Citation: Sundsøy, Pål, Johannes Bjelland, Asif M. Iqbal, Alex "Sandy" Pentland, and Yves-Alexandre de Montjoye. "Big Data-Driven Marketing: How Machine Learning Outperforms Marketers' Gut-Feeling." Social Computing, Behavioral-Cultural Modeling and Prediction (2014): 367–374.

As Published: http://dx.doi.org/10.1007/978-3-319-05579-4_45

Publisher: Springer-Verlag Berlin Heidelberg

Persistent URL: http://hdl.handle.net/1721.1/92459

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

MIT

Massachusetts Institute of Technology

Big Data-Driven Marketing: How machine learning outperforms marketers' gut-feeling

Pål Sundsøy[1], Johannes Bjelland[1], Asif M Iqbal[1],
Alex Sandy Pentland[2], Yves-Alexandre de Montjoye[2]

[1]Telenor Research, Norway
{pal-roe.sundsoy,johannes.bjelland,asif.iqbal}@telenor.com
[2]Massachusetts Institute of Technology, The Media Laboratory, USA
{pentland,yva}@mit.edu

Abstract. This paper shows how big data can be experimentally used at large scale for marketing purposes at a mobile network operator. We present results from a large-scale experiment in a MNO in Asia where we use machine learning to segment customers for text-based marketing. This leads to conversion rates far superior to the current best marketing practices within MNOs.

Using metadata and social network analysis, we created new metrics to identify customers that are the most likely to convert into mobile internet users. These metrics falls into three categories: discretionary income, timing, and social learning. Using historical data, a machine learning prediction model is then trained, validated, and used to select a treatment group. Experimental results with 250 000 customers show a 13 times better conversion-rate compared to the control group. The control group is selected using the current best practice marketing. The model also shows very good properties in the longer term, as 98% of the converted customers in the treatment group renew their mobile internet packages after the campaign, compared to 37% in the control group. These results show that data-driven marketing can significantly improve conversion rates over current best-practice marketing strategies.

Keywords: Marketing, Big Data, Machine learning, social network analysis, Metadata, Asia, Mobile Network Operator, Carrier

1   Introduction

For many people in Asia, mobile phones are the only gateway to the Internet. While many people have internet capable phones, they are often not aware of their capabilities. The overall penetration of internet in these countries is very small which causes a large digital discrepancy [1]. In the market of this study, internet penetration is less than 10%.

Mobile Network Operators (MNOs) commonly use texts as a way to raise customer awareness of new products and services - and to communicate with their customers. In Asian markets, MNOs typically run thousands of text campaigns a year, resulting in customers receiving several promotional texts per month. Making sure they are not seen as spammers by customers but rather as providing useful information is a major concern for MNOs. For this particular operator, the policy is to

not send more than one text per customer every 14 days. This limit is currently maxed and is preventing new campaigns from being launched.

Targeting, deciding which offer to send to which customer, often relies on the marketing team's "gut-feeling" of what the right audience is for this campaign. In a recent IBM study, 80% of marketers report making such decisions based on their "gut-feeling" [2]. A data-driven approach might lead to an increased efficiency of text-based campaigns by being one step closer to delivering "the right offer to the right customer". For example, previous research showed that data-driven approaches can reliably predict mobile phone and Facebook user's personality [3-5,18], sexual orientation [6], or romantic relationships [7].

Our data-driven approach will be evaluated against the MNO's current best-practice in a large-scale "internet data" experiment [8,9]. This experiment will compare the conversion rates of the treatment and control groups after one promotional text.

We show that this data-driven approach using machine learning and social network analysis leads to higher conversation rates than best-practice marketing approach. We also show that historical natural adoption data can be used to train models when campaign response data is unavailable

2   Best practice

The current best practice in MNOs relies on the marketing team's experience to decide which customers should receive a text for a specific campaign. The marketing team typically selects customers using a few simple metrics directly computed from metadata such as call sent, call received, average top-up, etc. For this particular "internet data" campaign, the marketing team recommended to use the number of text sent and received per month, a high average revenue per user (ARPU) [10], and to focus on prepaid customers. Table 1 shows the variables used to select the control group, the customers that, according to the marketing team, are the most likely to convert.

The control group is composed of 50 000 customers selected at random amongst the selected group.

Table 1: Variables used to select the control group

Sending at least four text per month
Receiving at least four text per month
Using a data-enabled handset
'Accidental data usage' (less than 50kb per month)
Customer in medium to high ARPU segment (spending at least 3.5 USD per month)

3   Data-driven approach

3.1   Features

For each subscriber in the experiment, we derive over 350 features from metadata, subscription data, and the use of value added services. Earlier work show the existence of influence between peers [11,12] and that social influence plays an important role in product adoption [13,14]. We thus inferred a social graph between customers to compute new features. We only considered ties where customers interacted more than 3 times every month. The strength of ties in this social graph is a weighted sum of calls and texts over a 2-month period. Using this social graph, we computed around 40 social features. These include the percentage of neighbors that have already adopted mobile internet, the number of mobile data users among your closest neighbors (ranked by tie strength), or the total and average volume of data used by neighbors.

3.2   Model

We develop and train the model using 6 months of metadata. As the outcomes of previous mobile internet campaigns were not stored, we train our model using natural adopters. We then compare these natural adopters, people who just started using mobile internet, to people who over the same period of time did not use mobile internet. Our goal is to identify the behavior of customers who 1) might be interested in using internet and who 2) would keep using mobile internet afterwards. We then select 50 000 natural adopters and 100 000 non-internet users at random out of these groups. Note that natural converters are only a way for us to extract characteristics of customers who are likely to convert. The conversion rates are likely to have been better if we had access to data about previous campaigns and previously persuaded adopters.

Table 2. Training set

Sample size	Classifier	Definition
50k	Natural adopters	Less than 50KB of data per month from December to March (accidental data usage). More than 1MB of data per month in April and May
100k	Reference users not using internet	No internet usage

We tested several modeling algorithms such as support vector machine and neural networks to classify natural converters. The final model is a bootstrap aggregated (bagging) decision tree [15] where performance is measured by accuracy and stability. The final model is a trade-off between accuracy and stability where, based on earlier experience, we put more weight on stability. The accuracy of the final model is slightly lower than other considered models. The bagging decision tree however turned out to be more stable when tested across different samples.
The final cross-validated model only relies on a few key variables. Only 20 features out of the initial 350 where selected for the final model. Table 3 shows the top 10 most useful features to classify natural converters as ranked by the IBM SPSS

modeler data mining software. The features tagged as binned are handled by the software optimal binning feature.

Table 3. Top 10 most useful features to classify natural converters. Ranked by importance in the model.

Rank	Type	Description
1	Social learning	Total spending on data among close social graph neighbors
2	Discretionary income	Average monthly spending on text (binned)
3	Discretionary income	Average monthly number of text sent (binned)
4	Discretionary income	Average monthly spending on value added services over text (binned)
5	Social Learning	Average monthly spending on data among social graph neighbors
6		Data enabled handset according to IMEI (Yes/No)
7	Social Learning	Data volume among social graph neighbors
8	Social Learning	Data volume among close social graph neighbors
9	Timing	Most used handset has changed since last month
10		Amount of 'accidental' data usage

3.3 Out-of sample validation

Before running the experiment, we validated our model on natural adopters in a new, previously unseen, sample using other customers and another time period. The performance on historical data is measured using lift curves. Fig. 1 shows a lift of around 3 among the 20% highest scored customers. This means that if we were to select the 20% highest scored customers, the model would predict 3 times better than selecting at random from the sample.
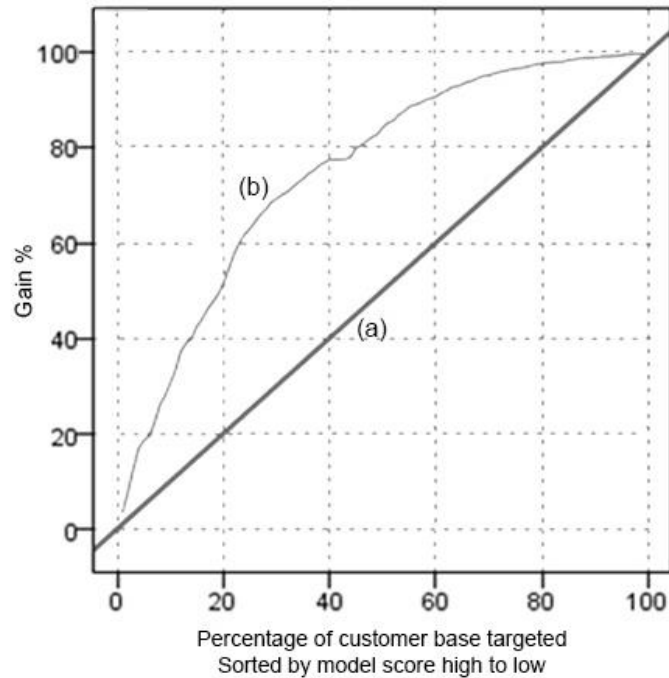
Fig 1. Out-of-sample validation of the model using lift curves. **(a)** is the gain if customer were selected at random while **(b)** shows selection by using the model.

We then select the treatment group using our model. We let the marketing team pick their best possible (control) group first and then specifically excluded them when selecting our treatment group. The treatment group is composed of the top 200 000 customers with the highest score. This represents approximately well under 1% of the total customer base.

4   Experiment

A large-scale experiment is then run to compare our data driven approach to the current best-practice in MNOs. The approaches will be compared using the conversion rates of the control and treatment group.

In this experiment, the selected customers receive a text saying that they can activate a 15MB bundle of data usage for half of the usual price. The 15MB have a limited validity and are only valid for 15 days. The customer can activate this offer by sending a text with a code to a short-number. This is a common type of campaign and is often used in this market. The text contains information about the offer and instructions on how to activate it.

The conversion rates between treatment and control group were striking. The conversion rate in the treatment group selected by our model is 6,42% while the conversion rate of the control group selected using the best-practice approach is only 0.5%, as shown in Fig. 2. The difference is highly significant (p-value < 2.2e-16). Our

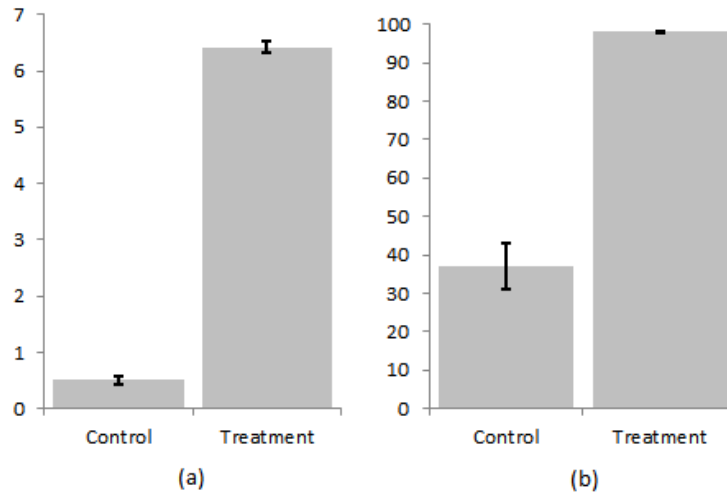data-driven approach leads to a conversion rate 13 times larger than the best-practice approach.

Fig 2. (a) Conversion rate in the control (best practice) and treatment (data-driven approach) groups. **(b)** the percentage of converted people who renewed their data plan after using the volume included in the campaign offer. Error bars are the 95% confidence interval on the mean using the blaker method

The goal of this campaign is not only to have customers take this offer but also to have them renew it after the trial period. We thus compare the renewal rate, customers buying another data plan after using the half-priced package, between the converted people in the two groups. Here too, the results are striking and highly significant (p-value < 2.2e-16). We find that while 98% of converted people in the treatment group buy a second, full price package, only 37% of the converted people in the control group renew their plan. This means that 6.29 % of the treatment group is converted in month two, compared to 0.19% of the control group.

5   Discussion

Although it was not a goal for our data-driven approach to be interpretable, a posteriori categorization of the features selected by our model leads to some interesting qualitative insights. Indeed, most of the features fall under three categories: discretionary income, timing, and social learning, see Table 3.
    Discretionary income was expected by the marketing team to be important overall. They hypothetised that customers with a high total ARPU would be more likely to convert. The model does however not select total spending as an important variable to help predict conversion. In fact, looking at the ARPU of those who received an SMS and then adopt, we see that the low ARPU segment is slightly overrepresented. Our text and data focused discretionary spending variables are however selected as important by the model. Text and data focused spending variables seem to contain relevant information to help predict adoption more than overall spending.
Timing measured through using a new phone is our 9th most useful feature.

Finally, the social learning features we computed for this study turn out to be very helpful to help classify the natural converters. The total spending on data among the closest social graph neighbors is our most predictive feature.

Using social features in selecting customers for the offer might have improved the retention rates. We speculate that the value customers derive from mobile data increases when their neighbors are also mobile data users. In other words that we expect that a network externality effect exists in mobile internet data. This means that selecting customers whose closest neighbors are already using mobile data might have locally used this network effect to create the lasting effect we observe with very high retention rate in the second month.

The success of this pilot study triggered new technical developments and campaign data are now being recorded and stored. Future models might be refined using this experimental data and new insights might be uncovered using statistical models or interpretable machine learning algorithms. We also plan to refine the different attribute categories; social learning, discretionary income, and timing and to use them a priori in further model building.

The marketing team was impressed by the power of this method and is now looking into how this can be implemented in the MNO's operations and used more systematically for future campaigns. They see such data-driven approach to be particularly useful to help launch new products where little prior experience exists. This was the case with this campaign as the overall mobile internet penetration rate in the country is very low. The marketing team usually learns what the right segment for a new product through extensive trial-and-error.

When performing research on sensitive data, privacy is a major concern. [16,17] showed that large scale simply anonymized mobile phone dataset could be uniquely characterized using as little as four pieces of outside information. All sensitive information in this experiment was hashed and only the local marketing team had the contacts of the control and treatment groups.

We believe our findings open up exciting avenues of research within data-driven marketing and customer understanding. Using behavioral patterns, we increased the conversion rate of an internet data campaign by 13 times compared to current best-practice. We expect such an approach will greatly reduce spamming by providing the customer with more relevant offers.

References

1. Chinn, Menzie D. and Robert W. Fairlie, The Determinants of the Global Digital Divide: A Cross-Country Analysis of Computer and Internet Penetration. Economic Growth Center (2004)
2. IBM Center for applied insight, `http://public.dhe.ibm.com/common/ssi/ecm/en/coe12345usen/COE12345USEN.PDF`

3. Predicting Personality Using Novel Mobile Phone-Based Metrics YA de Montjoye, J Quoidbach, F Robic, AS Pentland Social Computing, Behavioral-Cultural Modeling and Prediction, 48-55 (2013)

4. Friends don't lie: inferring personality traits from social network structure, J Staiano, B Lepri, N Aharony, F Pianesi, N Sebe, A Pentland, Proceedings of the 2012 ACM Conference on Ubiquitous Computing, 321-330 (2012)

5. Schwartz HA, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones SM, et al. Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach. PLoS ONE 8(9): e73791. doi:10.1371/journal.pone.0073791 (2013)

6. Jernigan, Carter; Mistree, Behram F.T.. Gaydar: Facebook friendships expose sexual orientation. First Monday, [S.l.], ISSN 13960466 (2009), http://journals.uic.edu/ojs/index.php/fm/article/view/2611/2302.

7. Backstrom, Lars, and Jon Kleinberg. Romantic Partnerships and the Dispersion of Social Ties: A Network Analysis of Relationship Status on Facebook, arXiv preprint arXiv:1310.6753 (2013).

8. Robert M. Bond, Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, James H. Fowler. A 61-million-person experiment in social influence and political mobilization. Nature, 489 (7415): 295 DOI: 10.1038/nature11421 (2012)

9. Sinan Aral, Dylan Walker, Identifying Influential and Susceptible Members of Social Networks Science vol. 337 no. 6092 July 20, page 337 (2012)

10. Wikipedia: http://en.wikipedia.org/wiki/Arpu

11. Social influence. Mapping social psychology series. Turner, John C.Belmont, CA, US: Thomson Brooks/Cole Publishing Co.xvi 206 pp (1991)

12. Christakis, Nicholas A., and James H. Fowler. The spread of obesity in a large social network over 32 years. New England journal of medicine 357.4, 370-379 (2007)

13. P. Sundsøy, J. Bjelland, K. Engø-Monsen, G. Canright, R. Ling, Comparing and visualizing the social spreading of products on a large-scale social network,"The influence on Technology on Social Network Analysis and Mining, Tanzel Ozyer et.al, XXIII, 643 p. 262 illus, Springer (2013)

14. P Sundsøy, J Bjelland, G Canright, K Engø-Monsen, R Ling, Product adoption networks and their growth in a large mobile phone network, IEEE Advanced in Social Network Analysis and Mining (2010)

15. Breiman, Technical Report No. 421, Department of Statistics University of California (1994)

16. de Montjoye, Yves-Alexandre; Hidalgo, César A.; Verleysen, Michel; Blondel, Vincent D., „Unique in the Crowd: The privacy bounds of human mobility", Nature Scientific Reports, Volume 3, id. 1376 (2013)

17. de Montjoye, Y.-A., Wang S., Pentland A., On the Trusted Use of Large-Scale Personal Data. IEEE Data Engineering Bulletin, 35-4 (2012)

18. de Montjoye, Y.-A.*, Quoidbach J.*, Robic F.*, Pentland A., Predicting people personality using novel mobile phone-based metrics. International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction, Washington, USA (2013)