Generating Ad Targeting Rules using Sparse Principal Component Analysis with Constraints

Mihajlo Grbovic
Yahoo! Labs
701 First Avenue
Sunnyvale, USA
mihajlo@yahoo-inc.com

Slobodan Vucetic
Temple University
1805 N Broad Street
Philadelphia, USA
vucetic@temple.edu

ABSTRACT

Determining the right audience for an advertising campaign is a well-established problem, of central importance to many Internet companies. Two distinct targeting approaches exist, the model-based approach, which leverages machine learning, and the rule-based approach, which relies on manual generation of targeting rules. Common rules include identifying users that had interactions (website visits, emails received, etc.) with the companies related to the advertiser, or search queries related to their product. We consider a problem of discovering such rules from data using Constrained Sparse PCA. The constraints are put in place to account for cases when evidence in data suggests a relation that is not appropriate for advertising. Experiments on real-world data indicate the potential of the proposed approach.

Categories and Subject Descriptors

H.2.8 [**Database applications**]: Database Applications Data Mining

Keywords

data mining, computational advertising, audience modeling

1. INTRODUCTION

The key players in display advertising are advertisers and publishers. Given a fixed budget, advertisers wish to advertise a product and maximize the campaign response in terms of user clicks or purchases. Publishers run the websites and want to supplement their income. They are typically able to track user behavior in great detail, including search queries, page views, email activity, ad clicks. This gives them the ability to identify subsets of users who are likely to respond to the campaign. This is referred to as ad targeting. In model-based targeting [1], a model is learned using behavioral data to estimate the probability of action for every user. In rule-based targeting, an audience is manually defined by a set of rules, and any user satisfying all the rules qualifies.

Main representatives of rule-based targeting are 1) site retargeting, which targets users that visited websites of certain companies; 2) email retargeting, which targets users that received emails from certain companies; 3) search retargeting, which targets users that searched certain keywords. Typical scenario is that the advertiser already knows the market well and can provide a list of targeting elements, i.e. websites, domains or keywords. However, most advertisers provide an incomplete list or settle for self-retargeting, thus missing opportunity to show ads to relevant users who visited websites or received emails from domains related to the advertiser.

In this paper we consider a problem of data-driven recommendation of targeting elements. Without loss of generality, let us consider email retargeting as an example. Given user profiles composed from counts of emails received per domain, e.g. during the last $T = 30$ days, we can use Sparse Principal Component Analysis [2] to find small groups of correlated domains. The domains which co-exist with the advertiser's domain in the resulting groups are recommended. The main problem with such correlation-based domain groupings is that not all groupings are meaningful for targeting. For example, a retail domain *nike.com* may be highly correlated with a financial domain *paypal.com* that typically processes the order, and/or *usps.com* that typically ships the order. Moreover, some groupings may not be appropriate from a privacy standpoint. For example, fast food domains were found to be correlated with plus-size clothing stores. Such profiling may offend the users. For this reason we propose a Constrained Sparse PCA variant that leverages taxonomy over domains, and penalizes grouping of domains which belong to categories that are far apart from each other.

2. PROPOSED APPROACH

Given advertiser's website(s), domain(s) or search keyword(s), consider suggesting a handful of targeting elements, websites, commercial domains or keywords to form a rule-based segment. We describe the proposed approach in the context of all three rule-based types, as it is applicable in any single scenario; combining them has not been considered.

We assume n targeting elements (websites, domains or keywords), and m users. Let us define the number of interactions between i-th element and j-th user in the last T days as $x_i(j)$. Considered interactions are: visits to website i by user j, emails sent by domain i to user j, searches for keyword i by user j. Combining information about all n targeting elements, we have a row vector $\mathbf{x}_j = [x_1(j), x_2(j), ..., x_n(j)]$, which serves as a user profile. Finally, combining all m users we have a data matrix $\mathbf{X} = [\mathbf{x}_1^T, \mathbf{x}_2^T, ..., \mathbf{x}_m^T]^T$ of size $m \times n$.

WWW'14 Companion, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/10.1145/2567948.2577351.

Let us assume the features are centered and denote the $m \times m$ covariance matrix of targeting elements as $\mathbf{S} = \mathbf{X}^T\mathbf{X}$. The problem at hand is to decompose $\mathbf{S}$ into K sparse principal components $\mathbf{u}_k, k = 1, ..., K$ of cardinality r. Each resulting PC defines a group, where non-zero element indices define group members. Targeting elements that co-occur with advertiser's query element are suggested as rules.

The problem of maximizing the variance of component $\mathbf{u}$ with a constraint on its cardinality is NP-hard. In [2], the authors proposed an approximate solution to the problem based on a convex relaxation in which the vector $\mathbf{u}$ is replaced by a matrix $\mathbf{U} = \mathbf{u}\mathbf{u}^T$,

$$\begin{aligned} &\text{maximize} && \mathbf{Tr}(\mathbf{SU}) \\ &\text{subject to} && \mathbf{Tr}(\mathbf{U}) = 1, \mathbf{1}^T|\mathbf{U}|\mathbf{1} \leq r, \mathbf{U} \succeq 0. \end{aligned} \tag{1}$$

where $\mathbf{1}$ is a vector of ones, matrix $|\mathbf{U}|$ contains absolute values of $\mathbf{U}$ elements, and $\mathbf{U} \succeq 0$ means that $\mathbf{U}$ is positive semidefinite. The first PC $\mathbf{u}_1$ is obtained as the most dominant eigenvector of $\mathbf{U}^1$, the solution to (1). Similarly to regular PCA, this process iterates by updating $\mathbf{S}$, $\mathbf{S}_2 = \mathbf{S}_1 - (\mathbf{u}_1^T\mathbf{S}_1\mathbf{u}_1)\mathbf{u}_1\mathbf{u}_1^T$ to obtain the second component.

In SPCA with Constraints (CSPCA) [3] we assign an additional cost on grouping features together, and enforce distance constraints on $\mathbf{u}_k$ parameterized by allowed distance d. Given a distance metric $\mathbf{D}$, the total cost associated with the first PC can be defined as $C = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} I(\mathbf{U}_{ij} \neq 0)\mathbf{D}_{ij}$, where I is an indicator function, $\mathbf{D}_{ij}$ is a distance between features i and j and $\mathbf{U}$ is the problem (1) solution. After appropriate convex relaxation due to non-convex cost [3], the resulting constrained Sparse PCA problem is

$$\begin{aligned} &\text{maximize} && \mathbf{Tr}(\mathbf{SU}) \\ &\text{subject to} && \mathbf{Tr}(\mathbf{U}) = 1, \mathbf{1}^T|\mathbf{U}|\mathbf{1} \leq r \\ & && \mathbf{1}^T|\mathbf{U} \circ \mathbf{D}|\mathbf{1} \leq d,\ \mathbf{U} \succeq 0, \end{aligned} \tag{2}$$

Leveraging taxonomy to form D. Given two targeting elements, e.g. domains, i and j and a function $f(\cdot) \rightarrow c$ that maps them into taxonomy nodes (categories), we define a distance between the targeting elements $\mathbf{D}_{ij}$ as the distance between the nodes $\mathbf{D}_{ij} = dist(c_i, c_j)$. In SPCA this distance can be used to enforce a penalty on grouping elements that are far from each other in the taxonomy. Such optimization problem will favor groupings with similar categories.

3. EMPIRICAL ANALYSIS

We chose email retargeting as an example application, and compared 3 targeting approaches: 1) self-retargeting, where only advertiser's domain is targeted, 2) SPCA-based and 3) CSPCA-based targeting. We constructed a matrix $\mathbf{X}$ of email counts that $10,336$ commercial domains sent to 11M anonymous users of Yahoo Mail during 1 month. It is important to stress that we used this data and the email retargeting example purely as a test bed for our framework. The point of our experiments is simply to show that the SPCA rule recommendation framework is effective. Only *from/to* email address data with anonymized user ids was extracted, without processing the email content. Data was extracted exclusively from users who opted in for such research.

In Constrained SPCA, we used a 3 levels deep, $1,733$ node Pricegrabber taxonomy[1] to categorize the domains. The choice of a distance for forming matrix $\mathbf{D}$ was *weighted graph distance* that takes into account the depth of the nodes in the hierarchy; starting from weight 1 for the level one edges, it cuts the weight in half at every next level.

[1]http://www.pricegrabber.com

Table 1: Examples of resulting PCs

SPCA results		CSPCA results	
retail+rewards pc	**food+plus size**	**retail pc**	**games pc**
sears.com	grubhub.com	e.kohls.com	steampowered.
citibank.com	lanebryant.com	landsend.com	gamestop.com
bankofamerica.com	catherines.com	sears.com	thinkgeek.com
o.macys.com	jimmyjohns.com	gap.com	playstation.net
e.bn.com	chipotle.com	jcrew.com	
travel+rewards pc	**food+games pc**	**vitamins pc**	**books pc**
starwoodhotels.com	github.com	vitacost.com	half.com
tdrewards.com	steampowered.com	luckyvitamin.com	adebooks.com
tdbank.com	playstation.net	swansonvitamins	biblio.com
expedia.com	gamestop.com	iherb.com	thriftbooks.com
southwest.com	thinkgeek.com	walgreens.com	alibris.com
	dominos.com	christianbook.com	
hunting+other pc	**furniture+other**	**outdoor pc**	**travel pc**
cheaperthandirt.com	potterybarn.com	rei.com	megabus.com
brownells.com	westelm.com	backcountry.com	greyhound.com
kingsizedirect.com	containerstore.com	campmor.com	enterprise.com
sportsmansguide.com	americanexpress.	orvis.com	airtran.com
usps.com	swgas.com	usoutdoor.com	

Table 1 shows examples of resulting PCs obtained by solving problems SPCA (1) and CSPCA (2) on the domain covariance matrix $\mathbf{S}$. Target cardinality was set to 5. It can be observed that SPCA may produce corrupted results, due to random noise, or connections between domains that are not useful for targeting, e.g. banks co-occur with retail or travel domains, i.e. rewards or processing, and shipping companies co-occur with shopping domains. DSPCA constraints handle these side effects better, producing much cleaner results.

We studied the performance of SPCA- and DSPCA-based targeting recommendations compared to the baseline self-retargeting in 138 advertising campaigns. Similarly to [1] we show offline evaluation based on the logs of user activities. Recommendations were generated using the 1-st month data $\mathbf{X}$, and tested on user email profiles and conversions collected during the 2-nd (following) month. The goal was to determine how many more converters would be found during the 2-nd month if targeting included the recommended domains found using SPCA vs. SPCA with Constraints.

Performance was measured in terms of the yield rate, $YR = \frac{\#\text{conversions}}{\#\text{impressions}}$. In our setting, the counts of ad impressions in self-retargeting was less than the counts of ad impressions in SPCA- or DSPCA-based targeting. To make the denominator in YR equal in both cases, we added random users to the self-retargeting segment. Compared to the self-retargeting, the DSPCA approach improves YR by 4.24%, averaged over all 138 campaigns, while the SPCA approach improvement was 2.58%. This justifies the use of constraints in domain suggestions for email retargeting. We omitted the results in which the performance was averaged by advertiser category. These results showed that CSPCA outperformed SPCA in most categories, such as retail and travel. There some were exceptions, such as fast food, indicating that the potentially offensive rules help in targeting.

4. REFERENCES

[1] M. Aly, A. Hatch, V. Josifovski, and V. K. Narayanan. Web-scale user modeling for targeting. In *WWW*, 2012.

[2] A. d'Aspremont, L. El Ghaoui, M. I. Jordan, and G. R. Lanckriet. A direct formulation for sparse pca using semidefinite programming. *SIAM review*, 49(3), 2007.

[3] M. Grbovic, C. R. Dance, and S. Vucetic. Sparse principal component analysis with constraints. In *AAAI*, 2012.