisation systems, could be "upgraded" to "multiple keywords per interest". [9] Also, each keyword or interest can be given weight based on the number of web page views. Graphical interface may take advantage of gradient functions being part of HTML canvas objects to make the model look more appealing and modern.

Bibliography

[1] OFC, A market study: Online Targeting of Advertising, London, UK: OFC, 2010

[2] IAB, A guide to Online Behavioural Advertising and Prices, London, UK: IAB, 2011

[3] J. Mayer, J. Mitchell, Third-Party Web Tracking: Policy and Technology, Stanford, US: Stanford University, 2012

[4] ICO, New EU cookie law (e-Privacy Directive), available at www.ico.gov.uk

[5] Palimonka J., Detective Increasing Transparency of TA, Mozilla Firefox, 2012 available at https://addons.mozilla.org/en-US/firefox/addon/detective-increasing-visibi (not updates since November 2012, may be incompatible)

[6] (not published, provided here only for reference) Palimonka J., Project Report, 2012 available at https://docs.google.com/open ?id=0BwMaPNkfLHPYaDhrekZSTWNoSGM

[7] Evidon, Ghostery application, available at http://www.ghostery.com/about

[8] AVG, Do not track application, available at http://www.avg.com/gb-en/do-not-track

[9] Gauch et all, User Profiles for Personalized Information Access, US: Information and Telecommunication Technology Center, 2010