The eyes of the beholder: Gender prediction using images posted in Online Social Networks

Quanzeng You*, Sumit Bhatia†, Tong Sun‡, Jiebo Luo*

*Department of Computer Science, University of Rochester, Rochester, NY

{qyou, jluo}@cs.rochester.edu

†IBM Almaden Research Centre, San Jose, CA

subhatia@us.ibm.com

‡PARC, A Xerox Company, Webster, NY

Tong.Sun@xerox.com

***Abstract*—Identifying user attributes from their social media activities has been an active research topic. The ability to predict user attributes such as age, gender, and interests from their social media activities is essential for personalization and recommender systems. Most of the techniques proposed for this purpose utilize the textual content created by a user, while multimedia content has gained popularity in social networks. In this paper, we propose a novel algorithm to infer a user's gender by using the images posted by the user on different social networks.**

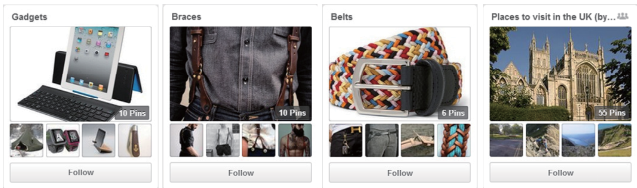
Keywords*—*Social Multimedia, Social Images, Gender Classification, Demographics

I. INTRODUCTION

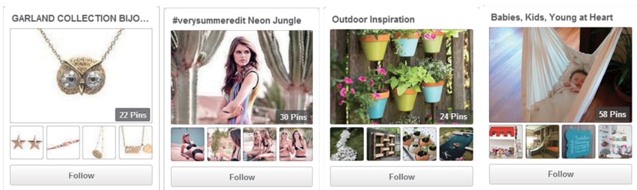
Online Social Networks (OSNs) such as Facebook, Twitter, etc. have become immensely popular with three out of every four adult internet users using at least one social networking site [1]. Such a large scale adoption and active participation of users has led to research efforts studying relationship between users' digital behavior and their demographic attributes such as age, gender, relationship status, etc.. Accurate techniques to predict these demographic attributes are useful for marketing purposes, and personalization and recommender systems. A large scale study of about 58,000 Facebook users performed by Kosinski et al. [2] reveals that digital records of human activity can be used to accurately predict a range of personal attributes such as age, gender, sexual orientation, political orientation, etc. Likewise, there have been numerous works that study variations in language used in social media with age, gender, personality, etc. [3], [4], [5]. While most of the popular OSNs studied in literature are mostly text based, some of them (e.g., Facebook, Twitter) also allow people to post images and videos. Recently, OSNs such as Instagram and Pinterest that are majorly image based have gained popularity with almost 20 billion photos already been shared on Instagram and an average of 60 million photos being shared daily [6].

Recent research efforts have provided indications that images posted by users on OSNs may prove to be useful to learn various personal and social attributes of users. Lovato et al. [7] proposed a method to learn users' latent preferences by extracting aesthetics and visual features from images favorited by users on Flickr. The learned models can be used to predict images likely to be favorited by the user on Flickr with reasonable accuracy. Cristani et al. [8] infer personalities of users by extracting visual patterns and features from images marked as favorites by users on Flickr. Can et al. [9] utilize the visual cues of tweeted images in addition to textual and structure-based features to predict the retweet count of the posted image. Motivated by these works, this paper investigates *if the images posted by users on online social networks can be used to predict their gender*. We frame gender classification as a binary classification problem (male and female categories) and evaluate the use of a variety of image based features.

Sumit Bhatia was at Xerox when this work was performed.



(a) Example pinboards from several different male users.



(b) Example pinboards from several different female users.

Figure 1: Example pinboards from Pinterest users of different genders.

In particular, we extract the features from users' posting behaviors and posted content. We utilize the images from users of Pinterest. Figure 1 shows some randomly selected pinboards for both male and female users. Even for those randomly chosen users, we can clearly see the difference between male and female users. For male users, they are more interested in electronics, buildings, mens clothes and so on. On the other hand, female users are more likely to post pinboards that are related to jewelry, womens clothes, gardening and so on. Based on these intuitive findings, for each user, we extract features from their collections of pins in a few different categories, such as travel, art, and technology. For posting behaviors, we mainly focus on the users' own labeled distribution of their collections of pins over the limited number of categories provided by Pinterest. Meanwhile, we employ visual topic model to obtain user-level visual topic distributions in different categories according to each user's posted images. Our results suggest that both posting behavior and posted content are beneficial for gender prediction. In particular, visual content provides competitive performance in terms of accuracy compared with the results from [10], which employ sociolinguistic model on user generated Tweets.

II. Related Work

In this section, we review works which are closely related to our motivation for online user gender prediction. Specifically, we focus on works in gender prediction and classification of visual personality.

A. Gender Prediction

Gender classification has been an interesting research topic. There has been related works which tried to predict gender using frontal facial images [11], [12], [13]. All of these works focus on using the features from facial images. Until recently, researchers started to focus on the gender prediction of online users by their generated content. Schler et al. [14] employed writing style and content based features for blogger users. Their results indicated that there are differences between male and female bloggers, which can be employed for gender and age classification. Meanwhile, Yan et al. [15] used additional profile related features, such as web page background colors and emoticons for gender classification of weblog users. The work in [16] further analyzed the stylistic differences between different age group of bloggers. They further proved that there stylistic features can be good at prediction of online users' geographical location and ethnic group.

With the popularity of online social networks, more and more researchers start to work on inferring online user attributes from informal and casual user generated content. In [10], a collection of sociolinguistic features are extracted from informal textual content (Tweets) and stacked-SVM-based algorithm is employed to classify different latent user attributes. The work in [3] further employed n-gram language model features from user profiles, including screen name, full name and description, to predict the gender of Twitter users. In this way, there are a total of 15 million distinct features in their model. Liu et al. [17] tried to estimate the gender distribution of commuting populations using different transportation. Their work suggested that gender classification may be beneficial for building smart cities. Different from the above works, which only use the user generated content, Al Zamal et al. [18] extends their works by leveraging the content from the users' friends. In other words, they use signals from Twitter user's neighborhood to infer the latent attributes of the current user. Their work suggested that features from neighbor's profiles and content are capable of predicting the current user's attributes. However, users' different neighbors may share different aspects with them. Thus, their methods may heavily rely on the choice of subset of neighbors. Recently, the authors from [19] evaluated the performance of using profile characteristics for Twitter user gender classification. Since they only utilized profile features, such as user name and different profile colors, their approach significantly reduced the number of features employed and achieved competitive performance with other existing approaches.

B. Visual Personality

Visual content becomes increasingly popular in all online social networks. Recent research works have indicated the possibility of using online user generated visual content to learn personal attributes. In [2], a total of $58,000$ volunteers provided their Facebook likes as well as detailed demographic profiles. Their results suggest that digital records of human activities can be used to accurately predict a range of personal attributes such as age, gender, sexual orientation, political orientation, etc. Meanwhile, the work in [20] tried to build the connection between cognitive effects and the consumption of multimedia content. Image features, including both aesthetics and content, are employed to predict online users' personality traits. The findings may suggest new opportunities for both multimedia technologies and cognitive science. More recently, Lovato et al. [21] proposed to learn users' biometrics from their collections of favorite images. Various perceptual and content visual features are proven to be effective in terms of predicting users' preference of images. The achievements of these works encourage us to exploit visual features for online gender prediction which has not been explored by other research works.

III. Proposed Approach

We frame the task of predicting users' gender from their posted images as a binary classification task. Given a set of images posted by a user on a social networking site, we predict whether the user is male or female. We posit that males and females differ in terms of their image *posting behavior* as well as in the *content* of posted images. We extract features to capture visual content of images as well as users' posting behavior as described in the following subsections.

A. User Posting Behavior

We define *posting behavior* of a user in terms of the types of images she choses to post. Intuitively, images posted by a person are representative of her interests and preferences. Often, males and females have inherently different interests and preferences and thus, images posted by them should be indicative of these differences in interests and preferences. For example, females (as a group) may post more images related to fashion whereas males (as a group) may post more images of bikes and cars. Often, at the time of posting, OSNs offer a medium to users to indicate the board category of the image(s) by means of tags, album names, category labels, etc. We use these associated tags/category labels to represent the posting behavior of users.

Pinterest, the OSN studied in this work, requires users to choose a category label from a list of 33 pre-defined labels for each pinboard they create. Table I lists the names of all the 33 category labels as provided by Pinterest. We assume that all the pins (images) in the same pinboard to have the same category label as that of the corresponding pinboard. Next, for a given user, we calculate the number of pins in each category and normalize these numbers to get a distribution of categories for the user. This way, we have a 33 dimensional *preference vector* for each user giving a category distibution for the images posted by the user. We use the category distibution thus obtained as 33 features, each feature measuring the fraction of images posted by the user in the corresponding category.

Table I: List of categories in Pinterest Boards.

Animals	Architecture	Art	Cars & Motorcycles	Celebrities	Design	DIY & Crafts
Education	Film, Music & Books	Food & Drink	Gardening	Geek	Hair & Beauty	Health & Fitness
History	Holidays & Events	Home Decor	Humor	Illustrations & Posters	Kids	Men's Fashion
Outdoors	Photography	Products	Quotes	Science & Nature	Sports	Tattoos
Technology	Travel	Weddings	Women's Fashion	Other		

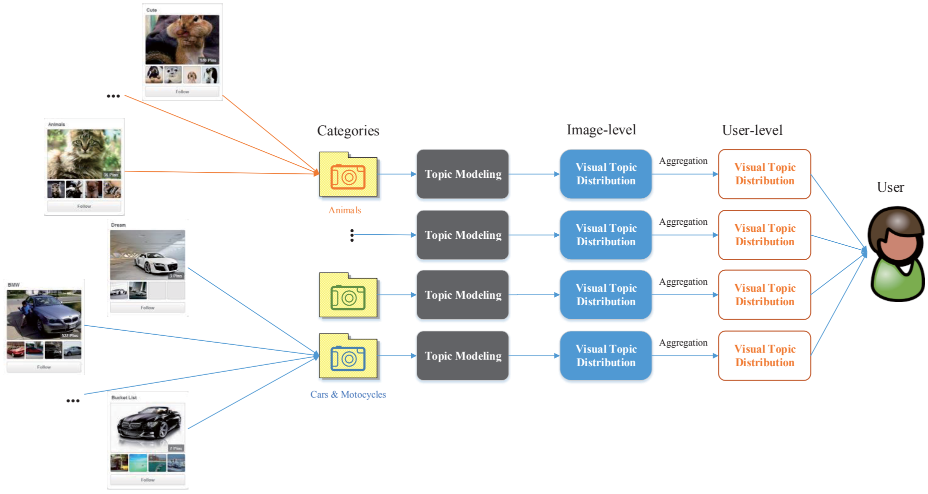


Figure 2: Framework for building user-level visual topic distribution in different categories.

B. Visual Content of Posted Images

In order to capture the differences in content of the images posted by males and females, we use Bag of Visual Words [22] model to represent the visual content of images posted by the users. Bag of visual words model, similar in principle to bag of words representation of text documents, represents an image as a vector of *visual words*. As the first step in constructing a bag of visual words representation, we need to define an appropriate *visual vocabulary*. Typically, a large number of local image features are used as visual words and the occurrence count of each such visual word (local image feature) in the image is used to create the bag of visual words vector for the image. We employ Scale-invariant feature transform (SIFT) [23] to discover local features for each image in the dataset. Visual words are discovered by clustering all the SIFT features and defining the centre of each cluster as a visual word. In this way, each image can be represented as the frequencies of visual words discovered. Each image posted by a user is thus represented as a vector of visual words. Next, for a fine-grained analysis, we discover the latent visual topic distribution of each user by employing probabilistic Latent Semantic Analysis (pLSA) [24]. In our implementation, we learn the visual topics on a per catehory basis. pLSA models the mixture distribution $p(i, v)$ of the image d and visual word v. By employing EM algorithm, the model can learn the topic distribution $p(z|i)$ of each image. Then, we aggregate the image-level topic distributions of all the images posted by the same user to obtain the user-level topic distribution. We set the number of visual topics to be 10 for all the categories. In other words, we learn a topic distribution for each image in all the categories. For each user, we aggregate the topic distributions in all different categories and normalize the distribution to obtain the user-level topic distribution in all different categories. Figure 2 summarizes the framework of extracting user-level visual topic distributions for different categories of images.

IV. Experiments

A. Data Preparation

In order to test the utility of using posting behavior and visual content based features for gender classification, we need a data set that contains the gender ground truth labels in addition to the images posted by the users. In order to obtain images posted by users, we use Pinterest which is one of the most popular image based OSN. Pinterest also allows its users to connect their accounts with other social network accounts, such as Twitter, Facebook, etc. and this information is generally available in users' Pinterest profiles. In order to obtain gender lables for the users, we took services of a commercial data vendor, Datasift[1] that provides access to data from Twitter and additional analytics services – such as users' gender label. Therefore, if we know the Twitter account of a Pinterest user, we can use Datasift to get the user's gender label.

Therefore, we crawled around 1500 such Pinterest users and extracted their corresponding Twitter profiles and queried Datasift to get the gender labels of these users . Since gender information is not available for all the users[2], we got labels only for a subset of users. From these users, we randomly selected a total of 160 users (80 females and 80 males) and crawled all the pinboards and pins for all of these 160

[1] http://datasift.com/

[2] depends on various factors such as users' privacy settings, confidence threshoulds employed by Datasift, etc.

pinterest users. Each of the pins in the same pinboard has the same category label, which is the category label of its pinboard. Table II shows the statistics of pins for these 160 users. In general, we see that females seems to be more active than males on Pinterest in all the 33 categories in terms of number of pins. Meanwhile, not every user will post images on every category. Different users may have different preferences towards different categories.

Table II: Statistics of number of pins for both males and females in different categories.

Category	Total		Average	
	Male	Female	Male	Female
Animals	1458	2155	18.2	26.9
Architecture	1405	3375	17.6	42.2
Art	4658	6456	58.2	80.7
Cars & Motorcycles	491	1007	6.1	12.6
Celebrities	3546	4693	44.3	58.7
Design	5035	10281	62.9	128.5
Diy & Crafts	5268	5461	65.9	68.3
Education	424	1058	5.3	13.2
Film, Music & Books	3146	4165	39.3	52.1
Food & Drink	6936	11074	86.7	138.4
Gardening	732	1844	9.2	23.1
Geek	598	324	7.5	4.1
Hair & Beauty	5043	6122	63.0	76.5
Health & Fitness	1413	2276	17.7	28.5
History	422	1126	5.3	14.1
Holidays & Events	1887	2492	23.6	31.2
Home & Decor	6040	14436	75.5	180.5
Humor	1732	1994	21.7	24.9
Illustrations & Posters	1175	1433	14.7	17.9
Kids	1964	2927	24.6	36.6
Mens & Fashion	3157	2005	39.5	25.1
Outdoors	395	2142	4.9	26.8
Photography	2693	4808	33.7	60.1
Products	2195	2197	27.4	27.5
Quotes	711	1057	8.9	13.2
Science & Nature	316	1819	4.0	22.7
Sports	550	738	6.9	9.2
Tatoos	127	652	1.6	8.2
Technology	685	966	8.6	12.1
Travel	2387	5793	29.8	72.4
Weddings	3378	5573	42.2	69.7
Women's Fashion	17645	21806	220.6	272.6
Other	8292	13422	103.7	167.8
Sum	95904	147677	1198.8	1845.96

B. Results

We evaluate the performance of different features using 10-fold cross validation. The performance is measured in terms of accuracy, precision, recall and F-Measure. All the results reported are obtained from simple logistic regression.

Table III summarizes the results of using features derived from user posting behavior. For each user, we count the number of images in each category. Next, we normalize the number of images in all categories. In this way, we build a distribution of the number of images over different categories. Each user is represented by a feature vector with length 33. The results show that posting behavior has better performance in female class in terms of recall. This may be due to the fact that females are more likely to post images than male, thus their posting behaviors are relatively stable compared with male users.

Table III: Performance of using posting behavior features.

Class	Accuracy	Precision	Recall	F-Measure
Female	0.763	0.629	0.763	0.689
Male	0.55	0.698	0.55	0.615
Avg	0.656	0.664	0.656	0.652

Table IV: Per-category accuracy for the task of gender classification.

Categories	Accuracy	Male	Female	Total
Tatoos	0.692	4	9	13
Diy & Crafts	0.618	32	36	68
Travel	0.612	30	37	67
Science & Nature	0.611	6	12	18
Education	0.6	8	12	20
Film, Music & Books	0.6	28	38	66
Animals	0.5897	20	19	39
Other	0.581	37	56	93
Outdoors	0.577	6	20	26
Illustrations & Posters	0.571	17	18	35
Quotes	0.571	9	12	21
Food & Drink	0.57	39	61	100
Home & Decor	0.563	41	55	96
Humor	0.548	19	23	42
Weddings	0.538	29	36	65
Hair & Beauty	0.537	34	48	82
Sports	0.529	6	11	17
Kids	0.519	22	30	52
Art	0.516	29	35	64
Health & Fitness	0.512	20	21	41
Photography	0.484	25	39	64
Women's Fashion	0.478	63	75	138
Products	0.468	22	25	47
Gardening	0.457	13	22	35
Celebrities	0.455	29	26	55
Design	0.441	32	36	68
Technology	0.435	11	12	23
Architecture	0.424	13	20	33
Mens & Fashion	0.411	30	26	56
Cars & Motorcycles	0.4	7	13	20
Holidays & Events	0.362	28	30	58
Geek	0.333	6	6	12
History	0.167	10	14	24

Meanwhile, Table IV shows the accuracy of using per-category visual topic features. The last three columns represent the number of users who own pinboards in different categories for male, female and total respectively. It is interesting to see that some categories such as Mean & Fashion and Women's Fashion cannot distinguish the two genders from the content. We argue that this may be due to that items in these categories are not highly distinguishable. While, males and females may prefer different kinds of images in other categories, such as Tatoos, Travels and so on. Table V shows the performance of using visual topic models from categories that have an accuracy of over 50% in Table IV. Before we use the simple logistic regression on these features, we first apply singular value decomposition (SVD) on the feature matrix for feature space reduction. We use the top 20 largest singular vectors as the new feature representation of posted content. Next, simple logistic regression is employed to evaluate the performance on these features. The results show that features derived from posted content perform better than features derived from user posting behavior. In particular, the performance on male users have been significantly improved. This observation provides support for our hypothesis that content analysis of images posted by males and females in OSN may help in gender prediction task.

We also evaluate the performance by combining both user posting behavior and visual content based features. The results are summarized in Table VI. We observe that the combining both types of features achieve better overall performance than that achieved by using either of posting behavior based features or content based features. We also note that the performance for male class is better than the performance of female class.

Table V: Performance of using posted content.

Class	Accuracy	Precision	Recall	F-Measure
Female	0.75	0.698	0.75	0.723
Male	0.675	0.73	0.675	0.701
Avg	0.713	0.714	0.713	0.712

Table VI: Performance of using both posting behavior and posted content.

Class	Accuracy	Precision	Recall	F-Measure
Female	0.688	0.733	0.688	0.71
Male	0.75	0.706	0.75	0.727
Avg	0.719	0.72	0.719	0.718

V. Conclusion

We proposed the problem of predicting user's gender based on images posted by her on an OSN. We framed the problem as a binary classification problem and evaluated features derived from users' posting behavior and visual content of the posted images. We achieved F-measure of around 72% using both types of features providing support to the hypothesis that there are differences between male and female users in terms of posting behavior and posted content. In our future work, we plan to extend our work to predict other user attributes such as age, user interests and preferences from user generated multimedia content. This can provide rich and valuable information for customized marketing and personalized recommendation systems.

Acknowledgment

This work was supported in part by an award from the Xerox Foundation.

References

[1] *Pew research internet project C Social Networking Fact Sheet.*, 2013 (accessed 10 June 2013). [Online]. Available: http://engineering.purdue.edu/ mark/puthesis

[2] M. Kosinski, D. Stillwell, and T. Graepel, "Private traits and attributes are predictable from digital records of human behavior," *PNAS*, vol. 110, no. 15, pp. 5802–5805, 2013.

[3] J. D. Burger, J. Henderson, G. Kim, and G. Zarrella, "Discriminating gender on twitter," in *EMNLP*, 2011, pp. 1301–1309.

[4] D. Bamman, J. Eisenstein, and T. Schnoebelen, "Gender identity and lexical variation in social media," *Journal of Sociolinguistics*, vol. 18, no. 2, pp. 135–160, 2014.

[5] H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, L. Dziurzynski, S. M. Ramones, M. Agrawal, A. Shah, M. Kosinski, D. Stillwell, M. E. Seligman *et al.*, "Personality, gender, and age in the language of social media: The open-vocabulary approach," *PloS one*, vol. 8, no. 9, 2013.

[6] *Instagram C press page.*, 2013 (accessed June 2013). [Online]. Available: http://instagram.com/press/

[7] P. Lovato, A. Perina, D. S. Cheng, C. Segalin, N. Sebe, and M. Cristani, "We like it! mapping image preferences on the counting grid." in *ICIP*, 2013, pp. 2892–2896.

[8] M. Cristani, A. Vinciarelli, C. Segalin, and A. Perina, "Unveiling the multimedia unconscious: Implicit cognitive processes and multimedia content analysis," in *ACM MM*. ACM, 2013, pp. 213–222.

[9] E. F. Can, H. Oktay, and R. Manmatha, "Predicting retweet count using visual cues," in *Proceedings of the 22nd ACM international conference on Conference on information and knowledge management*. ACM, 2013, pp. 1481–1484.

[10] D. Rao, D. Yarowsky, A. Shreevats, and M. Gupta, "Classifying latent user attributes in twitter," in *Proceedings of the 2nd international workshop on Search and mining user-generated contents*. ACM, 2010, pp. 37–44.

[11] B. Moghaddam and M.-H. Yang, "Gender classification with support vector machines," in *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*. IEEE, 2000, pp. 306–311.

[12] Z. Sun, G. Bebis, X. Yuan, and S. J. Louis, "Genetic feature subset selection for gender classification: A comparison study," in *Sixth IEEE Workshop on Applications of Computer Vision*. IEEE, 2002, pp. 165–170.

[13] E. Makinen and R. Raisamo, "Evaluation of gender classification methods with automatically detected and aligned faces," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, no. 3, pp. 541–547, 2008.

[14] J. Schler, M. Koppel, S. Argamon, and J. W. Pennebaker, "Effects of age and gender on blogging." in *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*, vol. 6, 2006, pp. 199–205.

[15] X. Yan and L. Yan, "Gender classification of weblog authors." in *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*, 2006, pp. 228–230.

[16] S. Goswami, S. Sarkar, and M. Rustagi, "Stylometric analysis of bloggers age and gender," in *Third International AAAI Conference on Weblogs and Social Media*, 2009.

[17] W. Liu, F. Al Zamal, and D. Ruths, "Using social media to infer gender composition of commuter populations," in *Int'l AAAI Conference on Weblogs and Social*, 2012.

[18] F. Al Zamal, W. Liu, and D. Ruths, "Homophily and latent attribute inference: Inferring latent attributes of twitter users from neighbors." in *ICWSM*, 2012.

[19] J. S. Alowibdi, U. A. Buy, and P. Yu, "Empirical evaluation of profile characteristics for gender classification on twitter," in *Machine Learning and Applications (ICMLA), 2013 12th International Conference on*, vol. 1. IEEE, 2013, pp. 365–369.

[20] P. Lovato, A. Perina, N. Sebe, O. Zandonà, A. Montagnini, M. Bicego, and M. Cristani, "Tell me what you like and ill tell you what you are: discriminating visual preferences on flickr data," in *Computer Vision–ACCV 2012*. Springer, 2013, pp. 45–56.

[21] P. Lovato, M. Bicego, C. Segalin, A. Perina, N. Sebe, and M. Cristani, "Faved! biometrics: Tell me which image you like and i'll tell you who you are," *Information Forensics and Security, IEEE Transactions on*, vol. 9, no. 3, pp. 364–374, March 2014.

[22] L. Fei-Fei and P. Perona, "A bayesian hierarchical model for learning natural scene categories," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2. IEEE, 2005, pp. 524–531.

[23] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, vol. 60, no. 2, pp. 91–110, 2004.

[24] T. Hofmann, "Probabilistic latent semantic indexing," in *SIGIR*. ACM, 1999, pp. 50–57.