

From Online Behaviors to Offline Retailing

Ping Luo[‡], Su Yan^{‡b}, Zhiqiang Liu[§], Zhiyong Shen[§], Shengwen Yang[§], Qing He[‡]

[‡]Key Lab of Intelligent Information Processing of Chinese Academy of Sciences
Institute of Computing Technology, CAS, Beijing, China.

^bUniversity of Chinese Academy of Sciences, Beijing, China.

[§]Baidu, Inc., Beijing, China.

luop@ict.ac.cn, shenzhiyong@baidu.com

ABSTRACT

To combat the ease of online shopping in pajamas, offline mall owners focus increasingly on driving satisfaction and improving retention by identifying customers' preferences. However, most of these studies are based on customers' offline consuming history only. Benefiting from the internet, we can also get customers' online behaviors, such as the search logs, web browsing logs, online shopping logs, and so on. Might these seemingly irrelevant information from two different modalities (i.e. online and offline) be somehow interrelated? How can we make use of the online behaviors and offline actions jointly to promote recommendation for offline retailing?

In this study, we formulate this task as a cross-modality recommendation problem, and present its solution via a proposed probabilistic graphical model, called *Online-to-Offline Topic Modeling* (O2OTM). Specifically, this method explicitly models the relationships between online and offline topics so that the likelihood of both online and offline behaviors is maximized. Then, the recommendation is made only based on the pairs of online and offline topics, denoted by (t, l) , with high values of *lift*, such that the existence of the online topic t greatly increases the response on the corresponding offline topic l compared with the average response for the population without the online topic t . Furthermore, we evaluate this solution in both live and retrospect experiments. The real-world deployment of this model for the anniversary promotion campaign of a famous shopping mall in Beijing shows that our approach increases the occurred customer purchases per promotion message by 29.75% compared with the baseline. Also, our model finds some interesting interpretable relationships between the online search topics and offline brand topics.

Keywords

Brands recommendation, Topic modeling, Recommendation explanation

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

KDD '16, August 13-17, 2016, San Francisco, CA, USA

© 2016 ACM. ISBN 978-1-4503-4232-2/16/08...\$15.00

DOI: <http://dx.doi.org/10.1145/2939672.2939683>

1. INTRODUCTION

Recommendation systems (RS) [16] have been popular for decades, improving the quality of our daily lives by facilitating the item selections from a large set of candidates. Here, we focus on the recommendation systems for *offline retailing*. Although *online shopping* has the tendency of booming development, offline retailing still covers 92% and 90% of the US and China retailing market in 2014 [1, 2], respectively. Thus, small improvements on offline retailing might draw great benefits. Meanwhile, with the penetration of IT systems into this traditional sector, the customers' offline consumption logs can be recorded to better understand customers' interests for driving satisfaction and improving retention. However, previous studies mostly consider the behaviors from offline and online separately [8, 12].

In this paper, we study the online and offline activities from a joint perspective. Here, online behaviors consist of all the activities on Internet, such as web browsing, web searching, online video watching, online music listening, and online shopping etc, while offline behaviors include all the activities in real lives, such as shopping in a mall, traveling, dining out, watching films etc. Figure 1 shows an example of the data we have from these two modalities. For a small fraction of users, we might know their activities on both online and offline sides. On the online side, we know what they searched in a search engine, what videos they watched, and what items they bought online. On the offline side, we know what brands they consumed in an offline shopping mall. With these data, we might ask could the videos users watch on Youtube interpret the tour sites they select for travelling? Could the news users read on internet be related with the books they borrow in a library? If we connect the online and offline behaviors together, could these two modalities of data fire the chemical reactions for more meaningful recommendation?

This study pioneers towards this direction. As a case study, here we consider the online search logs as the online behaviors, and the offline shopping history as the offline behaviors (note that the proposed model can be applied to any type of online and offline activities). Originally, these two sets of data are owned by Internet companies and local malls, separately. We will show how profitable if Internet companies with online data and local retailers with offline data cooperate for a win-win situation (with the consent from users).

Our basic assumption in this study is that customers' online behaviors could reflect their interests, which in turn influence their offline behaviors. From the real-world data, we

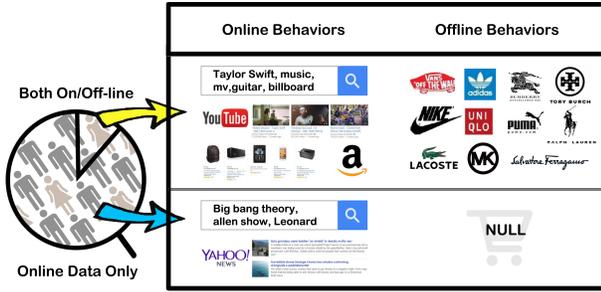


Figure 1: Offline behaviors and online behaviors.

do find some interesting examples which demonstrate that users’ search logs are associated with their offline shopping activities (even although people might not search on Internet the exact items they buy offline). Take a clothing brand *EVISU* as an example. Readers can perceive the style of this clothing brand by its in-store advertisement poster, shown in Figure 2. Also, Table 1 shows some key query words issued by the customers of this brand. It clearly demonstrates that people who buy *EVISU* are interested in Japanese animations, computer games, indicating that they are the fashionable new generation. This is consistent with the brand spirit of *EVISU*.

With this assumption, if the relationship between online and offline behaviors can be automatically discovered, it will greatly benefit the mall owners in both offline and online scenarios. A typical offline application is for the sales campaign of a shopping mall. To increase sales revenue, once or twice a year the mall carries out the promotion events by sending coupons to the targeted customers. So, which coupon should be sent to a given user? Setting aside the scale of the discount, it is reasonable to assume that the more a user like a brand the more likely she will use its coupons. Previously, this recommendation for offline coupons is based on users’ offline shopping history only. With the support of online behaviors, more accurate recommendation is expected to increase the revenue significantly. This is also validated by the live experiment conducted in a famous shopping center in Beijing (detailed in Section 5).

For the online application, with the connection between online and offline behaviors, the local retailers might conduct more effective online advertising to the target customers for a brand. For example, if we know that people who buy *EVISU* often watch Japanese animations, we can directly put the coupons of *EVISU* just before they watch the videos. Also, the representative words of the related topics are the important key words the brand manager should bid for the paid ads shown in the search results.

Motivated by these applications, here we propose the probabilistic graphic model to depict the online and offline behaviors jointly with the goal for offline retailing recommendation. Compared with the previous studies on RS, this task imposes us some new challenges.

- Multiple modalities and modality missing. The online and offline activities actually form the two modalities of user behaviors. Previous studies take use of only one modality (online or offline) for recommendation, but neglect the other part. Additionally, as shown in Figure 1, only a small



Figure 2: An in-store poster for *EVISU*

Word	Description
火影	<i>Naruto</i> , a Japanese anime
海贼王	<i>One Piece</i> , Japanese manga series
大学	campus
动漫	animations
lol	<i>League of Legends</i> , online battle arena game
ios	iphone operating system
ps4	a video game console from Sony Entertainment

Table 1: The search keywords of *EVISU* customers

fraction of users have both the online and offline activities. For the left large amount of users who are potential customers of a mall (living around the mall), they might only leave some online tracks, however, might not consume in the mall before. It means that their offline modality is missing. Thus, we need to transform these online tracks into the intentions and preferences for offline consumption.

- Interpretability. Previous recommendation methods, like matrix factorization [27], are usually less interpretable. In their methods, since users and items are mapped to the latent vectors, it is hard to explain what each dimension represents. However, for mall owners it will be much more acceptable if the reasons of retail recommendations are provided. Additionally, for the applications of online advertising we need to know the exact key words which are predictive to the consumption of a certain brand. In this study, the online tracks, especially the queries users search, provide us the unique opportunity for the explanation of recommendation. Thus, we need to not only make the recommendations, but also explicitly model the relationships between the online and offline behaviors such that the offline recommendations can be explained by the online search words. Later, we will give more meaningful examples to show the interpretability of our model.
- Predictive vs. unpredictable features. It is also worth emphasizing that not all the online behaviors are predictive to offline activities. For example, most people search the latest news from the search engine. Intuitively, the key words about the news, e.g., the local place names and the people names in breaking news, might not be predictive or causally related to any offline consumptions. Thus, the proposed model needs to automatically distinguish predictive online features from the unpredictable ones so that the recommendation is made only based on the predictive online features.

For past decades, many researches have been conducted for recommendation. The detailed discussion on these related works on recommendation techniques will be presented in Section 7. Here, we argue that although previous studies might address some of the challenges separately there still do not exist a model to tackle them all, especially for the issues of interpretability and predictive features. In this paper, we assume that both online and offline behaviors can be grouped into some semantically related topics. Then, we

develop the *lift*-based recommendation method, where the recommendation is made only based on the pairs of online and offline topics, denoted by (t, l) , with high values of *lift*, such that the existence of the online topic t greatly increases the response on the corresponding offline topic l compared with the average response of the population without the online topic t . To estimate these lift values, we propose a probabilistic graphic model, called *Online to Offline Topic Model* (O2OTM), to explicitly build the relationships between online and offline topics so that the likelihood to generate all the online and offline behaviors is maximized. We argue that this lift-based recommendation method will be more practically useful for the live recommendation scenarios.

Finally, we evaluate our model by both retrospect and live experiments to show its superiority over several baselines, and also give more examples to show its interpretability. For the live evaluation, we deployed our method into a sales campaign in a famous shopping center in Beijing. We find that for each promotion message our model brings 29.75% more purchases than the random method, and 12.5% revenue more than the method with the offline behaviors only. Therefore, we argue that by combining online and offline behaviors our solution has great potential of business impact for offline retailing.

2. PROBLEM FORMULATION

Here, we recommend brands of offline retailing to customers with the support of their online and offline behaviors. In this section, we describe the data we have and then formulate the problem. All the mathematical symbols are described in Table 2.

In each recommending task, users and items are two indispensable elements. For our problem, users are all the potential customers of a shopping mall while items are the product brands sold in the mall. \mathcal{B} is the set of all the brands, and \mathcal{U} is the set of all the customers. Meanwhile, we might know the search queries issued by these customers in a popular search engine. \mathcal{V} is the vocabulary set of all the search words.

For each customer $u \in \mathcal{U}$, its offline behaviors are represented by $\mathcal{B}_u = \langle b_{u,i} | b_{u,i} \in \mathcal{B} \rangle$, where $b_{u,i}$ is the i -th brand token bought by user u . Meanwhile, the customer’s online behaviors are represented by $\mathcal{W}_u = \langle w_{u,j} | w_{u,j} \in \mathcal{V} \rangle$ where $w_{u,j}$ is the j -th word token searched by user u . In other words, both the online and offline behaviors are represented in the bag-of-words manner in this study. Note also that \mathcal{B}_u or \mathcal{W}_u can be an empty set. Based on all these data, we focus on the recommendation of brands. Specifically, we give a ranking of the brands in \mathcal{B} for each user u .

3. ONLINE TO OFFLINE TOPIC MODEL

Before we detail the lift-based recommendation method, we propose the Online to Offline Topic Model (O2OTM), which models the relationships between online and offline topics explicitly based on the data corpus $\{\mathcal{W}_u, \mathcal{B}_u\}_1^U$ from both online and offline sides.

3.1 Modeling online and offline behaviors separately

First, the basic topic modeling method, LDA [5], can be applied to model the online and offline behaviors separately. LDA is developed to modeling document content based upon

Symbol	Description
\mathcal{B}	the set of all the brands
\mathcal{V}	the vocabulary set of all the search words
\mathcal{U}	the customer set
U	the total number of customers
\mathcal{B}_u	the brands bought by customer u
\mathcal{W}_u	the words searched by customer u
\mathcal{P}_u	the brands which are not consumed by u
θ	distribution over online topics for customers
ϕ	distribution over search words for online topics
π	distribution over offline topics for online topics
φ	distribution over brands for offline topics
π	distribution over offline topics for online topics
$\alpha, \beta, \eta, \lambda$	Dirichlet prior
x, y	online topic variable
z	offline topic variable
w	word variable
b	brand variable
t	index for online topics
l	index for offline topics
v	index for search words
k	index for brands

Table 2: The symbols.

the idea that the probability distribution over words in a document can be expressed as a mixture of topics, where each topic is a probability distribution over words.

In this study, since both online and offline behaviors are represented in a bag-of-words model, we can consider each \mathcal{W}_u or \mathcal{B}_u as a document to form the online corpus $\{\mathcal{W}_u\}_1^U$ and offline corpus $\{\mathcal{B}_u\}_1^U$. Then, LDA can be applied to these two corpora to generate online topics (denoted by x) and offline topics (denoted by z), separately. Here, an online topic $p(w|x)$ is a probability distribution over search words w while an offline topic $p(b|z)$ is a probabilistic distribution over brands b . This method helps to group search words and product brands into semantic topics, and can be used to answer queries about similar online users in terms of search words and similar customers in terms of consumed brands. However, this method does not provide any information about the dependent relationships between the online and offline topics, thus cannot perform modality transformation from online behaviors to offline activities.

3.2 Modeling online and offline behaviors jointly

The proposed O2OTM model draws upon the strengths of probabilistic graphic methods to model the online and offline behaviors jointly by explicitly considering the probabilistic dependency between online and offline topics. As in the O2OTM model, each user u adopts the following probabilistic process to generate its online and offline behaviors, denoted by $(\mathcal{W}_u, \mathcal{B}_u)$. This process can be divided into two parts. First, for each search word $w \in \mathcal{W}_u$, an online topic x is chosen from a distribution over online topics specific to u (denoted by θ_u), and the word is generated from the chose online topic. Second, for each brand $b \in \mathcal{B}_u$, again an online topic y is chose from θ_u . Based on y , an offline topic z is chosen from π_y , which is a distribution over offline topics specific to y ; and then the brand is generated from the chose offline topic.

The graphical model corresponding to this process in Figure 3. In this figure, the left part generates the online be-

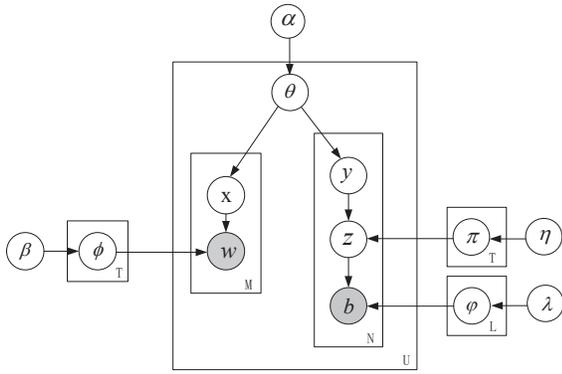


Figure 3: The graphical model of O2OTM

haviors while the right part generates the offline behaviors. In this model, each user u is associated with a distribution over online topics, θ_u (chosen from a symmetric *Dirichlet*(α) prior). The mixture weights corresponding to u are used to select an online topic x (or y). Then, based on x a search word is generated according to the distribution ϕ_x (drawn from a symmetric *Dirichlet*(β) prior). Meanwhile, based on y an offline topic z is generated according to the distribution π_y (drawn from a symmetric *Dirichlet*(η) prior), and a brand is generated according to the distribution φ_z (drawn from a symmetric *Dirichlet*(λ) prior).

Here, both x and y represent online topics. We use these two different symbols to distinguish the online topics which generates the search words from the ones which generates the offline brands. Additionally, both x and y are generated from the same distribution θ_u . Thus, θ_u is dependent on both online and offline behaviors. It is also worth mentioning that the dependent relationship between online and offline topics is modeled in π . It can be used for online-to-offline modality transformation.

3.3 Gibbs sampling algorithm to O2OTM

In this study we use Gibbs sampling [9] to estimate the parameters of O2OTM. By applying Gibbs sampling, we construct a Markov chain that converges to the posterior distribution on the latent variables. The transition between successive states of the Markov chain results from repeatedly drawing x, y, z from its distribution conditioned on all other variables, summing out the parameters using standard Dirichlet integrals. Specifically, we have the following three equations to update x, y, z .

First, we have

$$p(x_i = t | \vec{x}_{-i}, \vec{y}, \vec{z}, \vec{w}, \vec{b}) = p(x_i = t | \vec{x}_{-i}, \vec{y}, w_i = v) \propto \frac{O_{t,-i}^v + \beta_v}{\sum_{v=1}^V O_{t,-i}^v + \beta_v} \cdot \frac{O_{u,-i}^t + \alpha_t}{\sum_{t=1}^T O_{u,-i}^t + \alpha_t}, \quad (1)$$

where $x_i = t$ represents the assignment of the i -th word token in the corpus to the t -th online topic, $w_i = v$ represents the observation that the i -th word token is the v -th one in the word lexicon, \vec{x}_{-i} represents all assignments of online topics to the words, not including the i -th one. Additionally, $O_{t,-i}^v$ denotes the number of times that the v -th word being observed with the t -th online topic, not including the current instance. $O_{u,-i}^t$ denotes the number of times that the t -th online topic being observed in user u (who own-

s the search word w_i), not including the current instance. Here, the online topic x is assigned to each word token in the corpus.

Next, we have

$$p(y_j = t | \vec{y}_{-j}, \vec{x}, \vec{z}, \vec{w}, \vec{b}) = p(y_j = t | \vec{y}_{-j}, \vec{x}, z_j = l) \propto \frac{O_{t,-j}^l + \eta_l}{\sum_{l=1}^L O_{t,-j}^l + \eta_l} \cdot \frac{O_{u,-j}^t + \alpha_t}{\sum_{t=1}^T O_{u,-j}^t + \alpha_t}, \quad (2)$$

where $y_j = t$ represents the assignment of the j -th brand token in the corpus to the t -th online topic, $z_j = l$ represents the observation that the l -th offline topic is assigned to the j -th brand token, \vec{y}_{-j} represents all assignments of online topics to the brand tokens, not including the j -th one. Additionally, $O_{t,-j}^l$ denotes the number of times that the l -th offline topic being observed with the t -th online topic, not including the current instance. $O_{u,-j}^t$ denotes the number of times that the t -th online topic being observed in user u (who buys the brand b_j), not including the current instance. Here, the online topic y is assigned to each brand tokens in the corpus.

Finally, we have

$$p(z_j = l | \vec{z}_{-j}, \vec{x}, \vec{y}, \vec{w}, \vec{b}) = p(z_j = l | \vec{z}_{-j}, y_j = t, b_j = k) \propto \frac{O_{l,-j}^k + \lambda_k}{\sum_{k=1}^B O_{l,-j}^k + \lambda_k} \cdot \frac{O_{t,-j}^l + \eta_l}{\sum_{l=1}^L O_{t,-j}^l + \eta_l} \quad (3)$$

where $z_j = l$ represents the assignment of the j -th brand token to the l -th offline topic, $b_j = k$ represents the observation that the j -th brand token is the k -th brand in the brand lexicon, $y_j = t$ represents the observation that the j -th brand token is assigned to the t -th online topic, \vec{z}_{-j} represents all assignments of offline topics to the brand tokens, not including the j -th one. Additionally, $O_{l,-j}^k$ denotes the number of times that the l -th offline topic being observed with the k -th brand, not including the current instance. $O_{t,-j}^l$ denotes the number of times that the t -th online topic being observed with the l -th offline topic. Here, the offline topic z is assigned to each brand token in the corpus.

After the Markov chain converges, the model parameters can be estimated via

$$\theta_{ut} = \frac{O_u^t + \alpha_t}{\sum_{t=1}^T O_u^t + \alpha_t} \quad (4)$$

$$\phi_{tv} = \frac{O_t^v + \beta_v}{\sum_{v=1}^V O_t^v + \beta_v} \quad (5)$$

$$\pi_{tl} = \frac{O_t^l + \eta_l}{\sum_{l=1}^L O_t^l + \eta_l} \quad (6)$$

$$\varphi_{lk} = \frac{O_l^k + \lambda_k}{\sum_{k=1}^B O_l^k + \lambda_k} \quad (7)$$

Here, T, L are the numbers of online and offline topics, respectively. V, B are the sizes of the word vocabulary and brand lexicon.

In this study, we do not estimate the hyperparameters $\alpha, \beta, \eta, \lambda$. Instead, these smoothing parameters are fixed at 0.05, 0.01, 0.01, 0.01, respectively. The algorithm begins by assigning search words to random online topics, and assigning offline brands to random online and offline topics. Each iteration of the algorithm involves applying Equ. (1),(2),(3),

to every word token and brand token in the corpus. This sampling process performs at least 2000 rounds until convergence.

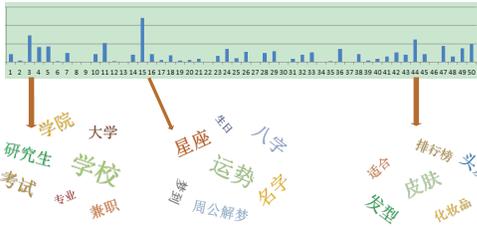
3.4 Model Inference

After we learn all the parameters θ, π, φ , we can infer the relationship between a brand and an online topic. Specifically, for an online topic $y = t$ and an offline brand $b = k$, we have

$$\begin{aligned}
 p(y = t|b = k) &= \sum_{l=1}^L p(y = t|z = l)p(z = l|b = k) \\
 &\propto \sum_{l=1}^L p(y = t)\pi_{tl}\varphi_{lk}
 \end{aligned}
 \tag{8}$$



(a) The in-store poster for *Teenie Weenie*



(b) The typical words of the 3 highly-related online topics

Figure 4: The brand *Teenie Weenie* and its related search words. The description of words is listed in Table 3.

Topic #	Word	Description
3	学校	campus
3	学院	school
3	大学	university
3	研究生	graduate students
3	兼职	part-time job
3	专业	majors in school
15	运势	horoscope
15	星座	12 zodiac signs
15	八字	Chinese horoscope according to the birthday
15	名字	name
15	周公解梦	understanding the dreams
44	皮肤	skin
44	头发	hair
44	发型	hair style
44	化妆品	cosmetics
44	排行榜	top charts

Table 3: English translation of words in Figure 4

With this derivation, here we show you an example of the relationship between a brand and online topics. *TEENIE WEENIE* is a Korean brand designed for young girls from 15 to 25 year-old. It is famous for its school-style dress like the poster in Figure 4(a). Figure 4(b) shows the distribution over online topics for this brand, and the representative words for the top 3 topics are also shown with English translation in Table 3. The first topic is mainly about campus life, which indicates that the brand is somehow related to college students. The second topic is about horoscope. Many people especially young women are interested in their horoscopes, which could forecast the love, fortune and career etc. The third topic is about hair style and cosmetics, which shows the brand is related to a group of female people who cares about their appearance. These three topics clearly show that the customers of *TEENIE WEENIE* are a group of young girls, especially college girls, who care about their appearance and long for love and good future. We argue that this kind of relationship could show the characteristics of the consumers of the brand, and help designers to adjust products according to their customers' interests and tastes.

4. CROSS-MODALITY RECOMMENDATION

In this section, we present the methods for cross-modality recommendation.

4.1 Recommendation for Retrospect Evaluation

With all the parameters θ, π, φ , we could make recommendation for customers. The probability of the i -th user buying the k -th brand is calculated as follows:

$$\begin{aligned}
 p(b = k|u = i, \theta, \pi, \varphi) &= \sum_{t=1}^T \sum_{l=1}^L p(b = k|z = l, \varphi)p(z = l|y = t, \pi)p(y = t|u = i, \theta) \\
 &= \sum_{t=1}^T \sum_{l=1}^L \varphi_{lj}\pi_{tl}\theta_{it}
 \end{aligned}
 \tag{9}$$

With the estimation of θ_i , this recommendation is made based on both its online and offline behaviors. Since this method can easily generate the ranking of brands for a user, we will evaluate it in the retrospect experiments. This method is deliberately developed only to show the online behaviors might provide extra value for this recommendation task, compared with the method using the offline behaviors only.

4.2 Recommendation for the Live System

The recommendation method in Equ. (9) has the following drawbacks. First, this method uses all the online topics for the offline recommendation although we believe that some online topics are not predictive or causally related to the offline consumption, as mentioned earlier. Additionally, these unpredictable online topics might occupy a large portion of the search content of a user while the topics with the consumption intentions are drowned out. For example, the search content from a primigravid woman might contain little about pregnancy care compared with the large portion about her work and everyday life. Thus, the online topic distribution for this user will have a low value for the topic

about pregnancy care, and the brands for babies and pregnant women might not be ranked high based on Equ. (9).

4.2.1 Lift measure

Therefore, we need a measure to automatically identify the pairs of predictive online topic and its corresponding offline one so that the recommendation is made only based on these pairs.

Specifically, we define the following *lift* measure between an online topic t and an offline topic l . Remind that the proposed O2OTM model outputs the relationship between online and offline topics. Namely, the probability of offline topic l given online topic t , denoted by $\pi_{tl} = p(z = l|y = t)$, is assigned so that the likelihood of all the data is maximized. Then, based on π we can also calculate $\pi_{-tl} = p(z = l|y = -t)$, measuring the probability of offline topic l given all the other online topics t , excluding t . Then, the *lift* is defined as follows

$$lift_{tl} = \frac{p(z = l|y = t)}{p(z = l|y = -t)} = \frac{\pi_{tl}}{\pi_{-tl}} \quad (10)$$

This lift value is simply the ratio of two values: the target response l with the existence of t divided by the response without t . It actually measures how much enhancement we can get in predicting l conditioned on t compared to the condition without t . The bigger this lift value, the more association between online topic t and offline topic l .

4.2.2 Examples of high-lift pairs

Table 4 shows some examples of high-lift pairs of online and offline topics discovered from our data. These examples are selected so that they can cover different groups of consumers.

- For Group I, the search words are about pet care, and this online topic corresponds with the lift value of 11.5 to the fashioned brands for young girls. Since most of these brands are imported from Japan and Korea, they are more expensive than the regular ones, indicating that these girls are affordable to raise pets.

- For Group II, it is matched to the group of young people (mostly boys) who love Japanese anime series and computer games. What they wear is mostly from some casual brands, such as VANS, ADIDAS ORIGINAL and so on.

- For Group III, it corresponds to the group of pregnant women and young mummies. They search for the knowledge about pregnancy care, newborn baby care and postpartum recovery. Also, they corresponds to two offline topics with high lift values. The first brand group refers to the ones for newborn children, while the second group refers to some brands for young girls, indicating that these new mummies still cannot stop buying something for themselves.

- For Group IV, it refers to the group of housewives with happy lives. On the online side, they search for the materials for keeping fit, popular TV series, netnovels and so on. On the offline side, they buy two groups of brands. One is for their children, and the other is for their husbands with the brands for outdoor sportswear.

- For Group V, we consider the customers of the high-end cosmetic and jewellery brands. We are surprising to find that the lift value reaches 19 when they search for the online materials for gongfu stories, love and urban novels, and literature. Also, the people, who search for the materials about plastic surgery and phone games, like to buy these expensive brands.

- For Group VI, we consider the people who search for stocks, funds, securities, and investment. These are rich people who own some extra money for asset management. The offline topic with the highest lift value to this online topic is about some fancy restaurants in China with different types of cuisines. It indicates that good food is the first priority of these people.

It is worth mentioning that all these pairs of online and offline topics with high values of lift are automatically discovered by the proposed O2OTM model. They are consistent with our basic understandings on online and offline behaviors.

4.2.3 Lift-based recommendation

Then, with a user-specified threshold $\rho > 0$ on lift, we can identify the pairs of online and offline topics whose lift values are not smaller than ρ . Namely, we have

$$\mathcal{R} = \{(t, l) | lift_{tl} \geq \rho\} \quad (11)$$

Then, with these high-lift pairs \mathcal{R} we develop the following recommendation method. Consider the scenario of the recommendation for weekend shopping. The search words in the recent weekdays are collected for inferring their online topics. Once an online topic t appears at least for q times, it is considered to match with \mathcal{R} to check whether t appears in $\mathcal{T} = \{t | (t, l) \in \mathcal{R}\}$. If yes, the top brands in the corresponding offline topics are identified for recommendation. Some postprocessing steps are needed to consider the user’s recent shopping history and recommendation logs to avoid repetitive recommendation. In this study, the two thresholds are set as $\rho = 4.5$ and $q = 10$.

In this method, the recommendations are made in a triggering manner. It means that recommendations only happen when the online topics with high lift values appear for enough times and the corresponding recommendations are considered to be reliable enough. It can automatically judge when to recommend, and when not even if many unpredictable search words appear. Since the pairs reflect clear intentions for offline consumption as shown in the examples in Table 4, we believe that the recommendations from this lift-based method will be more attractive to potential customers, while avoiding unnecessary disturbance to users.

5. LIVE EVALUATION

We deployed the proposed lift-based model to a real-world sales campaign in a famous shopping mall in Beijing. In order to celebrate its anniversary, the mall planed to send out lots of coupons from different brands to attract customers. These coupons were sent via short messages 3 days before the anniversary day, and each short message contained 3 coupons of different brands. For the VIP users of this mall, we knew both their offline consumption history and online search queries with their consent. We randomly divided these users into 3 groups, each of which contains around 10,000 users. With this big number of users in each group we assume that the demographic information (e.g. gender, education, spending power, etc.) are the same for these three groups of users.

Then, we evaluated the recommendation performances of three methods, which are responsible for the coupon recommendation for these 3 user groups respectively. These 3 methods include: 1) RAN, representing the random recommendation; 2) LDA, representing the method based on

the offline consumption history only. Specifically, the LDA method is applied on $\{\mathcal{B}_u\}_{u=1}^U$, and the recommendation for offline retailing can be made by $p(b|u) = \sum_z p(b|z)p(z|u)$; 3) LIFT, representing the lift-based method based on both online and offline data. The 3 methods are deliberately designed to answer the questions on whether the recommendation based on history data is effective compared with the random method, and whether the online behaviors can provide extra value for this recommendation compared with the offline data only.

Methods	IPR	APPC (RMB)	OCPPUP (RMB)
RAN	2.34%	1656.64	38.76
LDA	2.39%	1870.30	44.70
LIFT	2.59%	1942.10	50.30

Table 5: The live evaluation results

In this live evaluation, we recorded the following three 3 measures on the mall consumption for these 3 user groups on the anniversary day:

- IPR, *the instore purchase rate* which means the percentage of people consuming in the mall on the anniversary day.
- APPC, *average purchases per customer*.
- OCPPUP, *occurred customer purchases per unit promotion* which equals to $IPR \times APPC$ and means the purchases each promotion message could bring to the mall.

The results are shown in Table 5. Clearly, the consumptions from the LIFT group are better than the other two groups in terms of all these three measures. Although the increase on IPR is small, the measure of APPC increases greatly. It results in the OCPPUP value of 50.3, meaning that each short message with the coupons recommended by the LIFT method brings 50.3 RMB revenue to the mall in average. For the other two groups of users, the OCPPUP values are 44.7 and 38.76, respectively. Thus, the LIFT group achieves 29.75% increase on OCPPUP compared with the RAN group, and 12.5% increase compared with the LDA group.

Since the demographic statistics are the same for these 3 groups of users, we attribute this purchase increase to better coupon recommendation by LIFT. It indicates that with both the online and offline behaviors O2OTM models the users’ preferences more accurately to achieve more attractive recommendations. It clearly shows the great potential business impact if the recommendations based on both online and offline behaviors are introduced to offline retailing.

Furthermore, we check what kind of users the consumption increase comes from. Specifically, for each group of users, we rank them in terms of their purchases. Then, we divide the users, who purchased in this sales campaign, into 10 subgroups by their rankings. Namely, the top 10% users go to the 1st subgroup, the top 10%-20% users go to the 2nd one, and so on. Then, for each subgroup we calculate its APPC value. The results are shown in Figure 5, where the number on each bar represents the purchase percentage of the users in this subgroup compared to the total purchases of all the users.

We have the following observation. First, the users whose spending ranks in the top 10% contribute more than 54.6% of

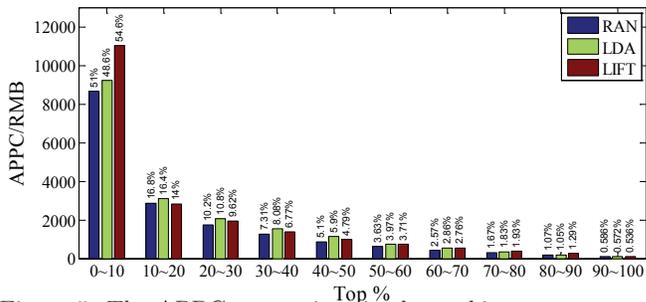


Figure 5: The APPC comparison in the ranking.

the total revenue for the LIFT user group. Second, the APPC increase from LIFT mainly comes from these top ranked high-valued customers. It indicates that these high-valued customers are more sensitive to the performance of coupon recommendation. The better the recommendation serves for high-valued customers, the more purchases it will bring.

6. RETROSPECT EVALUATION

In this section, we compare the performances of the recommendation methods in retrospect evaluation.

Each data set we use consists of online part and offline part. The online part contains users’ search queries from a widely used online search engine in China. Each search log has three attributes, namely user id, query and date. Note that a query is parsed into a set of words. The offline part contains consumers’ shopping history in a local shopping mall. Each shopping transaction consists of consumer id, brand she buys and date.

In this retrospect evaluation, we use two data sets from two different malls in China, denoted by Mall A and Mall B. In total, we get 6539 users from Mall A, with 732 offline brands and 30097 search words. For Mall B, there are 6961 users with 789 offline brands and 103597 search words. For each data set, we use the data from 2013/01/01 to 2014/12/31 for training, and the data from 2015/01/01 to 2015/06/01 for testing.

Many measures [21] can be used to evaluate the performance of recommendation systems, like mean absolute error (MAE), root mean squared error (RMSE), precision, recall. Since this problem is not for rating prediction, it is not suitable to adopt MAE and RMSE for evaluation. Neither, we cannot use recall since the ground truth for all the brands a user will buy is unknown. We can only calculate the percent of the recommended brands that a user really bought in the testing data. After getting the precision for each customer, we average them and derive the *precision@K* measure for the whole customers.

In this evaluation we focus on recommendation for the products which are not consumed by a user before, since these un-consumed products might reflect users’ preferences more precisely. We know that there are some previous studies focusing on repetitive consumption of products [3], which is mostly decided by the use experiences of the products, the product attributes (e.g. continuous or one-shot usage), and some user psychological issues (e.g. boredom) [19]. However, the recommendation for repetitive consumptions is beyond the range of this study.

Here, we have the two baseline methods, RAN and LDA, which are the same in the live evaluation. The method of

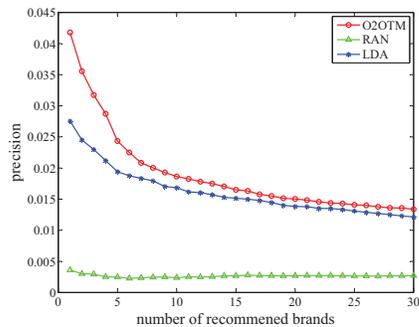
RAN refers to the random recommendation method, while LDA refers to the one using the offline behaviors only. For the method of RAN, its $precision@K$ measure can be analytically estimated as follows,

$$precision@K = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \frac{|\mathcal{T}_u - \mathcal{B}_u|}{|\mathcal{B} - \mathcal{B}_u|}, \quad (12)$$

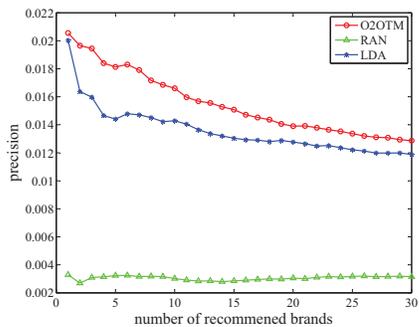
where $(\mathcal{B} - \mathcal{B}_u)$ is the set of all the un-consumed brands customer u and $(\mathcal{T}_u - \mathcal{B}_u)$ includes all the un-consumed brands that customer u buys in the testing set. Note that this measure for RAN is independent of K .

In this retrospect evaluation, we compare these two baseline methods with the recommendation method via Equ. (9), denoted as O2OTM. This method can generate the full ranking of all the un-consumed brands. However, the lift-based method might recommend nothing when its online topics do not hit the ones with high lift values. Thus, we do not include it for this retrospect evaluation.

Note also that there exists some strong baseline methods, such as factorization machine [15], which can embed the features of users and items (e.g. all the search words) for recommendation. However, these methods cannot provide any interpretability for recommendation. Additionally, the goal of this retrospect evaluation is only to show that the online activities can provide extra value for offline retailing. Thus, we do not include these methods for evaluation.



(a) Results on Mall A



(b) Results on Mall B

Figure 6: $precision@K$ measures for the three methods on two malls

After some initial experiments, we set the number of online topics as $T = 50$, and the number of offline topics as $L = 30$ for this retrospect evaluation. The results of the $precision@K$ measure on the two malls are shown in Figure 6. It clearly shows that O2OTM is significantly better than LDA. At Mall A, the improvement from O2OTM is

more clear when K is small, while this improvement becomes small after K reaches 15. At Mall B, the performances of O2OTM and LDA are similar when $K = 1$, however, when $K > 1$ the improvement from O2OTM becomes large. Also, the line for the method of RAN is flat, which is consistent with our analytical analysis.

7. RELATED WORKS

Our work is related to three areas, i) recommendation ii) topic modeling methods and iii) multi-modality learning.

Recommendation. Recommendation services have been deployed for many types of items such as books [14], music [7], mobile apps [26] and even point-of-interests [25]. Collaborative filtering (CF) is the well known recommendation approach with the core idea that users of similar interests may display similar ratings on similar items. Specifically, if a recommendation is made based on similar users (those who have the similar ratings on the same items as the target user has), it is referred to as user-based CF [10]. On the other hand, if a recommendation is made based on similar items (those that have been rated similarly by the same set of people), it is referred to as item-based CF [20]. Besides these nearest-neighbor techniques, matrix factorization was developed with the assumption that users and items are associated with latent factors and the observed user-item rating matrix is the multiplication of the two. These latent factors are learned by minimizing a loss function, which is defined as the sum-squared distance between observed and predicted ratings [22], or by maximizing the likelihood of the regression residual which is assumed to obey the standard Gaussian distribution [18].

To address the issue of data sparsity, different approaches have been developed to incorporate various features on users and items into recommendation. For example, social network was introduced to constrain the factorization process in such a way that socially connected people tend to have similar user feature vectors [13]. For the items with rich textual information, the fLDA model was proposed to combine matrix factorization with LDA [4]. To recommend scientific articles the CTR model was developed with the assumption that the latent factor of items is the sum of its document topic distribution and the latent offset vector [23]. Yin et al. [27] developed a model interpreting the silent viewing behaviors to “pseudo vote” to help recommending articles. Also, the general model of factorization machine [15] was proposed to introduce all kinds of features on both items and users for recommendation with promising performance.

However, all these related works seldom consider the recommendation from the multi-modality perspective, even when the two modalities are only weakly associated. For this cross-modality recommendation task, we explicitly consider the relationships between the two modalities and simultaneously perform topic modeling over the two modalities. The relationships help to not only make cross-modality transformation with clear interpretability, but also identify the associations with high lift values.

Topic modeling. PLSA [11] and LDA [5] are popular to find the topics over the co-occurrence matrix. Different variants of topic modeling were proposed to describe the extra features with the data. For example, the author topic model [17] was developed to consider the author information on the articles. The TOT model [24] was proposed to describe the topics with its booming time. These previous

works provide the basic idea of probabilistic generative process to describe the data we have, but none of them apply it for the cross-modality data.

Multi-modality learning. Multi-modality learning has attracted many attentions recently with the fast accumulation of the data with multiple modalities. Most of these previous works focus on the supervised or semi-supervised learning problems with the features from multiple modalities. Co-training [6] is the most famous method for this task. Recently, the importance of different modalities was also considered to extract the discriminative features for weak modality with the supervision from strong modality. Note that all these modalities are in the feature space. As to the cross-modality recommendation task in this study, the online modality is used as feature space while the offline modality is considered as labels for prediction.

8. CONCLUSIONS

In this paper, we study the problem of offline retailing recommendation based users' online behaviors. We propose the probabilistic model O2OTM, which performs the topic modeling for both the online and offline activities, and also build the relationships between the online and offline topics. These resultant relationships can directly support cross-modality recommendation, and also provide recommendation explanation. Additionally, the lift measure is developed to identify the pairs of online and offline topics, which have strong or even causally-related associations. The live experiments show that this cross-modality recommendation contains huge business impact. We emphasize that this model for cross-modality recommendation has wide application scenarios. For example, based on the search logs we can perform the recommendations for restaurants, mobile apps, tour sites for traveling, and so on. Also, we will make the proposed model scalable for large-scale data and build the general system for cross-modality recommendation.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 61473274, 91546122, 61573335), National High-tech R&D Program of China (863 Program) (No.2014AA015105), Guangdong provincial science and technology plan projects (No. 2015 B010109005).

9. REFERENCES

- [1] http://www.ripenecommerce.com/blog/ecommerce_survey#prettyPhoto.
- [2] http://www.pwccn.com/webmedia/doc/635687554653555153_consulting_delivering_o2o_jun2015_chi.pdf.
- [3] <http://ijcai-15.org/index.php/repeat-buyers-prediction-competition>.
- [4] D. Agarwal and B.-C. Chen. flda: matrix factorization through latent dirichlet allocation. In *WSDM*, 2010.
- [5] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *JMLR*, 2003.
- [6] A. Blum and T. Mitchell. Combining labeled and unlabeled data with co-training. In *COLT*, 1998.
- [7] H.-C. Chen and A. L. Chen. A music recommendation system based on music data grouping and user interests. In *CIKM*. ACM, 2001.
- [8] Y. H. Cho and J. K. Kim. Application of web usage mining and product taxonomy to collaborative recommendations in e-commerce. *Expert systems with Applications*, 2004.
- [9] G. Heinrich. Parameter estimation for text analysis. Technical report, Fraunhofer IGD, 2009.
- [10] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl. An algorithmic framework for performing collaborative filtering. In *SIGIR*, 1999.
- [11] T. Hofmann. Probabilistic latent semantic analysis. In *UAI*, 1999.
- [12] Y. Li, J. Hu, C. Zhai, and Y. Chen. Improving one-class collaborative filtering by incorporating rich user information. In *CIKM*, 2010.
- [13] H. Ma, H. Yang, M. R. Lyu, and I. King. Sorec: social recommendation using probabilistic matrix factorization. In *CIKM*, 2008.
- [14] R. J. Mooney and L. Roy. Content-based book recommending using learning for text categorization. In *Proceedings of the fifth ACM conference on Digital libraries*. ACM, 2000.
- [15] S. Rendle. Factorization machines with libFM. *ACM Trans. Intell. Syst. Technol.*, 2012.
- [16] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl. Grouplens: An open architecture for collaborative filtering of netnews. In *CSCW*, 1994.
- [17] M. Rosen-Zvi, T. Griffiths, M. Steyvers, and P. Smyth. The author-topic model for authors and documents. In *UAI*, 2004.
- [18] R. Salakhutdinov and A. Mnih. Probabilistic matrix factorization. In *NIPS*, 2007.
- [19] A. D. Sarma, S. Gollapudi, R. Panigrahy, and L. Zhang. Understanding cyclic trends in social choices. In *WSDM*, 2012.
- [20] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *WWW*, 2001.
- [21] G. Shani and A. Gunawardana. Evaluating recommendation systems. In *Recommender systems handbook*. 2011.
- [22] N. Srebro and T. Jaakkola. Weighted low-rank approximations. In *ICML*, 2003.
- [23] C. Wang and D. Bleig. Collaborative topic modeling for recommending scientific articles. In *KDD*, 2011.
- [24] X. Wang and A. McCallum. Topics over time: A non-markov continuous-time model of topical trends. In *KDD*, 2006.
- [25] M. Ye, P. Yin, W.-C. Lee, and D. L. Lee. Exploiting geographical influence for collaborative point-of-interest recommendation. In *SIGIR*, 2011.
- [26] P. Yin, P. Luo, W.-C. Lee, and M. Wang. App recommendation: a contest between satisfaction and temptation. In *WSDM*, 2013.
- [27] P. Yin, P. Luo, W.-C. Lee, and M. Wang. Silence is also evidence: interpreting dwell time for recommendation from psychological perspective. In *KDD*, 2013.

	Online words (English description)	Lift	Offline brands (English description)								
I	1. 狗狗 (dogs, puppies) 2. 泰迪 (<i>Teddy Bear Puppy</i> , a breed of small dog) 3. 宠物 (pet) 4. 狗粮 (dog food) 5. 绝育 (sterilization for pet) 6. 猫咪 (cats) 7. 宠物医院 (pet hospital) 8. 疫苗 (Pet vaccine) 9. 金毛 (<i>Golden Retriever</i> , a breed of large dog with thick yellow hair) 10. 拉布拉多 (<i>Labrador</i> , a breed of large dog)	11.5	1. TUTUANNA (a Japanese brand famous for lovely girl underwear) 2. PEACH JOHN (girl underwear brand of hot sexy and cute style) 3. INNISFREE (a Korean cosmetics brand) 4. 6IXTY SIGHT (women fashion brand, famous for women's lingerie, dresses) 5. SNIDEL (a Japanese girl clothing brand) 6. LILY BROWN (a Japanese girl clothing brand) 7. EARTH MUSIC (a Japanese girl clothing brand of simple and natural style) 8. ORBIS (a Japanese cosmetics brand) 9. The Body Shop (natural skin care, make-up brand) 10. Lowrys Farm (a Japanese girl clothing brand)								
	II		1. 游戏 (games) 2. 漫画 (comic series) 3. 怪物猎人 (<i>Monster Hunter</i> , a PSP game) 4. 盗墓笔记 (<i>The Secret of Grave Robber</i> , a net-novel) 5. 食尸鬼 (<i>Tokyo Ghoul</i> , a Japanese anime) 6. 火影忍者 (<i>Naruto</i> , a Japanese anime) 7. 最终幻想 (<i>Final Fantasy</i> , a role-playing video game) 8. 暴走 (<i>Rage Comic</i> , a comic series) 9. 战争 (wars) 10. 上古卷轴 (<i>The Elder Scrolls</i> , an action video game)	7.5	1. ADIDAS (a famous brand for sportswear) 2. NIKE (a famous brand for sportswear) 3. VANS (a brand for skate shoes and clothing for the young) 4. CONVERSE (a brand for sportswear and lifestyle footwear) 5. NIKE 360 (a classical sportswear series of NIKE) 6. ADIDAS NEO (a sports and leisure series of ADIDAS) 7. ADIDAS ORIGINAL (a classical sportswear series of ADIDAS) 8. LEVI'S (a famous brand for jeans and jackets) 9. PAUL FRANK (a fashion and lifestyle brand known for its bright color palette and stylish products) 10. MLB (a brand for baseball clothing)						
			III		1. 新生儿 (newborn baby) 2. 哺乳期 (suckling period) 3. 产后 (postpartum recovery) 4. 奶粉 (milk powder) 5. 母乳 (breast milk) 6. 睡觉 (sleep) 7. 大便 (baby stool) 8. 月子 (confinement in childbirth) 9. 剖腹产 (caesarean birth) 10. 满月 (babies' completion of its first month of life)	6.0	1. 乐友 (a brand special for newborn children related products) 2. ORCHESTRA (a children clothing brand) 3. TOYS R US (a kids store for all toys, video games, dolls, etc.) 4. MOTHERCARE (a brand for selected nursery accessories) 5. BABYLAND (a brand for children indoor interactive playground)				
						4.6	1. 6IXTY SIGHT (a women fashion brand, famous for young women's lingerie, dresses) 2. TEENIE WEENIE (a South Korean clothing brand special for the young women) 3. E-LAND (a South Korean clothing brand for the young women) 4. 阿童木 (a women and kids clothing brand) 5. Walker One (a brand for young women's shoes)				
					IV	1. 郑多燕 (a South Korean diet writer, famous for her fitness exercise) 2. 柔术 (human body flexibility exercise) 3. 天猫 (<i>TMall</i> , an online shopping site in China) 4. 奶粉 (milk powder) 5. 奇热 (a website for online TV series) 6. 钟汉良 (a famous TV series actor) 7. 新浪网 (<i>Sina</i> , a Chinese online media website) 8. 慢慢 (a shopping guide website) 9. 仙途 (a netnovel) 10. 生肖 (the Chinese zodiac)	7.6	1. 巴拉巴拉 (a brand special for children related products like children's shoes, toys and clothing) 2. OSHKOSH (an American brand for children's clothing) 3. 爱法贝 (a children clothing brand) 4. 法米尼 (a children clothing brand) 5. 斯乃纳 (a brand for pregnant women's clothing and care)			
							4.5	1. Columbia (an American outdoor sportswear brand) 2. Jack Wolfskin (a Germany outdoor sportswear brand) 3. 探路者 (a Chinese outdoor sportswear brand) 4. The North Face (an American outdoor sportswear brand, specializing in outerwear and equipment) 5. KAILAS (a brand in the manufacturing of technical climbing gear and apparel)			
						V	1. 武侠 (Chinese gongfu story series) 2. 文学 (literature) 3. 动漫 (comics and animations) 4. 激情 (emotional love novels) 5. 都市 (urban novels)	19.0	1. Dior (a French fashion brand famous for cosmetics, skin care, makeup and fragrances) 2. Chanel (a French fashion brand for perfume and bags) 3. BIOThERM (a brand of high-grade skin care products) 4. FANCL (a nature cosmetics brand from Japan)		
							1. 幼儿园 (kindergartens) 2. 双眼皮 (double fold eyelid) 3. 保卫萝卜 (<i>Carrot Fantasy</i> , a funny phone game) 4. 玻尿酸 (Hyaluronic Acid, used to keep skin elasticity and tension) 5. 整形 (plastic surgery)		5.8	5. DHC (a skincare brand from Japan) 6. Estee Lauder (An American skincare and makeup brand) 7. 周大福 (a Hongkong jewellery brand) 8. 安莉芳 (a Hongkong underwear brand) 9. L'Oreal (a famous French cosmetics, haircare brand) 10. Aupres (a cosmetics brand from Japan)	
							VI		1. 股票 (stocks) 2. 基金 (funds) 3. 投资 (investment) 4. 股份 (stock shares) 5. 证券 (securities) 6. 阿里巴巴 (<i>Alibaba</i> , a Chinese e-commerce company) 7. 金融 (finance) 8. 上市 (IPO) 9. 黄金 (gold) 10. 股价 (stock price)	9.7	1. 唐宫 (a fancy Chinese restaurant brand for Cantonese cuisines) 2. 熔宴 (a Chinese restaurant famous for Hunan cuisines) 3. 便宜坊 (a famous roast duck restaurant in Beijing) 4. 论道茶叶 (a Chinese tea brand) 5. 苗乡楼 (a Chinese restaurant famous for Guizhou cuisines) 6. 鹿港小镇 (a Chinese restaurant famous for Taiwan cuisines) 7. 蕉叶 (a restaurant famous for Thai food) 8. 恒香栈 (a Hongkong tea restaurant) 9. 釜山料理 (a Korean barbecue chain restaurant in China) 10. 至爱 (a Chinese restaurant for mixed cuisines)

Table 4: Example pairs of online and offline topics with the high lift values.