7-1-2013

The Value Of Users' Facebook Profile Data - Generating Product Recommendations For Online Social Shopping Sites

Jörg Gottschlich
Technische Universität Darmstadt, Darmstadt, Germany, gottschlich@emarkets.tu-darmstadt.de

Irina Heimbach
Technische Universität Darmstadt, Darmstadt, Germany, heimbach@emarkets.tu-darmstadt.de

Oliver Hinz
Technische Universität Darmstadt, Darmstadt, Germany, hinz@wi.tu-darmstadt.de

THE VALUE OF USERS' FACEBOOK PROFILE DATA – GENERATING PRODUCT RECOMMENDATIONS FOR ONLINE SOCIAL SHOPPING SITES

Gottschlich, Jörg[1], Technische Universität Darmstadt, Hochschulstr. 1, 64289 Darmstadt, Germany, gottschlich@emarkets.tu-darmstadt.de

Heimbach, Irina, Technische Universität Darmstadt, Hochschulstr. 1, 64289 Darmstadt, Germany, heimbach@emarkets.tu-darmstadt.de

Hinz, Oliver, Technische Universität Darmstadt, Hochschulstr. 1, 64289 Darmstadt, Germany, hinz@wi.tu-darmstadt.de

Abstract

Most online shops apply recommender systems, i.e. software agents that elicit the users' preferences and interests with the purpose to make product recommendations. Many of these systems suffer from the new user cold start problem which occurs when no transaction history is available for the particular new prospective buyer. External data from social networking sites, like Facebook, seem promising to overcome this problem. In this paper, we evaluate the value of Facebook profile data to create meaningful product recommendations. We find based on the outcomes of an experiment that already simple approaches and plain profile data matching yield significant better recommendations than a pure random draw from the product data base. However, the best recommendations are based on music/video, brands and product/category information that need to be extracted from the Facebook profile with more sophisticated approaches.

Keywords: Online Social Networking Sites, User Profile, new user cold start problem, Social Shopping, Facebook

[1] Corresponding author

1 Introduction

Nowadays most online shops apply *recommender systems*, i.e. software agents that elicit the users' preferences and interests with the purpose to make product recommendations (Xiao and Benbasat, 2007). Recommender systems foster add-on and cross-sales and have impact on sales diversity (Hinz and Eckert, 2010). Recommender systems can vary in the system's input, the data representation and the recommendation approach (Huang *et al.*, 2004). Most recommendation systems use past transactional data (e.g. on products and the user) to derive product recommendations (Adomavicius and Tuzhilin, 2005). These systems, however, usually suffer from a "cold start problem", i.e. it is difficult to make recommendations for new users where no transactional data is yet available (Huang *et al.*, 2004). Previous research proposed several solutions to this problem. First, a new user might get non-personalized recommendations built on top-seller rankings (Schafer *et al.*, 2001). Jannach et al. (2010) suggest explicitly asking new users for product ratings. It might be also possible to apply user's transactional data on-the-fly, such as online shop navigation history (Huang *et al.*, 2004). Adomavicius and Tuzhilin (2005) suggest using external information to build profiles of new users. A decade ago this might have been an exotic approach, but nowadays with the emergence of technologies which allow users to create and maintain online profiles, recommender designers have access to a huge body of user data.

Especially for the emerging field of social shopping sites the integration of external user data might create a promising opportunity to generate targeted product recommendations for new users. We define *social shopping sites* as online shops which integrate external online social networking sites like Facebook or offer their own features allowing users to build profiles, maintain their social relations (e.g. friendships), post their purchases on their walls or let friends evaluate their purchases. Well-known social shopping sites are caboodle.com and thisnext.com.

With respect to recommendation systems, social shopping sites face per se the same cold start problem as conventional online shops. It might be even worse if the social shopping site is an intermediary who does not offer a product portfolio himself, but provides a market place for a high number of sellers. To this business model the cold start problem is immanent: when buyers navigate to the partner stores and are making their deals there, the market platform cannot observe the transaction and hence can never build a detailed purchase history about its community members. However, social shopping sites have access to additional social information about the user. Based on such data, it might be possible to generate targeted product recommendations. To the best of our knowledge, no study has proposed and evaluated such an approach based on social data yet.

Because of the diversity and the vast amount of users' data on social networking sites, it is unclear which information allows generating valuable product recommendations. The purpose of this study is therefore to evaluate, what kind of data on a user's social networking site profile serve as a good base for product recommendations at a social shopping site. We build a recommender system which applies user's Facebook profile data to generate product recommendations. Together with the social shopping site of the world's largest mail order company (anonymous for confidentiality reasons), we conducted a small field experiment, asking participants to evaluate product recommendations generated on base of their Facebook profiles.

The remainder of this paper is structured as follows: In section 2 we summarize previous research on recommender systems and describe the new user cold start problem. Then we present the data, i.e. the parts of the Facebook profile we used and the product database. Section 4 describes our empirical study and the results in detail. Finally, we summarize our research and discuss future research venues.

2 New user cold start problem and the value of user factual data

When a new user visits an online shop, it is important to generate reliable recommendations and build good user profiles from the very beginning (Montaner *et al.*, 2003) to increase the user's perceived usefulness of and trust in the recommendation system (Xiao and Benbasat, 2007). The recommender systems research community proposes several approaches to deal with the new user cold start problem. Table 1 summarizes these approaches suggested by previous research. These approaches can be distinguished by means of three criteria. First, approaches are distinguished by the users' involvement: *implicit feedback* is generated with little user effort, and *explicit feedback* needs an active evaluation of a proposed product recommendation by the user (Huang *et al.*, 2004). Further, the used data can be *external,* obtained from outside the system, or *internal,* generated within the system. Finally, recommendations for new users can be *personalized,* i.e. customized for each user, and *non-personalized,* i.e. equal for all users.

One approach to solve the cold start problem is to offer non-personalized recommendations based on top-sellers list or editors advices (Schafer *et al.*, 2001). This approach might be insufficient for small start-up online shops, which have the problem of sparsely rated products. Another approach is to ask new users to evaluate a selected set of products (Jannach *et al.*, 2010). In this case the shop operators need to determine the set of products, which promise the highest information gain. Finally, it might helpful to explicitly ask users for their interests and preferences. However, the users are often not willing to provide such information (Montaner *et al.*, 2003) and thus customer data recorded with the registration form (like name, address, age, sex) may be used to classify a new user to some stereotype to generate initial product recommendations (Montaner *et al.*, 2003).

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">Type of User Feedback</td></tr>
<tr><td>Explicit</td><td>Implicit</td></tr>
<tr><td rowspan="2">Data Source</td><td>External data</td><td>Personalized:
• Raffles, Competitions
Non-Personalized:
• Public Statistics
• Market Research</td><td>Personalized:
• User data from social networking platforms</td></tr>
<tr><td>Internal data</td><td>Personalized:
• Register form
• Asking for interests and preferences
Non-Personalized:
• Explicit rating of a selected set of products</td><td>Personalized:
• Online shop navigation history
Non-Personalized:
• Top-seller lists
• Editors' advices</td></tr>
</table>

Table 1. A taxonomy of approaches for solving the new user cold start problem

Social networking platforms offer access to users' demographics (similar to the data gained by the register form) and users' interests and preferences (shared content, favorites etc.). The advantages of such data are, first, rather rich, complete and up-to-date data and second, users provide information about themselves voluntarily. Unfortunately, the richness of such data causes new challenges: Which kind of user data allows generating the most reliable product recommendations? This study evaluates the different data sources available on social networking platforms with respect to their usefulness in generating product recommendations.

3 Data

3.1 Facebook profile data

Facebook stores an extensive amount of data about each member. A complete list can be found at (Facebook.com, 2012). The availability and extent of the profile data depends on the user's attitude towards entering and making the information visible in his profile. Hence, we focus on a core set and base the product recommendations on the following subset of Facebook profile data:

- Date of Birth (or Age, respectively)
- Gender
- Likes: Whenever a user clicks on the "Like" button of a Facebook object (e.g. a fan page), that item is stored in the user's profile and are categorized into the following categories (if applicable): Music, Movies, Television, Activities, Books, Games, Athletes, Teams, Sports, Others, Admired people
- Groups: Membership of a user in a Facebook group
- Geodata: Hometown or Current City
- Posts: Status updates posted to his/her wall by the user (free-text)

Demographic data, such as Gender and Age, have a long tradition in marketing research and segmentation (Beane and Ennis, 1987; e.g. Zeithaml, 1985) and can be used to explain differences in adoption behavior (Aral and Walker, 2012), even though they might often be just substitutes for other latent factors (Fennell *et al.*, 2003). Nevertheless, as they are readily available and deliver a kind of an outer bound for user characterization, they should be included.

Certainly, Likes are the first part of the profile data which comes to mind when thinking about user preferences as they explicitly express affinity. The same is true for membership in groups. Geodata enables to handle location-related affinities. Posts, as the least structured data available due to their free-text nature, can however provide a broad insight into the user's daily life, habits and wishes, but also requires some effort to be made available for product recommendations.

3.2 Products

We aim to use the available social data to recommend products from the product database of our business partner. The product database contains a total of 1,942,857 products in the categories Lifestyle (425,334), Fashion (954,609), Habitation (390,691) and Not Specified (172,223). Products range from physical products to services like vouchers for events. Approximately, one third of the selected products have photos attached. Table 2 lists the fields included in the product data.

Field name	Description
ID	Product Identifier
Title	Name of the product
Description	Textual description
Gender/Age	Textual description of Gender/Age target group (e.g. Girl, Boy, Women etc.)
Categories	Hierarchical Product category (e.g. Lifestyle, Fashion and subcategories)
Brand	Manufacturer Brand name
Image URL	URL of product image (if any)

Table 2. Product data fields

4 Experiment

4.1 Product recommendation approaches

Considering the task of finding the right products for a user, the question is how we can exploit the data sources available from Facebook to arrive at an effective recommendation set. Having the data about the user is only the first step; the data's true value depends on the way how it is applied to derive product recommendations. Therefore, in our experiment, we consider different approaches of profile data application to arrive at a meaningful assessment of profile data value.

A plain and simple approach of matching products to profile entries is the direct search of profile characteristics in the product data. In the context of this paper, direct matching means that the comparison between product and profile is conducted without additional knowledge or further interpretation of the data's specific context. For example, a direct matching using "Likes" data simply tries to find keywords from the profile's "Likes" information in the product data. In detail, we implement the following approaches:

- **Likes**: Likes in a Facebook profile are categorized (see section 3.1) and hence provide additional Meta information about the kind of subject that is liked. Hence, it is reasonable, to try and find products whose description match terms that are "liked" by the user.
- **Gender/Age**: Mapping the gender and age taken from the Facebook profile to products matching the appropriate segment (like girls, boys, women, men) which can in turn be used to filter products, especially exclude those *not* matching well (e.g. girls' toys like Barbie for a male adult).
- **Groups**: By a membership in a Facebook group, users express their association to a certain topic which might be used to identify products of interest. We use the group name to search for matching products.
- **Geodata**: Some product offerings, e.g. event vouchers or souvenir articles, are location-specific. The hometown can be used to find products that match the user's home location and enable to offer products based on regional affinity.

We refrained from doing a direct matching on Posts, as the context and notion of keywords in Posts is not unambiguous per se.

More advanced approaches take the specific semantics of data fields into account, i.e. they interpret the contents and use a deeper understanding of the characteristics of product and profile data to make matches more meaningful than a simple keyword match. The specific approaches we examine here are:

- **Brand matching**: Brands offer a strong identification potential for consumers, especially in terms of communicating preferences and values (e.g. Ahearne *et al.*, 2005). Within Facebook, the brands a user relates to, can be "liked" (given the brand operates a Facebook page) or it can appear in Posts or Groups. As our product data also provides brand names along with the products, a match based on brand names is both promising and feasible. Especially as demographic data has its weaknesses in identifying brand preference (Fennell *et al.*, 2003, p. 242), this approach might overcome this shortcoming.
- **Product category matching**: Product categories provide an abstract description of the particular articles. As Facebook enables users to "like" generic terms as well, such as activities, a matching for product groups with the users profile entries enables to use those abstract information for matching (e.g. if a user likes "Hiking" and there is a product category "Fashion|Sports|Hiking", it is likely that the user will appreciate products belonging to this category).
- **Video/Music title matching**: Video/Movie/TV show or Music group names can be misleading when they are split into keywords which are then tried to match to a product (e.g. "House, MD", the TV show about a misanthropic, but ingenious doctor, is a definitive concept while a search for "house" would most likely result in many unrelated products being found). As Facebook offers specific categories (see section 3.1) to show affiliation with these kinds of products, it makes sense to treat their title as an atomic term when it comes to product matching.

Table 3 shows an overview of the different approaches and the Facebook data fields involved.

Filter Approach		**Facebook profile data involved**				
		Gender/Age	Likes	Groups	Hometown	Posts
Plain	Keyword match	X	X	X	X	
Specific	Brands		X	X		X
	Product Category Matching		X	X		X
	Video/Music		X (selection)			

Table 3. Overview of profile data used in recommendation approaches

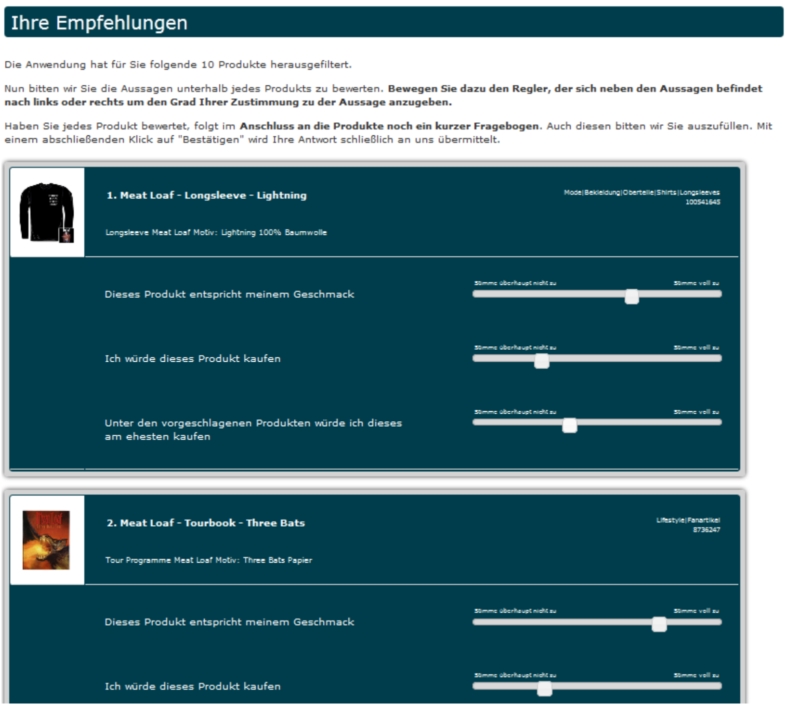

Figure 1. Screenshot of the test system

4.2 Experiment

We conduct a combined between- and within-subject design experiment (c.f. Hinz *et al.*, 2011). This means every subject receives ten product recommendations. A subject either receives recommendations based on a random draw or on a combination of the aforementioned approaches. The evaluation of the subjects who received the random selection serve as a benchmark. As the frontend interface presents all recommendations identically, the effect on the dependent variable can be attributed to the recommendation approach and the used data only. Such experiments allow an identification of causal effects which is often difficult e. g. with transactional data from the field.

The course of the experiments is as follows for a subject: The subject begins on a landing page with short instructions and starts the experiment by clicking on a start button. The subject then has to login to Facebook and authorize the Facebook application created for this experiment. If the authorization process is successful, the subject is returned to the experiment site. To create the product recommendations, the test-system calls the recommender subsystem (or the random generator) providing the Facebook access token to enable the Facebook profile access. The subsystem then returns a ranked set of ten product ids and the frontend provides these as recommendations to the subject (see Figure 1) in the specified order (best recommendation first). Each product is displayed with a title, a description, a category label, a product id and, if available, a picture. As our goal is to measure how well the products match the user's preferences, we do not provide information on prices to avoid a bias caused by different price levels.

Along with the products, a questionnaire asks the user to evaluate the product selection. For each product, the user rates whether the product meets the subject's taste (item from McAlexander *et al.*, 2002), as well as the propensity to purchase (Pereira, 2000) the product on a scale of "strongly disagree" to "strongly agree" by moving a slider. The slider's range was internally translated to a scale of 1 to 100. For control purposes, we also asked for the participant's age, gender, the experience with online shopping and their Internet usage behavior. After submission, the questionnaire cannot be changed anymore.

Every Facebook profile could participate only once in the test. To a large extent, participants were recruited via Facebook by sending the URL for online participation. As an incentive, we raffled Amazon vouchers among completed questionnaires. We conducted the experiment in July and August 2012. When participants visited the site, a short text explained the course of the experiment and then we asked the subjects to log in to their Facebook account to authorize the application. Subsequently, they received product recommendations they were expected to evaluate using slider inputs next to the product recommendations (cf. Figure 1).

Three participants denied the authorization of the Facebook application explicitly (not counting those who just dropped out and closed the window without explicitly denying; this number cannot be determined with certainty). Over the course of the experiment, we collected 86 completed questionnaires rating 860 product recommendations (comprising 788 different products due to multiple selections of some of the products). 58 of the respondents were male, 28 were female. The mean age was 27. Over 95% of the respondents use the Internet every day. Table 4 shows the descriptive statistics of the respondents' evaluations of the recommended products.

Every recommendation was made based on one of the approaches introduced in section 4.2 and thus makes use of specific Facebook data. In the following we will examine which type of Facebook data led to the most successful recommendation. This analysis will yield first insights on the value of Facebook data for recommendation systems.

Recommendation Source	Number of Recommen-dations	Taste		Purchase Propensity	
		Mean	Std. Dev.	Mean	Std. Dev.
Demographics	86	35.17	29.28	29.09	25.98
Groups	3	23.67	24.70	11.67	9.71
Likes	255	33.20	32.50	24.84	30.84
Geodata	5	20.20	41.82	20.80	42.05
Product Category	98	36.96	31.51	29.22	30.71
Brands	62	39.08	35.54	30.16	30.57
Video/Music	51	45.78	34.64	33.71	31.86
Random	300	23.07	27.25	19.52	26.64
TOTAL	**860**	**31.35**	**31.38**	**24.75**	**29.27**

Table 4. Descriptive statistics of the product recommendation evaluation. Taste and Purchase Propensity are measured on a 100-point-Likert scale.

4.3 Model and Results

As quality metric for the recommendation we measure how good a recommendation meets the subject's taste (McAlexander *et al.*, 2002) and the propensity to purchase (Pereira, 2000) , which also constitute our dependent variables. As independent variables we introduce dummy variables for the different types of Facebook data used, e.g. *Likes_D* is 1 if the recommendation is based on Likes data from Facebook, 0 otherwise or *Demographics_D* is 1 if the recommendation is based on demographic information like gender and age and 0 otherwise.

We further include demographic covariates and a dummy variable whether the recommended product included a picture because previous research found that pictures can have a significant influence on economic decisions (Dewally and Ederington, 2006). We also categorized all recommended products in search goods and experience goods manually and use this information as control variable.

Equation (1) and (2) summarize our models where *i* indicates the subject and *j* indicates the *j*-th recommendation for subject *i*:

(1) $taste_{i,j}=\beta_1+\beta_2\cdot Demographics\_D_{i,j}+\beta_3\cdot Groups\_D_{i,j}+\beta_4\cdot Brands\_D_{i,j}+\beta_5\cdot Likes\_D_{i,j}+\beta_6\cdot GeoData\_D_{i,j}+\beta_7\cdot ProductCategory\_D_{i,j}+\beta_8\cdot VideoMusic\_D_{i,j}+\beta_9\cdot Picture\_D_{i,j}+\beta_{10}\cdot Gender_i+\beta_{11}\cdot Age_i+\varepsilon_{i,j}$

(2) $purchase\_propensity_{i,j}=\beta_1+\beta_2\cdot Demographics\_D_{i,j}+\beta_3\cdot Groups\_D_{i,j}+\beta_4\cdot Brands\_D_{i,j}+\beta_5\cdot Likes\_D_{i,j}+\beta_6\cdot GeoData\_D_{i,j}+\beta_7\cdot ProductCategory\_D_{i,j}+\beta_8\cdot VideoMusic\_D_{i,j}+\beta_9\cdot Picture\_D_{i,j}+\beta_{10}\cdot Gender_i+\beta_{11}\cdot Age_i+\varepsilon_{i,j}$

The estimates for β_2-β_8 thus reflect the difference to the benchmark, e.g. a product that was randomly drawn from the product base. We pre-analyzed our data and the White-Test ($p<.05$) indicates heteroscedasticity and we therefore estimate our models with heteroscedasticity-consistent standard errors. The variance inflation factors (VIF) are well below 4 (mean VIF = 1.13, max VIF=1.27) and thus multicollinearity does not seem to be a problem in our dataset. We arrive at the results for equation (1) summarized in Table 5.

	Coef.	Robust Std. Err.	T	P>|t|
Constant (***)	31.69	5.46	5.80	.000
Demographics_D (***)	12.49	3.61	3.46	.001
Groups_D	1.19	12.18	0.10	.922
Likes_D (***)	10.40	2.57	4.04	.000
GeoData_D	.35	18.30	0.02	.985
ProductCategory_D (***)	14.51	3.58	4.05	.000
Brands_D (***)	16.30	4.79	3.40	.001
VideoMusic_D (***)	21.29	5.46	3.90	.000
Age (**)	-.40	.20	-1.97	.049
Gender_D (0:male/1:female)	1.87	2.26	.83	.408
Picture_D	1.03	2.24	.46	.645
ExperienceGood_D (0:no/1:yes)	1.28	1.43	.89	.373
F = 4.50 (p<.01)				

Table 5. Impact of Facebook Data on Meeting the Subject's Taste.

The F-value for all models allow us to reject the null hypothesis that the sets of coefficients are jointly zero (p<.01). Interestingly we find that Facebook data used for recommendations can significantly improve the recommendation quality. The most valuable data in our context is utilized by a specific understanding of Facebook information on music, film and TV shows. Recommendations based on this information yield a +21.29 higher score on a 100-points-Likert scale (p<.01). A similar improvement can be made if recommendations are based on a semantic understanding of product categories that might be interesting for the subject (p<.01). Recommendations based on this deeper understanding of product categories receive +14.51 points on the used scale. A specific approach is also useful to identify brands and make recommendations based on this information. This yields +16.3 points with respect to the recommendation quality (p<.01). But even simpler approaches that use direct matching can already lead to better recommendation than a random selection, e.g. the use of demographic data (p<.01) or Likes data (p<.01).

We further find that it is harder to meet the preferences of older subjects (-.40, p<.05). This can result either from poorer data quality as older users tend to spend less time on their Facebook activities (Joinson, 2008, p. 1032) or results from latent differences captured by age or it could be caused by the product base that might fit younger people better.

Making recommendations that meet the subjects' preferences is however only an antecedent of the propensity to purchase which we analyze by estimating equation (2). Table 6 summarizes the estimates.

	Coef.	Robust Std. Err.	T	P>|t|
Constant (***)	16.83	5.54	3.04	0.002
Demographics_D (**)	8.95	3.31	2.71	0.007
Groups_D (*)	-6.95	5.07	-1.37	0.171
Likes_D	5.11	2.48	2.06	0.039
GeoData_D	-2.08	16.46	-0.13	0.899
ProductCategory_D (**)	10.12	3.49	2.90	0.004
Brands_D (**)	10.53	4.19	2.52	0.012
VideoMusic_D (**)	12.49	4.99	2.50	0.013
Age	0.05	0.21	0.23	0.821
Gender_D (0:male/1:female)	1.90	2.19	0.87	0.386
Picture_D	0.70	2.14	0.33	0.742
ExperienceGood_D (0:no/1:yes)	1.14	1.34	0.85	0.396
F = 2.36 (p<.01)				

Table 6. *Impact of Facebook Data on Propensity to Purchase* (Pereira, 2000)

First, we observe a smaller impact of the data used with respect to purchase propensity which is not surprising: Products that meet the subjects' preferences do not necessarily convert to a purchase. This result provides some face validity.

Second, we find again that the recommendations that semantically interpret Facebook data lead to significantly better recommendations, e.g. understanding the brands preferences out of Facebook data and using this information to make recommendations, lead to an increase of +10.53 points (p<.05) on the 100-Likert scale with respect to purchase propensity. Similarly a semantic understanding on Video/Music and product categories seem to be valuable (p<.05). Data from this category can increase the purchase propensity by more than +74% (=12.49/16.83) which is a huge improvement and therefore is a promising solution for the recommender cold start problems.

Simpler approaches like the use of demographics can also lead to better recommendations and a higher propensity to purchase (+8.95, p<.01), but they can also lead to consequential misinterpretations and ultimately inappropriate recommendations. We observe in our dataset that the simple use of Groups data can lead to recommendations with significantly lower propensity to purchase (p<.1). Therefore, we recommend being very careful when using this information right away without trying to get a semantic understanding of its meaning.

Although the Likes data can be useful to recommend products that match the subjects' preferences, this information only delivers a small surplus to the purchase propensity. This is an interesting finding, as the Like data explicitly state the subject's preferences and hence would be expected to yield higher approval.

With respect to control variables we do not find a significant impact. The purchase propensity for experience goods seems to be slightly higher as is the purchase propensity of female subjects, but these findings are not statistical significant (p>.1).

5 Conclusion

As some online retailers do not have access to transaction histories and all retailers face the problem of a cold start problem, the exploitation of Facebook data to gain some first insights on the prospective buyer seems promising. We therefore conducted an experiment to assess the value of Facebook data for product recommendations. By conducting an experiment we were able to determine the causal effect of the Facebook base data for recommendation quality.

Interestingly, we find that the data in the Facebook profile is of value for product recommendation. Even very simple approaches like direct matching keywords from Facebook profiles with the product database lead to recommendations that match the prospective buyer's preferences significantly better than neglecting this information.

However, these simple approaches can also lead to misunderstanding and ultimately to recommendations that are totally inappropriate. We find for example that using the Groups data from Facebook and match keywords with the product database can lead to significantly worse recommendation than a pure random draw. This finding emphasizes that developers and business practice have to pay attention when they want to make use of Facebook data.

Approaches that try to interpret the data semantically and try to understand the specific meaning of the Facebook data seem to be very promising. We find that such information can increase the purchase propensity by more than +74% which is a huge improvement and, therefore, seems to be a promising solution for typical recommender cold start problems.

Our study does not come without limitations: First, the entry decision in our context is not totally comparable to the real decision. In our experiment, the subjects indeed participated voluntarily but may refrain from using the system in a real setting. It is therefore not clear whether this setting led to a self-selection bias and we cannot make any conclusions whether such systems would be accepted by prospective users. The social shopping site that provided the data, however, offers a similar Facebook app and found a substantial numbers of users in the market.

Second, beside the data quality the quality of the algorithms also impacts the recommendation quality. Therefore, we recommend being careful when looking at the magnitude of the particular coefficients. We are, however, confident that the sign and significance of the estimated coefficients are reliable evidences for the value of Facebook data for product recommendations. Future research could try to improve the algorithms that semantically interpret the data which should ultimately lead to even better recommendations.

This study contributes to research on recommender systems as our results can be used as basis for designing recommenders which use external data, especially from social networking sites. In addition, this study delivers starting points for developing alternative approaches for solving the cold start problem using external user data. Finally, we evaluate different sources of user data with respect to their usefulness for deriving product recommendations. To best of our knowledge, it is the first attempt to systematically evaluate the usefulness of Facebook profile data for product recommendations.

References

Adomavicius, G. and Tuzhilin, A. (2005). Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions, *Knowledge and Data Engineering, IEEE Transactions on* 17(6): 734 – 749.

Ahearne, M., Bhattacharya, C. B. and Gruen, T. (2005). Antecedents and consequences of customer-company identification: expanding the role of relationship marketing., *The Journal of applied psychology* 90(3): 574–85.

Aral, S. and Walker, D. (2012). Identifying influential and susceptible members of social networks., *Science (New York, N.Y.)* 337(6092): 337–41.

Beane, T. P. and Ennis, D. M. (1987). Market Segmentation: A Review, *European Journal of Marketing* 21(5): 20–42.

Dewally, M. and Ederington, L. (2006). Reputation, Certification, Warranties, and Information as Remedies for Seller-Buyer Information Asymmetries: Lessons from the Online Comic Book Market, *The Journal of Business* 79(2): 693–729.

Facebook.com. (2012). Accessing Your Facebook Info | Facebook Help Center. Retrieved December 5, 2012, from https://www.facebook.com/help/405183566203254/

Fennell, G., Allenby, G. M., Yang, S. and Edwards, Y. (2003). The Effectiveness of Demographic and Psychographic Variables for Explaining Brand and Product Category Use, *Quantitative Marketing and Economics* 1(2): 223–244.

Hinz, O. and Eckert, J. (2010). The Impact of Search and Recommendation Systems on Sales in Electronic Commerce, *Business & Information Systems Engineering* 2(2): 67–77.

Hinz, O., Hann, I. and Spann, M. (2011). Price discrimination in E-commerce? An examination of dynamic pricing in name-your-own price markets, *MIS Quarterly* 35(1): 81–98.

Huang, Z., Chung, W. and Chen, H. (2004). A graph model for E-commerce recommender systems, *Journal of the American Society for Information Science and Technology* 55(3): 259–274.

Jannach, D., Zanker, M., Felfernig, A. and Friedrich, G. (2010). *Recommender systems: an introduction*, Cambridge University Press.

Joinson, A. (2008). Looking at, looking up or keeping up with people?: motives and use of facebook, In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1027–1036.

McAlexander, J., Schouten, J. and Koenig, H. (2002). Building brand community, *The Journal of Marketing* 66(1): 38–54.

Montaner, M., López, B. and Rosa, J. D. La. (2003). A taxonomy of recommender agents on the internet, *Artificial Intelligence Review* 19: 285–330.

Pereira, R. (2000). Optimizing human-computer interaction for the electronic commerce environment, *Journal of Electronic Commerce Research* 1(1): 23–44.

Schafer, J., Konstan, J. and Riedl, J. (2001). E-commerce recommendation applications, *Data mining and knowledge discovery* 5: 115–153.

Xiao, B. and Benbasat, I. (2007). E-commerce product recommendation agents: Use, characteristics, and impact, *Mis Quarterly* 31(1): 137–209.

Zeithaml, V. A. (1985). The New Demographics and Market Fragmentation, *Journal of Marketing* 49(3): 64–75.