Decision Support Systems 70 (2015) 60–72

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Online profiling and clustering of Facebook users

Jan-Willem van Dam, Michel van de Velden *

Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

ARTICLE INFO

Article history:
Received 20 November 2013
Received in revised form 22 September 2014
Accepted 1 December 2014
Available online 9 December 2014

Keywords:
Online profiling
Social networks
Customer relationship management
Correspondence analysis
Cluster analysis
Facebook

ABSTRACT

In a relatively short period of time, social media have acquired a prominent role in media and daily life. Although this development brought about several academic endeavors, the literature concerning the analysis of social media data to investigate one's customer base appears to be limited. In this paper, we show how data from the social network site Facebook can be operationalized to gain insight into the individuals connected to a company's Facebook site. In particular, we propose a data collection framework to obtain individual specific data and propose methodology to explore user profiles and identify segments based on these profiles. The proposed data collection framework can be used as an identification step in an analytical customer relationship management implementation that specifically focuses on potential customers. We illustrate our methodology by applying it to the Facebook page of an internationally well-known professional football (soccer) club. In our analysis, we identify four clusters of users that differ with respect to their indicated "liking" profiles.

1. Introduction

Social networks and their role played in daily life increased considerably over the last few years. As illustrated by the editorials of two recent special issues [3,8], recent academic publications cover a broad spectrum of topics related to social media. Some examples concern the potential of social media and its effect on customer loyalty [4]; how to use Facebook to activate customers in sharing product/service recommendations [7,16,19], the role of social networks, in particular Facebook, on intentional social actions [6]; the relationship between personal networks and patterns of Facebook usage [29]; the effectiveness of user generated content in stimulating sales [10,17,39]. This short list of topics and references is by no means exhaustive. It only serves to illustrate the recent interest and range of applications relating to firms and social networks. One common element among extant literature is that none of these studies build on directly observed individual level social network data. Instead, either aggregate data or focus groups and/or (online) surveys were employed in order to answer the research questions. This limitation was recently also observed by [31] In their study of the effect of social media participation on visit frequency and profitability, survey respondents were linked to their social media (i.e. Facebook) profiles by matching of names. In this paper we add to the existing literature by explicitly considering the retrieval and analysis of profile data directly obtained from social network sites. The proposed methodology can be implemented into an analytical customer relationship management (CRM) framework aimed at the analysis of customer characteristics that may help improve a firm's customer management strategies. Moreover, by focusing on data from public profiles from the social media platform Facebook, we are able to identify potential rather than actual customers. That is, in contrast to typical CRM implementations that rely on data directly obtained from customers, we consider a much broader group of individuals that indicated an interest in a firm even when an actual purchase has not yet been recorded.

The contribution of this paper is threefold: First, we show how Facebook users that "like" a firm can be identified. As also observed by [31] this is not a trivial task. Second, using the information volunteered by such Facebook users through their publicly available pages, we show how segments of Facebook users can be identified through data visualization and cluster analysis methods. Clustering of a firms' Facebook fans, may improve understanding of strategic segmentation of social media users connected to a firm [28] Moreover, the cluster results and visualizations can be used to improve targeting of marketing efforts. For example, a company may consider seeking cooperation with another brand or a popular media figure based on the popularity of such a brand or person with the (potential) customers. Moreover, such efforts could be targeted directly at specific groups of (potential) customers rather than at all (potential) customers. Third, we apply our methodology to a (large) data set of Facebook users that indicated liking a popular and successful international football club. This football club granted us administrator rights, under provision of not revealing the name of the club and any results indicative of the club's name. The results of our analysis show that, based on the Facebook users' liking behavior, clusters can be obtained. Given the differences between liking patterns

* Corresponding author at: Michel van de Velden, Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands.
E-mail addresses: jwvdam@gmail.com (J.-W. van Dam), vandevelden@ese.eur.nl (M. van de Velden).

http://dx.doi.org/10.1016/j.dss.2014.12.001
0167-9236/

in these clusters, differentiated marketing strategies for the different clusters can be developed.

The remainder of this paper is structured as follows. First, we briefly review previous research on analytical CRM and online profiling. Next, we briefly discuss specific data considerations for Facebook. We introduce some terminology and review Facebook's data analysis and programming facilities. In Section 4, we show how specific Facebook users can be identified. Next, we analyze the individual level data using a combined multiple correspondence analysis and k-means cluster analysis method. We show how results can be visualized and interpreted. The paper concludes with a discussion of our results, implications for research and practice, and future directions.

2. Customer relationship management and online profiling

Customer relationship management (CRM) has become widely recognized as an important business approach [27,31] defines CRM as an "enterprise approach to understanding and influencing customer behavior through meaningful communications in order to improve customer acquisition, customer retention, customer loyalty, and customer profitability". Hence, customer acquisition (or: identification) can be seen as the first step in a customer relationship management cycle that, together with retention and customer development form a complete cycle geared at creating a better understanding of (potential) customers in order to increase long term customer value to the firm [22,33,30].

Customer identification is typically based on information directly available to a firm [38] For example, customers may be required to provide certain background information upon purchasing a product. In addition, companies may ask customers to volunteer information by completing a survey or persuading them to join a loyalty program. Based on the available data, a customer profile, that is, a model of the customer, can be constructed. Based on such a customer profile, a marketeer decides on appropriate strategies and tactics to meet the specific needs of the consumer [32]. Hence, possessing accurate information about preferences and background characteristics of your (potential) customers makes it possible to improve targeting of, possibly individual specific, marketing efforts [25].

Obtaining direct customer information requires an existing relationship with the firm. That is, customers need to have either purchased a product or made contact with the firm in such a way that identification is possible so that additional information can be collected. In the case of yet unidentified potential customers, it is not possible to acquire data in this fashion. Moreover, except for the observable transactional data (i.e., purchase time and amounts, etc.) customers may decline to provide additional information.

Social media offer a new source of customer profile information. In particular, social media offer opportunities to identify *potential* customers. Through social media, individuals often express preferences for brands, products, services, persons, political parties, etc., in a free unsolicited way. Thus, if one is able to collect such information from potential customers of a certain firm, for example by focusing on users that indicated an interest in that firm, online profiles can be created that allow for better, individualized, targeting.

Although it has been suggested that the rise of new media requires novel approaches to successfully manage customer relationships [18], applications in which customer background data from social network sources are used to gain insight into customer backgrounds, appear to be under represented in the academic literature. There are some studies [24,23,15,5] in which social network data were used that contained personal information of the users in the data set. However, the goal of these studies was to study network ties [24], privacy issues [5,15], or relating the number of friends to the amount of information available on a person's Facebook page [23]. None of the studies used the data for online profiling: The collection of information from the Internet for the purpose of formulating a profile of users' habits and interests [37]. In this paper, we fill this gap by proposing a data collection framework for the purpose of online profiling.

Online profiling can be divided into two categories: reactive and non-reactive data collection [37]. *Non-reactive data collection* focuses on the collection of data concerning Web usage behavior, e.g., IP addresses of visitors, time spent on certain Web pages, and clicking behavior information. These data are used to gain insight into Web user behavior, and thus, characteristics of individual visitors or visitor groups. Non-reactive data form a large and potentially interesting source of online profile information. However, for the construction of online user profiles, the observed usage behavior must first be transformed into meaningful variables. The construction and definition of such variables is not always a easy task. In our study, we therefore primarily focus on the retrieval and analysis of online profiles based on reactive rather than non-reactive data.

Reactive data collection zooms in on visitor characteristics which cannot be collected through tracking Web usage behavior of a visitor. Instead, reactive information is collected by using forms and selection menus, which have to be filled in by visitors themselves. Reactive data requires little to no recoding of the original variables and they are immediately collected at the user level. Moreover, in the case of Facebook, providing reactive data requires very little effort from the users. For example, when joining Facebook, users are asked to provide certain personal background information (e.g., name, gender, date of birth). Users provide this information by selecting the appropriate options. This basic background information can be supplemented by more personal information concerning, for example, hobbies, relationship status etc. Finally, by "liking" other pages, personal preferences for persons or objects can be indicated.

The resulting online profiles can be of great value for marketeers, as they can be used to identify different (segments of) users (customers) that require different marketing approaches. Moreover, it enables the company to know its potential customers, that is, individuals that indicated a preference towards the product/brand by "liking" it on Facebook.

3. Facebook data

Facebook users put personal information on their Facebook page. Some examples are someone's name, gender, date of birth, e-mail address, sexual orientation, marital status, interests, hobbies, favorite sports team(s), favorite athlete(s), or favorite music. Furthermore, it is possible to specify your Facebook friends, post messages, publish pictures or other content. Consequently, the potential value to marketeers and researchers of the information available through Facebook is substantial. However, extracting the information is no trivial task as:

1. Facebook users are able to make certain information not publicly available and therefore not visible to non friends.
2. Facebook users are not obliged to fill in fields and therefore, many users do not specify all possible information about themselves.
3. The default statistics that Facebook offers for Facebook page administrators are limited.
4. It is not obvious how Facebook users who "like" your page can be identified.

The first two points are a result of the design and policy of Facebook.com and therefore we take these points as these are. Instead, we focus on the extraction of available data from Facebook and consider a user profile data collection framework taking into account the abovementioned issues.

The Facebook data collection framework that we propose consists of three steps: 1) identification of "fans" of the Facebook page, 2) retrieval of relevant data for the identified fans, and 3) preparation of the data. The first step of this framework requires administrator rights to the page, in the other steps public information from the relevant pages needs to be collected. Before we show how to implement the data