

Learning to Blend Vitality Rankings from Heterogeneous Social Networks

Jiang Bian^a, Yi Chang^a, Yun Fu^b, Wen-Yen Chen^b

^a*Yahoo! Labs, 701 First Ave, Sunnyvale, CA 94089*

^b*Yahoo! Inc., 701 First Ave, Sunnyvale, CA 94089*

Abstract

Heterogeneous social network services, such as Facebook and Twitter, have emerged as popular, and often effective channels for Web users to capture updates from their friends. The explosion in popularity of these social network services, however, has created the problem of “information overload”. The problem is becoming more severe as more and more users have engaged in more than one social networks simultaneously, each of which usually yields different friend connections and various sources of updates. Thus, it has made necessity to perform effective information filtering to retrieve information really attractive to web users from each of social networks and further blend them into a unified ranking list. In this paper, we introduce the problem of blending *vitality* rankings from heterogeneous social networks, where *vitality* denotes all kinds of updates user receives in various social networks. We propose a variety of content, users, and users correlation features for this task. Since vitalities from different social networks are likely to have different sets of features, we employ a divide-and-conquer strategy in order to fully exploit all available features for vitalities from each social network, respectively. Our experimental results, obtained from a large scale evaluation over two popular social networks, demonstrate the effectiveness of our method for putting vitalities that really interest users into higher orders in the blended ranking list. We complement our results with a thorough investigation of the feature importance and model selection with respect to both blending strategy and ranking for each social network.

Keywords: Social vitality, Ranking blending, Heterogeneous social networks

1. Introduction

Social network services on the Web are now emerging as a new medium of communication: users are enabled to compose and broadcast messages with various types, such as text, links, images, and videos, to their friends in the portal of social network. In contrast to traditional Web portals that publish well-formed and static Web content, social network services, such as Facebook ¹ and Twitter ², feature much more lightweight and real-time information, which usually includes status updates of friends, emerging news, and other contents that are interesting to the publisher and its friends. As the fast and convenient channel for information sharing, social network services have gained its explosive popularity among Web users. For example, Facebook has already had more than 500 million registered users at the end of 2010, and Twitter claims that it enjoys over 108 millions registered users as of April, 2010.

Email addresses: jbian@yahoo-inc.com (Jiang Bian), yichang@yahoo-inc.com (Yi Chang), yfu@yahoo-inc.com (Yun Fu), wychen@yahoo-inc.com (Wen-Yen Chen)

¹<http://facebook.com>

²<http://twitter.com>

Such explosion in popularity of those social network services, however, leads to the problem of “information overload”, namely, the sheer amount of information received by ordinary users can easily go beyond their processing capabilities. For instance, it is estimated that active Twitter users received over 300 tweets on average per day as of early 2010. This problem has been becoming more severe since there is growing body of users who actively engage in more than one popular social network services, simultaneously. Thus, to let users more efficiently surf on the Web, it has made necessity to introduce effective information filtering mechanism to identify information most interesting to web users from each social network, and taking one more step, to build a blending method for aggregating interesting information from heterogeneous social networks.

There have been several previous works studying information retrieval in the context of social networks, such as [1, 2]. However, most of them paid attention to only single social network, without considering blending various types of content from different social networks into one unified ranking. Inherently, it is quite a challenging problem. While on the surface many social networks look similar, each individual user has various friends connections and brings in quite different attitudes for obtaining information. Particularly, according to previous studies [3], users on Twitter tend to connect with someone they don’t know and are more interested in breaking news or new discovery; while users on Facebook usually connect with others they know and are more apt to post and see local events and issues needed feedback. Therefore, it is very hard to normalize the users’ interests on information updates from heterogeneous social networks, which makes it even harder to blend those various types of information into a unified ranking framework.

To address these problems, we propose a new learning framework for blending vitality rankings from heterogeneous social networks, where we use “*vitality*” to represent all various types of updates users receive from different social networks. In particular, we first generalize a couple types of features, describing the content of vitality, the characteristics of the vitality viewer, and the correlation between viewer and vitality poster, as the signals to imply viewer’s interests on the vitality. However, since different vitalities are not generated from the same social network service, there are a number of features which are good indicator for user’s interests on the vitality under one social network service, while they might be invalid in another social network. For example, the “*like*” behavior in Facebook is a strong signal to indicate that the user *likes* to see the vitality, but Twitter does not include this feature. In our paper, we address this challenge by employing a divide-and-conquer strategy, which fully exploits the available features for each individual vitality, respectively. By using this strategy, we can apply learning-to-rank based algorithms to obtain calibrated and comparable ranking scores for vitalities from different social networks, which then directly leads to a unified ranking list. The results of a large scale evaluation over two popular social networks demonstrate the effectiveness of our method for identifying user-interested vitalities from heterogeneous social networks and blending them into a unified ranking list. We also complement our results with a thorough investigation of the feature importance and model selection for this blending framework.

The main contributions of this paper can be summarized as: (1) Formalizing the problem of blending vitality rankings; (2) Extracting a couple types of features for implying users’ interests on vitalities; (3) A divide-and-conquer strategy for ranking and blending vitalities from heterogeneous social networks.

2. Related Work

The recent growth and popularity of online social network services such as Facebook, Twitter, etc., has lead to a surge in the research community. Much of this work has focused on analyzing the network structure and growth patterns. For example, [4] studied the evolution of network structure and group membership in MySpace. [5] studied the topological properties of the social network formed by Twitter users. And, [6] analyzed the relationship strength in the social network of Facebook and LinkedIn. Beyond link structure,

[2] studied how to identify influential users in Twitter. However, most of them focused on descriptive analysis and generative models of link structure, without demonstrating how to explore those features for information retrieval and filtering in online social network services. Most recently, [1] investigated using Twitter data to improve the effectiveness of real-time web search. But, they still did not address how to retrieve information from the inside of online social network. In this work, we will investigate both content and network structure in the context of social networks to extract a couple types of features, which are then used for retrieving and filtering content generated inside the social network. In this paper, we will address this problem by applying learning-to-rank techniques, which have been introduced to order things [7].

In recent years, the ranking problem is frequently formulated as a supervised learning problem. These learning-to-rank approaches are capable of combining different kinds of features to train ranking functions. Especially, pair-wise learning-to-rank approaches, including RankSVM [8], RankNet [9], RankBoost [10], and GBRank [11], have become very popular. They learn the ranking function from pair-wise preference data by minimizing the number of contradicting pairs in training data. Most recently, list-wise learning-to-rank approaches, such as ListNet [12], LambdaMart [13, 14], have been proposed and been shown effective in various ranking applications. In this paper, we will apply both pair-wise and list-wise approaches.

Furthermore, there are several studies which have discussed blend ranking or rank aggregation [15, 16]. But, they targeted the problem of merging the different rankings on the homogeneous set of items, i.e. items in their work belong to the same domain. Recently, some work on desktop search [17] also focused on ranking aggregation over multiple types of documents. Specifically, it first generated specialized ranking results for each document type, and then proposed a method to merge them into a single ranking list. Note that this work still need an extra merging algorithm to make the rankings scores of different document types be comparable. In this paper, we investigate how to blend rankings of vitalities from heterogeneous social networks, and we will propose a new divide-and-conquer strategy for blending rankings, which can transfer the relevance signals between heterogeneous social vitalities so as to ensure their respective ranking scores naturally calibrated and comparable.

3. Problem Statement

We now formalize the problem of blending vitality rankings from heterogeneous social networks. Currently, there are a couple of popular social network services on the Web, which are denoted as SN_1, SN_2, \dots, SN_k . These social networks are inherently heterogeneous in the sense that each of them consists of its own set of users and the corresponding network structure. Most recently, there is an increasing number of users who actively engage in more than one social networks. At a certain time point, the user can receive different sets of status updates or other information items from her friends in different social networks. In this paper, we use “*vitality*”, denoted as v , to represent each of these status updates or information items. Then, we denote the set of vitalities received by user u from one social network, e.g. SN_i , as $\mathcal{V}_i(u) = \{v_1^i, v_2^i, \dots, v_{|\mathcal{V}_i|}^i\}$. If user u does not engage into the social network SN_i , then $\mathcal{V}_i(u) = \emptyset$.

As a result, at a specific time point, we can represent the set of all the vitalities user u receives from heterogeneous social networks as $\mathcal{V}(u) = \mathcal{V}_1(u) \cup \mathcal{V}_2(u) \cup \dots \cup \mathcal{V}_k(u)$. Inspired by the overload of the new coming vitalities for the user that easily happens, the goal of this work is to identify those vitalities that are more important or interesting to the user, followed by one more essential step to blend all of them into a unified ranking list. We illustrate the big picture of the problem in Figure 1.

In this paper, we will fulfill this goal by introducing a learning based framework. Inherently, it is very challenging for building an effective learning based approach for filtering and blending vitalities from heterogeneous social networks: First, the heterogeneity of social networks usually induces that some good signals/features for indicating user’s interests on vitalities under one social network might not be valid

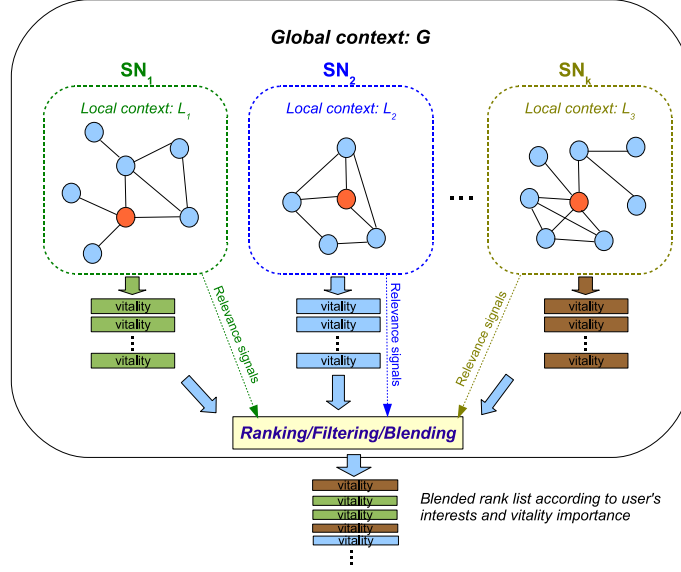


Figure 1: The general framework for filtering and blending vitalities from heterogeneous social networks (SN).

under another social network. Second, even if all vitalities can be normalized into a unified feature space, the range of one feature might be different with respect to different social networks. It is therefore difficult to build a ranking model which can compute calibrated and comparable ranking scores for vitalities from heterogeneous social networks. Furthermore, we need to find an effective way to extract the relevance judgment for each vitality which is then used as ground truth when learning the model. In this paper, we will attempt to address these problems by employing a divide-and-conquer strategy to fully exploit all available features for vitalities from heterogeneous social network, which will be discussed in details in the subsequent section.

4. Blending Vitality Rankings

We now introduce our learning based approach for blending vitality rankings from heterogeneous social networks. We focus on the specific characteristics of two heterogeneous social networks, Facebook and Twitter. Generally, we follow a learning to rank framework. Given a user u and the set of vitalities $\mathcal{V}(u)$ she receives from heterogeneous social networks at a specific time point, we first derive features for each user-vitality tuple $\langle u, v \rangle (v \in \mathcal{V}(u))$, (e.g. text of the vitality, user profile, correlation between user and vitality poster), as signals for predicting whether the user feels interested to see the vitality. Then, we take advantage of many types of user feedbacks in two social networks to infer preference judgments for the set of vitalities. After extracting features and preference judgments, we propose a divide-and-conquer strategy for learning the ranking models, which can address the problem of heterogeneous feature sets for different vitalities due to the heterogeneity of multiple social networks. In the following of this section, we will discuss each of these three aspects, respectively.

4.1. Feature Extraction

Our features are organized around the basic entities for each *user-vitality* tuple, including *vitality*, *user* (i.e. vitality viewer) as well as another latent one, *vitality poster*. We generalize features into a couple of groups, which are reviewed below.

- **Content Textual Features:** refers to those which are functions of the textual content of the vitalities, such as “*character length of the text*”, “*does text contain URL?*”, etc. We expect these features to be available in both Facebook and Twitter.

- **Vitality Non-Textual Features:** refers to those which describe non-textual characteristics of each vitality, such as “*vitality type in Facebook*”, “*number of existing comments in Facebook*”, “*number of retweets in Twitter*” and some temporal features, including “*lifetime of the vitality*” “*if the vitality consists of emerging popular words*”, etc. Availability of most of features in this group is dependent on the social network the vitality comes from.
- **Viewer Features:** refers to those which are functions of the viewer’s history in either Facebook or Twitter, such as “*number of friends*”, “*number of posted updates*”. Since our study focuses on the viewers have both Facebook and Twitter account, we can extract such features from both social networks no matter where the vitality comes from.
- **Vitality Poster Features:** refers to those which are functions of the vitality poster’s history in either Facebook or Twitter. Some of features in this group are available for both Facebook and Twitter, such as “*number of friends*”, while others are unique to only one social network, such as “*total number of received comments in Facebook*”, “*total number of received retweets in Twitter*”, etc.
- **Viewer-User Relationship Features:** refers to those which represent the communication, profile similarity, and mutual behaviors between the viewer and the user. Some of features in this group are available for both Facebook and Twitter, such as “*number of mutual friends*”, while others are unique to only one social network, such as “*similarity in terms of Facebook profile (age, gender, location, interests...)*”, “*number of mutual retweets in Twitter*”, etc.
- **Word Unigram Features:** beyond the above 5 types of features, we also derive word unigram features from the text of vitalities (from both Facebook and Twitter). As a simple feature selection method, only the most frequent 1000 words are included. And, we have excluded the stop-words.

Note that, social vitalities can be generated from heterogeneous social networks, but, since different social networks usually yield similar structures and utilities, social vitalities from heterogeneous social networks can still have a certain fraction of mutual features. In this paper, more than 60% features as introduced above are valid and available for both Facebook and Twitter.

4.2. Preference Extraction

In social network services, there are several other kinds of user behaviors beyond posting vitalities. For example, users can reply or provide evaluation to the vitality. All these behavior can imply that the user is interested in the content of the vitality. Under different social network services, the user behaviors can also be diverse. Particularly, in Facebook, after browsing the list of received vitalities, user can express her interests on it by “*commenting*” a sentence, clicking the “*like*” button, or “*sharing*” it again; while in Twitter, user can express her interests by “*retweeting*” or “*replying*”.

In this paper, we examine such users behaviors data to extract a set of preference judgment. In particular, for each user u and her received vitality list from one social network $\mathcal{V}_i(u)$, ($i = 1, \dots, k$) at a specific time point (we use Facebook as the example), if there is a pair of vitalities, (v_1, v_2) , where v_1 was “*commented*”, “*liked*”, or “*shared*” while v_2 was not, we then say v_1 is preferred over v_2 , denoted by $v_1 \succ v_2$. Similarly, we use the corresponding user behaviors “*retweeting*” and “*replying*” in Twitter to extract preference. All the extracted preference information will be directly integrated into the pairwise learning-to-ranking approach.

Note that, various user behaviors might infer different levels of user interests, and this problem is out of the range of this paper. We will explore it in future. Also, note that, it is not necessary to extract preference

Algorithm 1 Ranking Vitalities Using Union Feature Space

Data: \mathcal{D} : the whole training data set

$\mathcal{D}_{\text{facebook}}$: the training data set from Facebook

$\mathcal{D}_{\text{twitter}}$: the training data set from Twitter

Tools: TRAINMLR(\mathcal{D}, \mathcal{F}): learning to rank algorithm based on training set \mathcal{D} using feature space \mathcal{F} .

TRAIN-MODEL:

1. $\mathcal{F}_{\text{union}} = \mathcal{F}_{\text{mutual}} \cup \mathcal{F}_{\text{facebook}} \cup \mathcal{F}_{\text{twitter}}$
 2. Map each vitality into $\mathcal{F}_{\text{union}}$, with setting 0 value for invalid features.
 3. $\mathcal{M}_{\text{union}} \leftarrow \text{TRAINMLR}(\mathcal{D}, \mathcal{F}_{\text{union}})$
-

between vitalities from different social networks, as our learning method described in the following will ensure the predicted scores of vitalities from heterogeneous social networks naturally calibrated and comparable, even without explicit preference between vitalities from different social networks during training.

4.3. Ranking Models

Due to the heterogeneity of multiple social networks, even though there are a large fraction of overlapping features for vitalities from distinct social networks, a set of features available for vitalities from Facebook are missing for those from Twitter, and vice versa. Formally, according to above introduction to feature extraction, we can divide the whole feature space \mathcal{F} into three subsets: $\mathcal{F}_{\text{mutual}}$ (including features available for both Facebook and Twitter), $\mathcal{F}_{\text{facebook}}$ (including features available only for Facebook), $\mathcal{F}_{\text{twitter}}$ (including features available only for Twitter).

If we train separate ranking models for vitalities from Facebook and Twitter respectively, it cannot ensure that the ranking scores of all vitalities are calibrated and comparable for a direct blending. The straightforward method to address this problem is to map all vitalities from different social networks into the same feature space. As shown in Algorithm 1, we union all features together to form up a new feature space, $\mathcal{F}_{\text{union}}$. We then map each vitality into $\mathcal{F}_{\text{union}}$. For those features that are invalid for the vitality, we simply set the value as 0.

Since Algorithm 1 sets many features as 0 value, especially those Facebook- or Twitter-specific features, it still cannot fully exploit all available features for learning the ranking model. In contrast, it even reduces the effects of those Facebook- or Twitter-specific features. However, many of such features are important signals to indicate users' interests under the corresponding social network. To address this problem, we further propose a divide-and-conquer strategy, which can be summarized as Algorithm 2.

For this new algorithm, in Step 1 and 2, we learn separate ranking models, $\mathcal{M}_{\text{facebook}}$ and $\mathcal{M}_{\text{twitter}}$, using vitalities from Facebook and Twitter, respectively. Step 3 learns a ranking model $\mathcal{M}_{\text{mutual}}$ using data from both Facebook and Twitter but with only the features available for both social networks. Then, by using $\mathcal{M}_{\text{mutual}}$, we can map Facebook vitalities $\mathcal{D}_{\text{facebook}}$ and Twitter vitalities $\mathcal{D}_{\text{twitter}}$ into respective scores with only using mutually available features, as stated in Step 4 and 5. After that, we can use these scores as additional features to learn composite ranking models for two social networks, $\mathcal{M}_{\text{facebook_comp}}$ and $\mathcal{M}_{\text{twitter_comp}}$. Note that each individual of these two composite models has leveraged data from both Facebook and Twitter.

To rank vitalities (both vitalities from Facebook and those from Twitter) for the certain user, we apply $\mathcal{M}_{\text{facebook_comp}}$ and $\mathcal{M}_{\text{twitter_comp}}$ to Facebook vitalities and Twitter ones respectively. Then, we can obtain the ranking by sorting their ranking scores. Since these two models are both trained based on all data, the predicted grades are naturally calibrated and comparable. As a result, we can directly blend Facebook vitalities and Twitter vitalities according to their ranking scores.

Algorithm 2 A Divide-and-Conquer Strategy for Blending Vitality Rankings

Data: \mathcal{D} : the whole training data set

$\mathcal{D}_{\text{facebook}}$: the training data set from Facebook

$\mathcal{D}_{\text{twitter}}$: the training data set from Twitter

Tools: TRAINMLR(\mathcal{D}, \mathcal{F}): learning to rank algorithm based on training set \mathcal{D} using feature space \mathcal{F} .

PREDICT(\mathcal{D}, \mathcal{M}): compute the ranking scores for the dataset \mathcal{D} using model \mathcal{M} .

TRAIN-MODEL:

1. $\mathcal{M}_{\text{facebook}} \leftarrow \text{TRAINMLR}(\mathcal{D}_{\text{facebook}}, \{\mathcal{F}_{\text{mutual}}, \mathcal{F}_{\text{facebook}}\})$
 2. $\mathcal{M}_{\text{twitter}} \leftarrow \text{TRAINMLR}(\mathcal{D}_{\text{twitter}}, \{\mathcal{F}_{\text{mutual}}, \mathcal{F}_{\text{twitter}}\})$
 3. $\mathcal{M}_{\text{mutual}} \leftarrow \text{TRAINMLR}(\mathcal{D}, \mathcal{F}_{\text{both}})$
 4. $\mathbf{y}_{\text{facebook}} \leftarrow \text{PREDICT}(\mathcal{D}_{\text{facebook}}, \mathcal{M}_{\text{mutual}})$
 5. $\mathbf{y}_{\text{twitter}} \leftarrow \text{PREDICT}(\mathcal{D}_{\text{twitter}}, \mathcal{M}_{\text{mutual}})$
 6. $\mathcal{M}_{\text{facebook_comp}} \leftarrow \text{TRAINMLR}(\mathcal{D}_{\text{facebook}}, \{\mathbf{y}_{\text{facebook}}, \mathcal{F}_{\text{facebook}}\})$
 7. $\mathcal{M}_{\text{twitter_comp}} \leftarrow \text{TRAINMLR}(\mathcal{D}_{\text{twitter}}, \{\mathbf{y}_{\text{twitter}}, \mathcal{F}_{\text{twitter}}\})$
-

Note that our propose algorithm is quite general, and it can apply any state-of-the-art learning to rank method as TRAINMLR. In this paper, we employ a pair-wise algorithm, Gradient Boosted Decision Tree (GBDT) [18], and a list-wise algorithm, ListNet [12], to learn the ranking function for TRAINMLR.

5. Experimental Setup

5.1. Datasets

User set: To collect vitalities from Facebook and Twitter, we first collect a set of users who have registered on both Facebook and Twitter. This user set is obtained from a commercial online portal service which can allow users to integrate their Facebook and Twitter accounts into one single portal. All users have been anonymized, and each user is represented using a ID without any meaning. From the whole set of users, we select a subset by filtering those who have no behavior in consecutive two days. By doing that, we keep the users who are active on both Facebook and Twitter. From this subset, we totally sample 5000 users.

User-vitality tuples: Our dataset was collected in order to simulate a user’s experience within this online portal service. For one user, when she accesses this portal service, we collected 20 most recent vitalities from Facebook and Twitter, respectively. Then, these 40 vitalities form up one *user access session*. One user could access this online portal service for several times one day. To avoid duplicate vitalities in different sessions for one user, we only record one session for one user per day. In our experiment, we record the first one for those users who have more than one access in one day. And, we totally collect one week data in Jan, 2011. In our experiment, we use the first 5 days data as training set, the 6th day data for validation, and the 7th day data as testing set.

Relevance judgments: In our experiment, all user-vitality tuples are labeled automatically based on user behaviors information. In particular, for one user-vitality tuple, we check whether this user have taken any behavior (“like”, “comment”, “retweet”, etc) on this vitality two days later after we collect this tuple. If there has been any behavior, we label that this user is “interested” in this vitality, otherwise, we label it as “not-interested”.

Note that, we further filter those user access session which does not have any labeled “interested” tuple. Finally, we have totally 29642 user access sessions, where 21089 sessions are used as training set, 4256 are used for validation, and the other 4297 are used as testing set. In Table 1 demonstrates the ratio of both “interested” and “not-interested” vitalities in training, validation, and testing dataset.

Table 1: Data distribution in sense of relevance labeling for vitalities

Dataset	interested	not-interested
Training	19.8%	80.2%
Validation	21.2%	78.8%
Testing	20.7%	79.3%

5.2. Evaluation Metrics

We adapt the following information retrieval metrics to evaluate the performance of blending vitality rankings

Mean Reciprocal Rank(MRR): The MRR of each session is the reciprocal of the rank at which the first *interested* vitality was returned, or 0 if none of the top N results contained a user *interested* one. The score for a sequence of sessions is the mean of the individual session’s reciprocal ranks. Thus, MRR is computed as $MRR = \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} \frac{1}{r_s}$, where \mathcal{S} is a set of sessions, r_s is the rank of the first *interested* vitality in session s .

Precision at K: for a given session, $P@K$ reports the fraction of vitalities ranked in the top K results that are labeled as *interested*. This metric measures overall user potential satisfaction with the top K results.

Mean Average of Precision(MAP): Average precision for each session is defined as the mean of the precision at K values calculated after each user *interested* vitality was retrieved. The MAP value is defined as the mean of average precisions of all sessions in the test set, i.e. $MAP = \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} \frac{\sum_{r=1}^N (P@r \cdot rel(r))}{|R_s|}$, where R_s is the set of *interested* vitalities for session s , r is the rank, N is the number of retrieved vitalities, $rel()$ is a binary function on the “*interested*” of a given rank.

5.3. Compared Methods

In our study, we compare our proposed approach with several baseline methods, as listed in Table 2.

Table 2: Compared methods in the experiments

Method	Description
(Time, Time)	Rank Facebook and Twitter vitalities based on timestamp, and blend two rankings using round-robin method.
$(\mathcal{M}_{\text{union}}, \mathcal{M}_{\text{union}})$	(Baseline) All the vitalities are represented using the union feature space. Use $\mathcal{M}_{\text{union}}$ to compute ranking scores for all vitalities, and then blend them directly based on calculated scores.
$(\mathcal{M}_{\text{facebook}}, \mathcal{M}_{\text{twitter}})$	Use $\mathcal{M}_{\text{facebook}}$ and $\mathcal{M}_{\text{twitter}}$ to compute ranking scores for facebook and twitter vitalities, respectively, and then blend them directly based on calculated scores.
$(\mathcal{M}_{\text{mutual}}, \mathcal{M}_{\text{mutual}})$	Use $\mathcal{M}_{\text{mutual}}$ to compute ranking scores for both facebook and twitter vitalities, and then blend them directly based on calculated scores.
$(\mathcal{M}_{\text{facebook}}, \mathcal{M}_{\text{mutual}})$	Use $\mathcal{M}_{\text{facebook}}$ and $\mathcal{M}_{\text{mutual}}$ to compute ranking scores for facebook and twitter vitalities, respectively, and then blend them directly based on calculated scores.
$(\mathcal{M}_{\text{mutual}}, \mathcal{M}_{\text{twitter}})$	Use $\mathcal{M}_{\text{mutual}}$ and $\mathcal{M}_{\text{twitter}}$ to compute ranking scores for facebook and twitter vitalities, respectively, and then blend them directly based on calculated scores.
$(\mathcal{M}_{\text{facebook_comp}}, \mathcal{M}_{\text{twitter_comp}})$	Use $\mathcal{M}_{\text{facebook_comp}}$ and $\mathcal{M}_{\text{twitter_comp}}$ to compute ranking scores for facebook and twitter vitalities, respectively, and then blend them directly based on calculated scores.

6. Experimental Results

6.1. Blending Vitality Rankings

In this experiment, we train all kinds of ranking models based on the training dataset, with parameter tuning on the validation dataset. Then, we test the respective results on the remainder hold-out testing dataset.

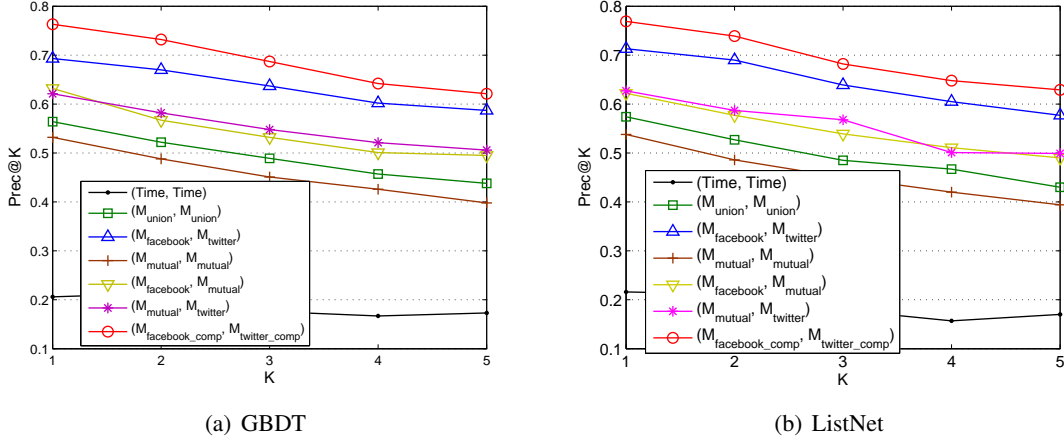


Figure 2: Precision at K for all the compared methods as listed in Table 2 using GBDT or ListNet

Figure 2(a) and 2(b) illustrate the Precision at K of all compared methods as listed in Table 2, where we apply GBDT in Figure 2(a) and ListNet in Figure 2(b). We also illustrate the MAP and MRR scores for all compared methods in Table 3. From these figures and tables, we can observe that ListNet can reach a little better performance than GBDT. More importantly, we can find that $(\mathcal{M}_{\text{facebook_comp}}, \mathcal{M}_{\text{twitter_comp}})$ can reach better performance than other methods. In particular, by using ListNet, the Precision@1 of $(\mathcal{M}_{\text{facebook_comp}}, \mathcal{M}_{\text{twitter_comp}})$ can reach 77%, compared to 71% Precision@1 exhibited by $(\mathcal{M}_{\text{facebook}}, \mathcal{M}_{\text{twitter}})$, 54% exhibited by $(\mathcal{M}_{\text{mutual}}, \mathcal{M}_{\text{mutual}})$, and 57% obtained by $(\mathcal{M}_{\text{union}}, \mathcal{M}_{\text{union}})$. Table 3 also demonstrates that $(\mathcal{M}_{\text{facebook_comp}}, \mathcal{M}_{\text{twitter_comp}})$ can perform better than the other methods. After conducting t-test in terms of MAP with respect to both learning algorithms, we find that the improvements of $(\mathcal{M}_{\text{facebook_comp}}, \mathcal{M}_{\text{twitter_comp}})$ over other methods are statistically significant ($p\text{-value} < 0.03$).

Table 3: MRR and MAP for all compared methods

Method	MRR (GBDT)	MAP (GBDT)	MRR (ListNet)	MAP (ListNet)
$(\mathcal{M}_{\text{union}}, \mathcal{M}_{\text{union}})$	0.612	0.424	0.617	0.426
$(\mathcal{M}_{\text{facebook}}, \mathcal{M}_{\text{twitter}})$	0.782	0.485	0.790	0.488
$(\mathcal{M}_{\text{mutual}}, \mathcal{M}_{\text{mutual}})$	0.586	0.389	0.585	0.387
$(\mathcal{M}_{\text{facebook}}, \mathcal{M}_{\text{mutual}})$	0.662	0.441	0.670	0.442
$(\mathcal{M}_{\text{mutual}}, \mathcal{M}_{\text{twitter}})$	0.670	0.453	0.675	0.455
$(\mathcal{M}_{\text{facebook_comp}}, \mathcal{M}_{\text{twitter_comp}})$	0.814	0.502	0.821	0.507

6.2. Feature Importance

From the above experimental results, we have demonstrated that using Facebook- or Twitter-specific features ($\mathcal{F}_{\text{facebook}}$ or $\mathcal{F}_{\text{twitter}}$) can significantly boost the performance of blending vitality rankings over the baseline method. It is thus worth investigating which features are highly valued by those compared ranking models, as presented in Algorithm 2. We rank features by the descending order of the importance in ListNet and show the top five in Table 4.

From Table 4, we can find that, for both $\mathcal{M}_{\text{facebook_comp}}$ and $\mathcal{M}_{\text{twitter_comp}}$, the composite features $\mathbf{y}_{\text{facebook}}$ and $\mathbf{y}_{\text{twitter}}$ have played an important role in ranking. If we learn separate ranking models for Facebook and Twitter vitalities, i.e. $\mathcal{M}_{\text{facebook}}$ and $\mathcal{M}_{\text{twitter}}$, some overlapping important features will replace the roles of composite features. This table can also indicate that different sets of features are more important for vitalities from different social networks. However, for the baseline approach $\mathcal{M}_{\text{union}}$, since all vitalities are mapped into one feature space and are used together to learn the ranking model, the ranking

Table 4: The top-5 most important features for compared ranking models.

$\mathcal{M}_{\text{facebook_comp}}$	$\mathcal{M}_{\text{twitter_comp}}$	$\mathcal{M}_{\text{facebook}}$	$\mathcal{M}_{\text{twitter}}$	$\mathcal{M}_{\text{union}}$
whether the facebook vitality type is photo?	whether the twitter vitality type is "retweet"?	does vitality text contain viewer's username?	whether the twitter vitality type is "retweet"?	number of mutual communicated vitalities during a certain time
$\mathcal{Y}_{\text{facebook}}$	whether the twitter vitality contain "@viewer's username"?	whether the facebook vitality type is photo?	does vitality text contain viewer's username?	does vitality text contain viewer's username?
number of facebook comments the viewer post during a certain time	$\mathcal{Y}_{\text{twitter}}$	number of mutual communicated vitalities during a certain time	whether the twitter vitality contain "@viewer's username"?	one popular unigram word feature
number of existing "like" for the facebook vitality	number of "retweet" the viewer post during a certain time	number of emotional symbols	number of mutual friends	one popular unigram word feature
number of existing facebook comments for the facebook vitality	number of existing retweet for the twitter vitality	number of facebook comments the viewer post during a certain time	number of mutual communicated vitalities during a certain time	number of emotional symbols

performance depends on more on overlapping features, while the effects of Facebook- or Twitter-specific features are diluted.

Based on our experiments, we also find that word unigram features do not yield much contribution to vitality ranking except $\mathcal{M}_{\text{union}}$. Table 5 illustrates the MRR and MAP values for all compared methods by using GBDT with or without word unigram features. From the table, we can find that the removing word unigram features results in very little decreasing in MRR and MAP except $\mathcal{M}_{\text{union}}$. After computing the feature importance, we find that some words, such as "birthday", "deals", yield much contribution to the performance of $\mathcal{M}_{\text{union}}$, which indicates some words can imply users' interest when there is few complex features. And, we can observe the similar results when using ListNet.

Table 5: MRR and MAP for all compared methods with and without word unigram features

Method	MRR	MRR (w/o unigram)	MAP	MAP (w/o unigram)
$(\mathcal{M}_{\text{union}}, \mathcal{M}_{\text{union}})$	0.612	0.502	0.424	0.338
$(\mathcal{M}_{\text{facebook}}, \mathcal{M}_{\text{twitter}})$	0.782	0.779	0.485	0.483
$(\mathcal{M}_{\text{mutual}}, \mathcal{M}_{\text{mutual}})$	0.586	0.575	0.389	0.380
$(\mathcal{M}_{\text{facebook}}, \mathcal{M}_{\text{mutual}})$	0.662	0.661	0.441	0.441
$(\mathcal{M}_{\text{mutual}}, \mathcal{M}_{\text{twitter}})$	0.670	0.667	0.453	0.451
$(\mathcal{M}_{\text{facebook_comp}}, \mathcal{M}_{\text{twitter_comp}})$	0.814	0.811	0.502	0.501

6.3. Discussions

To complement the empirical finding in this section, we will provide some insightful discussions that further justify the advantage of our new strategy for blending vitality rankings.

The baseline approach, $(\mathcal{M}_{\text{union}}, \mathcal{M}_{\text{union}})$, maps all vitalities into a union feature space and learns a single model for both Facebook and Twitter vitalities. This causes many 0-valued features, especially many of those in $\mathcal{F}_{\text{facebook}}$ or $\mathcal{F}_{\text{twitter}}$ which are specific for individual social network. This will accordingly reduce the effects of these special features in $\mathcal{M}_{\text{union}}$. The approach using only overlapped features, $(\mathcal{M}_{\text{mutual}}, \mathcal{M}_{\text{mutual}})$, even underperforms the baseline approach, because it does not use any Facebook- or Twitter-specific features, which are important signals for indicating users' interests in the context of their associated social networks. As observed, $(\mathcal{M}_{\text{facebook}}, \mathcal{M}_{\text{mutual}})$ and $(\mathcal{M}_{\text{mutual}}, \mathcal{M}_{\text{twitter}})$ can improve the performance over the baseline approach, since they leverage the representational strength of one specific type of vitalities. But, the absence of specific features for the other type of vitalities still hurts the ranking and blending performance. Since $(\mathcal{M}_{\text{facebook_comp}}, \mathcal{M}_{\text{twitter_comp}})$ leverages all specific features to enrich the representation of both types of vitalities, it can reach better performance.

Furthermore, we take more discussions on comparison between $(\mathcal{M}_{\text{facebook_comp}}, \mathcal{M}_{\text{twitter_comp}})$ and $(\mathcal{M}_{\text{facebook}}, \mathcal{M}_{\text{twitter}})$. As shown in Table 6, $(\mathcal{M}_{\text{facebook_comp}}, \mathcal{M}_{\text{twitter_comp}})$ can outperform $(\mathcal{M}_{\text{facebook}}, \mathcal{M}_{\text{twitter}})$.

$\mathcal{M}_{\text{twitter}}$) in terms of MRR and MAP metrics, and the t-test illustrates that the improvement are statistically significant (p-value < 0.04).

Table 6: MRR and MAP for $(\mathcal{M}_{\text{facebook_comp}}, \mathcal{M}_{\text{twitter_comp}})$ compared with $(\mathcal{M}_{\text{facebook}}, \mathcal{M}_{\text{twitter}})$

Method	MRR	Gain	MAP	Gain
$(\mathcal{M}_{\text{facebook}}, \mathcal{M}_{\text{twitter}})$	0.782	4.09%	0.485	3.51%
$(\mathcal{M}_{\text{facebook}}, \mathcal{M}_{\text{twitter}})\text{norm}$	0.790	3.04%	0.491	2.24%
$(\mathcal{M}_{\text{facebook_comp}}, \mathcal{M}_{\text{twitter_comp}})$	0.814	-	0.502	-

We hypothesize the major reason is that, $\mathcal{M}_{\text{facebook}}$ and $\mathcal{M}_{\text{twitter}}$ use different sets of features, and even the overlapping features may yield different distributions over Facebook v.s. Twitter. Thus, the ranking scores computed by $\mathcal{M}_{\text{facebook}}$ and $\mathcal{M}_{\text{twitter}}$ are likely to be incomparable in nature. To address this problem, it is necessary to learn an extra combining model to calibrate ranking scores from $\mathcal{M}_{\text{facebook}}$ and $\mathcal{M}_{\text{twitter}}$. Previous studies, e.g. Kim et al. [17], have proposed some methods, such as the type scores in [17], to combine ranking scores of different types of content. In our experiments, we also apply a straightforward combining method, i.e. the min-max normalization, to calibrate the ranking scores from $\mathcal{M}_{\text{facebook}}$ and $\mathcal{M}_{\text{twitter}}$, respectively. Table 6 demonstrates the performance of this method, denoted as $(\mathcal{M}_{\text{facebook}}, \mathcal{M}_{\text{twitter}})\text{norm}$, when we apply GBDT. From the table, we can find that $(\mathcal{M}_{\text{facebook_comp}}, \mathcal{M}_{\text{twitter_comp}})$ can still outperform $(\mathcal{M}_{\text{facebook}}, \mathcal{M}_{\text{twitter}})\text{norm}$, and the t-test shows the significant improvement (p-value < 0.05). And, we can observe the similar results when using ListNet.

Indeed, it is very possible to find a better combining model that can achieve better performance. However, $\mathcal{M}_{\text{facebook_comp}}$ and $\mathcal{M}_{\text{twitter_comp}}$ can ensure the predicted scores naturally calibrated and comparable without the extra combining model. As shown in Table 4, $\mathbf{y}_{\text{facebook}}$ and $\mathbf{y}_{\text{twitter}}$ play important roles in $\mathcal{M}_{\text{facebook_comp}}$ and $\mathcal{M}_{\text{twitter_comp}}$. Since they are predicted by $\mathcal{M}_{\text{mutual}}$ that inspired by mutual features, we can view $\mathbf{y}_{\text{facebook}}$ and $\mathbf{y}_{\text{twitter}}$ as relevance signals transferred between different types of vitalities. As a result, the further predicted grades $\mathcal{M}_{\text{facebook_comp}}$ and $\mathcal{M}_{\text{twitter_comp}}$ are naturally calibrated and comparable. Moreover, since they use $\mathbf{y}_{\text{facebook}}$ and $\mathbf{y}_{\text{twitter}}$ in training, both $\mathcal{M}_{\text{facebook_comp}}$ and $\mathcal{M}_{\text{twitter_comp}}$ take advantage of additional training data from the other social network to serve learning so as to benefit ranking and blending performance more by fully exploiting all available data.

7. Conclusion and Future Work

In this paper, we presented, to our best knowledge, an early attempt to blending vitality rankings from heterogeneous social networks. We introduced the formalized problem of blending vitality rankings from heterogeneous social networks, and proposed a variety of content, users, and users correlation features for this task. Due to the heterogeneity of vitalities, we employ a divide-and-conquer strategy in order to fully exploit all available features for vitalities from each social network, respectively. A large scale evaluation over two popular social networks demonstrated the effectiveness of our method for blending vitality rankings. In the future, we will first investigate more composite features to better serve the blending performance. Furthermore, we will explore incorporating personalization into the blending process. In particular, we will examine divide users into several group based on topic analysis on their interests, so as to generate separate ranking models for each individual user group.

References

- [1] A. Dong, R. Zhang, P. Kolari, J. Bai, F. Diaz, Y. Chang, Z. Zheng, Time is of essence: improving recency ranking using twitter data, in: Proc of WWW, 2010.
- [2] J. Weng, E. Lim, J. Jiang, Q. He, Twiterrank: finding topic-sensitive influential twitterers, in: Proc of WSDM, 2010.

- [3] S. Gordhamer, When do you use twitter versus facebook?, in: <http://mashable.com/2009/08/01/facebook-vs-twitter/>, 2009.
- [4] L. Backstrom, D. Huttenlocher, J. Kleinberg, X. Lan, Group formation in large social networks: membership, growth, and evolution, in: Proc of KDD, 2006.
- [5] A. Java, X. Song, T. Finin, B. Tseng, Why we twitter: understanding microblogging usage and communities, in: Proc of WebKDD/SNA-KDD, 2007.
- [6] R. Xiang, J. Neville, M. Rogati, Modeling relationship strength in online social networks, in: Proc of WWW, 2010.
- [7] W. Cohen, R. Schapire, Y. Singer, Learning to order things, Journal of Artificial Intelligence Research.
- [8] R. Herbrich, T. Graepel, K. Obermayer, Large margin rank boundaries for ordinal regression, in: Advances in Large Margin Classifiers, 2000.
- [9] C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, G. Hullender, Learning to rank using gradient descent, in: Proc of ICML, 2005.
- [10] Y. Freund, R. Iyer, R. Schapire, Y. Singer, An efficient boosting algorithm for combining preferences, in: Journal of Machine Learning, 2003.
- [11] Z. Zheng, K. Chen, G. Sun, H. Zha, A regression framework for learning ranking functions using relative relevance judgments, in: Proc of SIGIR, 2007.
- [12] Z. Cao, T. Qin, T.-Y. Liu, M.-F. Tsai, H. Li, Learning to rank: from pairwise approach to listwise approach, in: Proc of ICML, 2007.
- [13] C. Burges, K. Svore, P. Bennett, A. Pastusiak, Q. Wu, Learning to rank using an ensemble of lambda-gradient models, in: JMLR: Workshop and Conference Proceedings - Yahoo! Learning to Rank Challenge, 2011.
- [14] C. Burges, From ranknet to lambdarank to lambdamart: An overview, in: Microsoft Research Technical Report MSR-TR-2010-82, 2010.
- [15] C. Dwork, R. Kumar, M. Naor, D. Sivakuma, Rank aggregation methods for the web, in: Proc of WWW, 2001.
- [16] Y.-T. Liu, T.-Y. Liu, T. Qin, Z.-M. Ma, H. Li, Supervised rank aggregation, in: Proc of WWW, 2007.
- [17] J. Kim, W. B. Croft, Ranking using multiple document types in desktop search, in: Proc of SIGIR, 2010.
- [18] Z. Zheng, H. Zha, T. Zhang, O. Chapelle, K. Chen, G. Sun, A general boosting method and its application to learning ranking functions for web search, in: Proc of NIPS, 2007.