Board Recommendation in Pinterest

Krishna Y. Kamath[1], Ana-Maria Popescu[2] and James Caverlee[1]

[1] Texas A&M University, College Station TX 77840, USA,
krishna.kamath@gmail.com, caverlee@cse.tamu.edu
[2] Research Consulting
anamariapopescug@gmail.com

Abstract. In this paper we describe preliminary approaches for content-based recommendation of Pinterest boards to users. We describe our representation and features for Pinterest boards and users, together with a supervised recommendation model. We observe that features based on latent topics lead to better performance than features based on user-assigned Pinterest categories. We also find that using social signals (repins, likes, etc.) can improve recommendation quality.

Keywords: recommendation, social network, interest network

1 Introduction

This paper focuses on the task of *recommending relevant boards to Pinterest users.* Pinterest is a fast-growing interest network with significant user engagement and monetization potential. One of the important aspects of Pinterest is encouraging pinning activity by recommending relevant, high-quality information to the site's users. We use a content-based filtering approach and report encouraging initial results. More specifically, we focus on three aspects of content recommendation for Pinterest boards. First, we describe our representation and features for Pinterest boards and users. Second, we describe our computation of potential board *relevance* to a user based on given features. Finally, we describe a supervised recommendation model which incorporates various relevance scores for good overall performance.

2 Related Work

Item recommendation is a well-studied problem [1]; general recommendation approaches include collaborative filtering [2], content-based filtering [11] or hybrid approaches [9]. Recently, recommender systems for users and content (tweets, topics, tags, etc.) in social networks have become an active area of interest [7, 8, 17, 14, 5, 15, 12]. Our work focuses on a particular recommendation task specific to Pinterest, a newer interest network, and leverages insights from both content-based filtering and from user modeling for social content recommendation. Pinterest is receiving additional attention from the research community, with recent work investigating other aspects such as global site analysis [3], gender roles and behaviors [13] and initial content quality measures [6].

3 Boards, Users and Board Relevance

In the following we describe our representation for Pinterest boards and users as well as our approach for assessing user-specific board relevance.
Let U be the set of Pinterest users, B the set of boards and $B_u \subseteq B$ be the set of boards created by user u. Each board is represented by means of a vector $\mathbf{b}$:

$$\mathbf{b} = \langle f_1, f_2, \ldots, f_n \rangle \tag{1}$$

where, $f_1, f_2, \ldots, f_n$ are features of $\mathbf{b}$ extracted from Pinterest data. Each user is represented as the mean of the board vectors for his set of boards B_u:

$$\mathbf{u} = \frac{1}{|B_u|} \langle \sum_{B_u} f_1, \sum_{B_u} f_2, \ldots, \sum_{B_u} f_n \rangle \tag{2}$$

To make sure this method of representing a user accurately captures his interests we exclude *community* boards from B_u. Pinterest users can turn a board into a *community* board by allowing others to pin to it. In previous experiments related to our recent work [6], we found that community boards have very high topical diversity and do not necessarily reflect the user's category-specific interest.
Given the above representation for a user u and board b, we can compute a measure of b's relevance to u by computing the cosine similarity between their corresponding vectors.

4 Feature Space

This section gives an overview of the features used to represent boards and users. We employ both *local* features (derived from a single board) and *global* features derived by leveraging a large set of boards.

4.1 Local Feature Extraction Methods

We use two methods to extract features directly from a given board by employing the user-supplied category label and, respectively, the board's pins.
Features From Board Category: When creating a board, users can assign to it one of 32 fixed categories (e.g., Art, Technology). Each board can be represented as a vector in a 32-dimensional space -e.g., a board in the Art category can be represented by a vector $\langle 1, 0, 0, \ldots, 0 \rangle$, where the first dimension corresponds to Art. A user vector is derived by combining board vectors as in (2).
Features From Pin Descriptions: Pins usually have free-text descriptions. A board can be represented as a vector using the bag-of-words model based on the content of the descriptions for all the board pins. Board vectors are again used to derived a final user vector as in (2).

4.2 Global Feature Extraction Methods

We next describe the use of information outside of a given board's content for feature extraction: (i) we account for Pinterest users interacting with a board and its owner; and (ii) we annotate a board with latent topics from a set learned from a collection of Pinterest boards.

Features From Social Interactions: We are interested in the social impact of a candidate board which may indicate the board is useful and recommendation worthy. We define the board *social score* as a linear function of its social impact (S_b) and the board user's social impact (S_u): $SocialScore(b) = w_b \cdot S_b + w_u \cdot S_u$. In later experiments we use $w_b = 0.9$ and $w_u = 0.1$. S_b is determined using social annotations from other users, in the form of repins, likes and follower count [3].

$$\begin{aligned} S_b = & \ w_{\text{re-pins}} \cdot \mathcal{F}(\text{mean re-repins for b}) \cdot \mathcal{F}(\text{std. re-repins for b}) + \\ & w_{\text{likes}} \cdot \mathcal{F}(\text{mean likes for b}) \cdot \mathcal{F}(\text{std. likes for b}) + \\ & w_{\text{followers}} \cdot \mathcal{F}(\text{\# of board followers}) \cdot \frac{\text{\# of board followers}}{\text{\# of user's followers}} + \\ & w_{\text{pins}} \cdot \mathcal{F}(\text{\# of pins on board}) \end{aligned}$$

where, $\mathcal{F}$ is a function maps which maps a real number to a value in $[0, 1]$ and the weights sum to 1. We experimented with logistic and double logistic functions for $\mathcal{F}$. Using this definition for S_b, we determine user's impact as:

$$\begin{aligned} S_u = & \ w_{\text{board scores}} \cdot [\text{Mean of social impact } (S_b) \text{ for all boards of } u\ ] + \\ & w_{\text{followers}} \cdot \mathcal{F}(\text{\# of user's followers}) + w_{\text{followers}} \cdot \mathcal{F}(\text{\# of boards}) \end{aligned}$$

Features From Latent Dirichlet Allocation (LDA): We previously described using Pinterest's board categories. However, users frequently skip the labeling step[4]. Additionally, generic categories (Outdoors, DIY & Crafts) lead to only a surface understanding of the board's content. These two reasons motivate us to also use features based on latent or hidden topics present in a board. Inspired by past work [18], we experiment with a *LDA-based topic discovery* method [19]. We generate one document per board by concatenating the board description, title, and pin descriptions. Topics are learned from a training set of 25,000 boards (> 9 pins each), and the learned model is used to label test boards. We compared LDA methods with two different values for number of topics - 100 and 200 and found that LDA with 200 (LDA-200) topics discovered latent topics on Pinterest better [6]. Hence, we used it to extract features from Pinterest to represent board vectors. Given a board, we first find board topics using LDA-200. We then represent the board as a vector in 200 dimensions, each for one topics in the LDA model. The user vector is then determined using (2).

5 Supervised Board Recommendation

We now describe our initial results for the task of recommending boards to Pinterest users. We describe our dataset, the *supervised board recommendation framework* and two sets of experiments.

[3] We include information about board size to penalize very sparse boards

[4] In our experience, with $> 290{,}000$ crawled boards, 47% lacked a user assigned category.

5.1 Data

For our analysis, we started with a sample of 4032 users and sampled $18,998$ of their boards. We then extracted features from these boards, using the four methods we described in Section 4, and built the corresponding board vectors. While using LDA-200 to discover topical features, we found that we could determine vectors for only $14,543$ (or 72%) of the boards. We analyzed the remaining boards and found that they were either very sparse (61% of the rest had at most 5 pins) or too incoherent; in some cases, topics outside of the learned set were required (e.g., a WWE board). Note that given the output of the LDA inference step for a test board, we only retain *core topics*, i.e. topics whose probability is greater than a threshold (0.05). Hence, for our experiments we used a dataset consisting of 4032 users and $14,543$ boards.

5.2 Supervised Board Recommendation

We now describe our board recommendation approach. Initially, we directly used the cosine similarity score to determine board-user similarity and recommend boards to users. However, this approach was not very effective, especially when combining different types of information (e.g., pin descriptions and LDA topics). Hence, we experimented with a supervised approach to board recommendation.

Generating Labeled Data: For scalability purposes, we automatically generated labeled data. We used a balanced data set with 50% positive and 50% negative recommendation examples. A second evaluation of a model trained on such data and used to produce recommendations judged *manually* will confirm the quality of the automatically derived labeled set. To obtain the labeled data, we first generate a set of similarity scores for each available (board, user) pair. Each corresponds to a class of basic features (e.g., LDA topics,etc.). We then select top$-k$ and bottom-k board-user pairs for each type of similarity score as *positive* and respectively *negative* examples. For each example in the final set, the attributes are represented by the similarity types (and their values by the similarity scores). Given a specific k, we generate a labeled dataset with $2k \times 4 = 8k$ labeled instances. For the experiments below, we set $k = 1000$ to generate a balanced set of 8000 recommendation examples.

Learning Recommendation Models: We employ the labeled data for learning recommendation models. We experimented with an SVM-based regression model; given a test example, the model will assign a score indicating a potentially good recommendation (if close to 1) or a bad recommendation (if close to 0). Potential board suggestions can be ranked according to the predicted score. For one of our evaluations we also used SVM-based classification to make a binary decision about a board being a good or bad suggestion for a user.

5.3 Experiments

We evaluate the value of the various feature classes (and their combinations) for board recommendation. In addition to methods testing the 4 feature classes in Section 4, we evaluated 2 other methods combining feature classes. The first, (*non-soc*), combines features based on board categories, pin descriptions and LDA topics, while the second (*all*) assesses the added impact of social features.

Table 1. Results: Feature classes' contributions to board recommendation quality. Combining feature classes and including social signals improves performance.

Method	F1	AUC	UIM	Compare
Board category (cat)	0.60 (0%, 1.00)	0.69 (0%, 1.00)	0.33 (0%, 1.00)	cat
Pin description (pin)	0.70 (17%, 0.00)	0.70 (1%, 0.05)	0.13 (-61%, 0.00)	cat
Social metrics (soc)	0.73 (22%, 0.00)	0.66 (-4%, 0.00)	0.12 (-64%, 0.00)	cat
LDA-200 topics (lda)	0.76 (27%, 0.00)	0.78 (13%, 0.00)	0.16 (-52%, 0.00)	cat
Non-social features non-soc: pin+cat+lda	0.83 (9%, 0.00)	0.84 (8%, 0.00)	0.22 (38%, 0.00)	lda
All features all: non-soc+ soc	**0.87 (5%, 0.00)**	**0.88 (5%, 0.00)**	**0.21 (-5%, 0.09)**	non-soc

We perform two types of experiments: (i) an evaluation using the automatically constructed 8000-example dataset and (ii) a second evaluation in which learned recommendation models are used to recommend boards for a small set of test users. The suggested boards are manually labeled and the various models are compared on this data.

Models: We compare 6 recommendation models. The first 4 models correspond to the 4 basic feature classes. For each such class, the resulting similarity score is used as a final aggregate feature by the model (e.g., *lda* only uses the similarity score based on LDA topics as basic features, etc.). Additionally, a mixed non-social model *non-soc* uses three similarity scores based on the pin descriptions, user-assigned categories and, respectively, latent topics. Finally, a full model *all* uses all 4 similarity scores. SVM classification is used in the first evaluation and SVM regression in the second.

Evaluation: Automatically Derived Gold Standard We start with an intrinsic evaluation using the automatically constructed balanced gold standard. We use SVM classification and the standard metrics F_1 and AUC. We also define another metric called *User Interest Match* (UIM) score which measures the match between a board labeled as *relevant* and the set of explicit user interests for u:

$$\text{UIM} = \frac{1}{|B_u^r|} \sum_{b \in B_u^r} \text{\% of boards with category } \mathcal{C}(b) \text{ in the account of user } u$$

where, B_u^r is the set of boards recommended to a user u. Higher UIM values correspond to recommended boards from categories of particular interest to the target user. We used Student's t-test to determine stat. significance for reported improvements.

Table 1 summarizes our results. In addition to assessing each method separately, we compare it with another relevant method (indicated in last column). Single feature class methods are compared against the *cat* baseline, *non-soc* against the best single class method (*lda*) and the final *all* method against *non-soc*. The first value in the parenthesis is the % improvement w.r. to the reference method

Table 2. Results: Board recommendation evaluation with human judgments. Combining feature classes leads to better recommendations.

Method	Precision@5	Precision@10	NDCG@5	NDCG@10
Board category (cat)	0.68	0.56	0.89	0.86
Pin description (pin)	0.62	0.60	0.92	0.90
Social metrics (soc)	0.37	0.40	0.77	0.66
LDA-200 topics (lda)	0.78	0.72	0.94	0.92
Non-social (non-soc)	**0.90**	0.81	**0.98**	**0.97**
All features (all)	**0.90**	**0.82**	0.97	0.96

and the second value is the p-value from t-test. We find that: (i) *lda* performs best among single feature class methods; (ii) combining feature classes leads to better performance than using single feature types; and (iii) social interaction information improves recommendation results.

Evaluation: Human judgments In a second experiment, we evaluate the recommendation models learned on automatically generated training data using manual judgments. We set aside a subset of our labeled dataset for testing purposes. We learned 6 SVM regression models on a balanced subset of the remaining data and then used them to make board recommendations for 12 users in the test set. Specifically, we retained the $top-10$ recommendations for each of the 12 users. 2 annotators independently labeled them as *relevant* or *not relevant* (with 70.5% agreement). After resolving disagreements, we used the manual judgments to evaluate the 6 models using precision (% of good recommendations) and the normalized discounted cumulative gain (NDCG), which takes into account the rank of the recommendation as well. Table 2 summarizes the results for the 6 models using top-5 and top-10 recommended boards.
Based the results in Table 2 and the human judgments, we find that: (i) Board category labels are helpful when accurate, but if absent or wrong they can hurt the similarity score relying on this feature. The latent topics discovered by LDA lead to better performance. (ii) Not surprisingly, using social signals by themselves leads to poor performance, as they do not contribute any topical relevance information. A user who liked a popular Wedding board may not like a popular Technology board. (iii) Methods which combine features perform best - the impact of social features was muted in this smaller-scope evaluation leading to small differences between the two.

6 Conclusion

This paper investigates content-based recommendation of Pinterest boards, with a focus on 4 classes of features user for board representation. Our initial experimental results show that latent topics discovered by LDA correspond to the most valuable single feature class, but combining different feature classes leads to best overall results. Our current work focuses on better ways of incorporating direct and indirect social network information in the recommendation model.

References

1. Adomavicius, G. and Tuzhilin,A.:Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions.In: IEEE Transactions on Knowledge and Data Engineering, 2005
2. Das, A.. , Datar M, Gard A., Rajaram S.: Google News Personalization: Scalable Online Collaborative Filtering. In: Proceedings of WWW 2007.
3. Gilbert, E., Bakhshi, S., Chang, S., Terveen, L.: I Need to Try This: A Statistical Overview of Pinterest. In: Proceedings of CHI-2013.
4. Hagberg A., Schult D., Swart P.: Exploring network structure, dynamics and function using NetworkX. In: Proceedings of SciPy2008, 2008.
5. Hannon, J., Bennet, M., Smyth, B., Recommending Twitter users to follow using content and collaborative filtering approaches. In: Proceedings of RecSys10.
6. Kamath, K., Popescu, A., Caverlee, J. : Board Coherence: Non-visual Aspects of a Visual Site. In: Proceedings of WWW 2013 (poster).
7. Kyew, S.M., Lim, E., Zhu, F. : A survey of recommender systems in Twitter. In: Proceedings of SocInfo 2012
8. Liang, H., Xu, Y., Tjondronegoro, D., Christen P.: Time-aware topic recommendation based on micro-blogs. In: Proceedings of CIKM 2012
9. Melville, P., Mooney, R., Nagarajan, R., Content-Boosted Collaborative Filtering for Improved Recommendations. In Proceedings of AAAI 2002
10. Liu, Lu., Tang, J., Han, J., Jiang, M., Yang, S.: Mining Topic-level influence in heterogeneous networks. In: Proceedings of CIKM 2010.
11. Pazzani, M.J., Billsus, D.: Content-based Recommendation Systems. In: The Adaptive Web: Methods and Strategies of Web Personalization., 2007
12. Pennacchiotti, M., Silvestri, F., Vahabi, H., Venturini, R.: "Making your interests follow you on Twitter". In: Proceedings of CIKM 2012.
13. Ottoni, R., Pesce, J.P., Las Casas, D., Franciscani, G., Kumaruguru, P., Almeida, V.: Ladies First: Analyzing Gender Roles and Behaviors in Pinterest. In: Proceedings of ICWSM 2013
14. Rae, A., Sigurbjrornsson, N., van Zwol, R., Improving Tag Recommendation using Social Networks. In: Proceedings of RIAO 2010.
15. Sharma, A.: Network-centric Recommendation: Personalization with and in social networks. In: SocialCom/PASSAT, 2011.
16. Wikipedia: http://en.wikipedia.org/wiki/Pinterest
17. Yan R., Lapata, M., Li, X.: Tweet recommendation with graph co-ranking. In: Proceedings of ACL 2012
18. Hong L., and Davison B. D.: Empirical study of topic modeling in Twitter. In: Proceedings of the First Workshop on Social Media Analytics
19. Phan X.: GibbsLDA++. https://gibbslda.sourceforge.net/