Data Mining and Predictive Analytics - Assignment 1
Image Popularity Prediction on Social Networks

Wei-Tang Liao and Jong-Chyi Su
Department of Computer Science and Engineering
University of California, San Diego
La Jolla, CA 92093
{wel144,jcsu}@eng.ucsd.edu

1 Introduction

Posts on social networks usually have various contents, including image, text, and metadata (such as user information). An example post from Chiclopia website is shown in Figure 1. Using these information, our goal is to predict an image's popularity. For this purpose, we explored the Chictopia dataset[1] from [4] and implemented the paper's main tasks. We combined features in various ways and performed regression with different models. In addition, we experimented with a multi-label, multi-class tag recommendation task: given the available post features, we predict possible tags that a user may associate with the post image. Our results show that using features from social networks can help us predict image popularity.

Figure 1: *An example of a post on Chictopia.com. There are four general fields: Tags, Clothes, User Information, and Popularity.*

[1]The dataset is available at http://vision.is.tohoku.ac.jp/∼kyamagu/research/chic-or-social/

2 Previous Work

Most of the posts in social networks combine images and text. Moreover, there are other metadata such as friends, followers and the number of likes. As the social network getting more and more popular, some characteristics like image popularity is important to be predicted. Combining image features and metadata, the popularity or quality can be predicted and ranked as described in [4, 1, 2]. In addition, the function of tagging images is popular in social networks. Using face detection and recognition, people can tag their friends more easily. However, some image features are hard to be defined, such as the dressing style. Some recent works leverage metadata for tag recommendation [7, 6]. Moreover, image classification tasks can be done with only metadata but any image features [3].

Another application for images on social media is cloth parsing and trend prediction. In [9], tagged fashion images are used to help predict tags and transfer parsing results. To further leverage information from social networks, popularity is predicted [4] by network, visual and textual content. The result shows that features from networks can help predict image popularity. Metadata can also be used for similarity comparison tasks such as similarity comparison for images [5]. In [8], it shows using metadata and image feature can predict similarity results gathering by Amazon Mechanical Turk (MTurk).

In conclusion, It is an emerging trend to use data mining techniques in image dataset. Relative works show that leveraging social network metadata can help predict computer vision tasks like popularity prediction, similarity comparison and tag recommendation.

3 Dataset

Our work is based on the Chic or Social paper [4] and its Chictopia dataset collected from Chictopia, a social network that users can post images having fashion clothes. For costumers, there are links to each clothes items appears on the image, so it is easy for them to purchase a set of clothes. There are also tags to let them find single item more quickly. Like other social media, users can comment, bookmark, or vote (like) a post. Users can also follow a fashion designer, or make friend with each other. For fashion designers, they can collaborate with clothing companies by posting clothes of specific brands and give the purchasing link to users.

The Chictopia dataset consists of two parts: 1) in-network part, including 328,604 Chictopia posts 2) out-of-network part, including 3000 Chictopia posts as well as crowd voting results collected by MTurk. Each posts has tags, clothes, user information, and popularity information, as shown in Figure 1. Note that each word in sentences describing the clothes are regarded as tags in the dataset. Therefore, in Figure 1, the count for the tag "T-shirt" is 2. After parsing the dataset, we find ten most frequent tags as shown in Table 1. There are differences between using total frequency and document frequency. For example, one post may contain several cloth items with brand "H&M", therefore the tag "H&M" is ranked the 10^{th} frequent tag by total frequency but not by document frequency. There are three information can be treated as popularity: the number of votes, comments, and bookmarks. Each vote means one user liked this post. Since it is the most direct signal for popularity, one of our goals is to predict the number of votes, given other data.

In our experiments, we randomly choose 65,721 posts from the Chictopia dataset, and split the training and test data by $90\% - 10\%$ ratio. We then analyze the statistics of each feature. The histogram plots are shown in Figure . All the features show the long-tail characteristic. The histogram of friends and followees have some spikes, which may be caused by some popular designers who have biased number of posts in the dataset.

Most Frequent Tags (by total freq.)	black	everyday	white	blue	shoes	dress	casual	brown	vintage	H&M
Frequency	137597	137491	58459	49583	88546	82251	78210	35552	51884	53257
Most Frequent Tags (by document freq.)	black	everyday	shoes	dress	casual	bag	skirt	top	white	boots
Frequency	379501	138361	121404	105614	89647	86970	82109	78822	73276	69218

Table 1: *Ten Most Frequent Tags.*

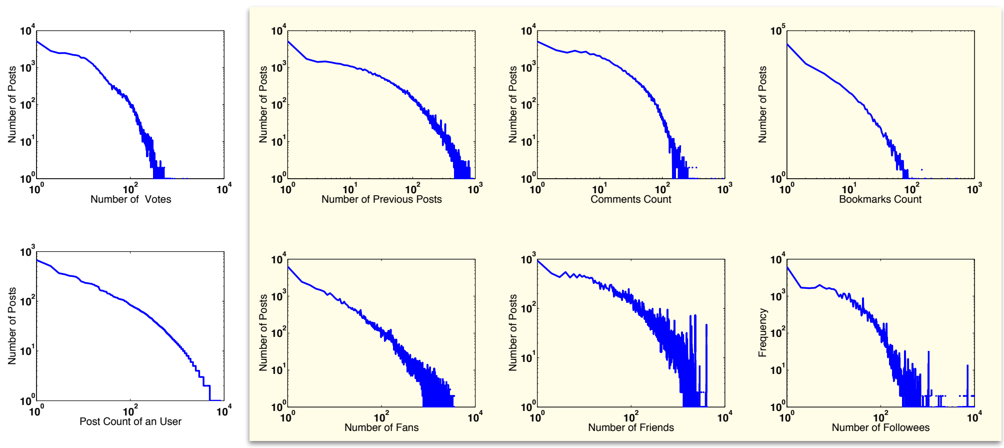

Figure 2: *Distribution of votes, previous posts, comments, bookmarks (from left to right, top row), and distribution of posts per user, fans, friends, and followees (from left to right, bottom row), in Chictopia dataset. The y-axis is the frequency of posts. Both x-axis and y-axis are in log scale. The relations between features in shaded six figures and the number of votes are compared in Figure 2. The distributions of number of friends and followees have some spikes. It is reasonable because the dataset may contain multiple posts from the same user, and one user has a fixed number of friends. There may be some people who post regularly and have lots of friends and followees on the social network.*

4 Features

The features we have from Chictopia dataset are shown in 2. We did not add or create other features so most of the description are similar to [4]; however, the features we used is slightly different from it. We separate features in three types: social, social*, and content. The social features are already described in Sec.3. The social* features are originally treated as a popularity measurement in [4]. They decided to only use vote count and discard the number of comments and bookmarks. We now briefly introduce content features:

Tag TF-IDF As shown in Fig 1, each post on Chictopia website has tags and few sentences describe the clothes. The dataset extract both unigrams(from tags and sentences) and bigrams(from sentences), then compute their TF-IDF weights. To reduce the feature dimensionality, only the first 1,000 most frequent n-grams are used.

Style descriptor This feature of clothing representation is parsed from the image directly, using the algorithm described in [9]. The feature includes color, texture, shape, and skin-hair probability.

Type	Name	Modality	Size
Social	Previous posts	Metadata	1
	Number of friends	Network	1
	log (Number of friends + 1)	Network	1
	Number of followers	Network	1
	log (Number of followers + 1)	Network	1
	Number of fans	Network	1
	log (Number of fans + 1)	Network	1
Social*	Number of comments	Network	1
	Number of bookmarks	Network	1
Content	Tag TF-IDF	Textual	1,000
	Style descriptor	Visual	411
	Parse descriptor	Visual	1,060
	Color entropy	Visual	6
	Image composition	Visual	6

Table 2: *Features.*

Parse descriptor This feature is also computed by clothing parsing algorithm [9]. First compute superpixels and assign each pixel with one of the 10 masks (labels). For each mask, find the RGB color, Lab color, texture response, and HOG descriptor, etc. All the features are concatenate into a 1,060 vector.

Color entropy It includes the entropy of RGB and Lab color from the image.

Image composition After detect the bounding box of human in the image, this feature contains the size of the bounding box, and the displacement from the bounding box to the center of the image.

In Figure , we want to investigate the correlation between the number of votes and six features (shaded in Figure). It shows that the comment count and bookmarks count have strong positive correlation with votes. It is reasonable because both comment and bookmarks are regarded as popularity measurements in [4] (but eventually they only use vote and discard comments and bookmarks). Other features have positive correlations (similar to Fig. 2 in [3]), but not very strong.

5 Predictive Task

Our predictive tasks have three parts: 1) Popularity prediction - regression, 2) Popularity prediction - classification, and 3) Tag recommendation.

5.1 Popularity Prediction - Regression

For popularity prediction, we use the same criteria as described in [4]. We want to use social, content, and both features to predict the number of votes in each post. It is a regression task , and we chose three criteria for the result: R^2, Spearman, and root-mean-squared-error (RMSE). R^2 and Spearman are rank correlation coefficients, which can be used to measure the dependence between two random variables. The higher value they are, the higher positive correlation between two variables. RMSE measures the squared error of predicted number of votes. We also turn this regression problem into binary classification task.

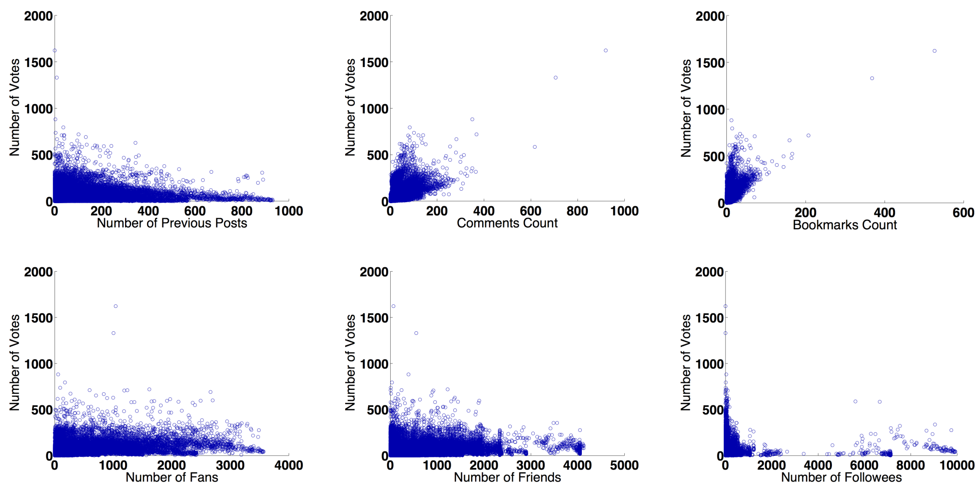

Figure 3: *The scatter plots showing the correlation between six different features and the number of votes. We can see that the comment count and the bookmarks count have strong positive correlation with the votes. They are reasonable because a post is more popular if more people bookmarked it or gave comments.*

5.2 Popularity Prediction - Classification

We now reformulate the regression problem into a classification problem. We calculate the 25% and 75% quantile of the number of votes. If a post has more votes than the 75^{th}-quantile, it has a label of +1; otherwise, it has a label of -1. In a separate task, if a post has less votes than the 25^{th}-quauntile, it has a +1 label; the label is -1 otherwise. The two classifiers identifies the most popular posts and the least popular posts in the dataset, respectively. The quantitative analysis is the misclassification rate. We experimented with different feature combinations and identified the best combinations by the lowest error rate.

5.3 Tag Recommendation

Tag recommendation for images is an important task on social networks [7, 6], since it assists users at finding interesting content and possibly new friendship connections. Both image content and metadata are commonly used to achieve automatic image tagging. In our implementation, we utilized social, social*, and content features to predict and recommend possible tags for a new post. We have omitted the tag TF-IDF content features in this task because the tags (or their presence) are now our target labels. For practicality purposes we have reduced our pool of possible tags down to the 1,000 most frequent tags in the entire Chictopia dataset. We have formulated this recommendation task into a multiclass, multilabel classification task. A class is defined by the presence of a tag, so there are a total of 1,000 possible classes. However, note that a post can contain multiple tags and thus multiple class labels. Instead of mapping a vector to a single label, we seek to define an appropriate model that maps each post vector x to a tag label vector y. This is achieved by breaking down the classification task into 1,000 independent binary classification tasks. Each binary classifier is trained to identify the presence of only one tag. The final list of all predicted tag labels for a post is the union of all tags that are predicted to be present by each of the binary classifiers. We use the average Hamming score to evaluate the quality of our model. For every post, the Hamming score is defined as the number of correctly predicted labels divided by the number of labels in the union of the true and predicted labels. Note that our formulation of the multilabel classification task assumes tags appear independently in the same post. One possible way of improving our prediction accuracy utilizes classifier chains that exploits correlation between tags.

6 Model

6.1 Popularity Prediction - Regression

For the popularity prediction regression task, we experimented with two types of models: the Linear Regression model (LinR) and the Support Vector Regression model (SVR). Our result is shown in Fig. 4. To train our regression models, we have divided our dataset of 65,721 posts into a training set and a testing set, with a ratio of 90%-10%. The posts are parsed into fixed-length floating point vectors with social-network-based and image-content-based features as described in the previous sections. The target value we fit to is the number of votes the post has received, which we believe is an appropriate measure of popularity.

6.2 Popularity Prediction - Classification

For popularity prediction classification task, we use support vector machine (SVM) model. We also tried logistic regression (LogR) and kernel support vector machine (kSVM); however, they need way more processing time without significant benefit.

6.3 Tag Recommendation

The tag recommendation task is a multiclass, multilabel classification task. We use One-Versus-the-Rest strategy to train 1,000 independent binary classifiers. Each binary classifier is a Linear Support Vector Classifier that learns from our training post vectors whether a tag is present (positive label) or absent (negative label) in any given post vector. The choice of LinearSVC is primarily due to its empirically faster training speed.

7 Results

7.1 Popularity Prediction - Regression

Our regression prediction results are shown in Table. 3. There are three different criteria: R^2, Spearman coefficient, and RMSE. Generally speaking, using social and social* features gives better result. For Spearman coefficient, content features actually aggravate performance instead of enhancing it. Note that the Linear Regression model consistently yields better results than the Support Vector Regression model, even though it is a rather straight-forward approach to solving the problem. We show the top five most/least popular images we have predicted in Fig. 4. On the bottom row, we can see that the tonality is monotonous; on the top row, there are more colorful clothes.

7.2 Popularity Prediction - Classification

We show the classification results in Table. 4. It is obvious that incorporating social and social* features yields the best results. This is because the comments count and bookmarks count have strong positive correlations to the number of votes (as shown in Fig. 3). Another notable discovery is that the results for combined social and content features is worse than the results for using only social features. We speculate that this is because the high-dimensional image content features, such as HOG, has no direct relation to people's perception for aesthetic ideal.

Criteria	R^2		Spearman		RMSE	
Model	LinR	SVR	LinR	SVR	LinR	SVR
Social	0.3587	0.0672	0.7050	0.6277	42.47	55.86
Content	0.2655	**0.0687**	0.6137	0.3988	45.03	55.95
Social + Content	0.4314	0.0621	0.7307	0.6004	36.88	**52.04**
Social + Social*	0.6080	0.0511	**0.8591**	**0.6901**	32.92	55.76
Social + Social* + Content	**0.6220**	0.0683	0.8410	0.6428	**32.29**	56.08

Table 3: *Popularity rediction - regression result. The best results in each column are shown in bold.*

Figure 4: *Top five predicted most popular and least popular images (Note that some images are removed from the website, and we discard them). The predicted scores and the ground truth votes are shown as reference. We can see that our prediction results corresponds to the votes. We also found that clothes with monotonous black and white colors usually have lower scores, such as the bottom left and bottom right images.*

Label	$>75\%$ Error Rate	$<25\%$ Error Rate
Social	15.2%	16.4%
Content	27.9%	26.4%
Social + Content	22.0%	22.4%
Social + Social*	**11.4**%	**10.6**%
Social + Social* + Content	15.4%	16.4%

Table 4: *Popularity prediction - classification result. The best results in each column are shown in bold.*

7.3 Tag Recommendation

Our tag recommendation result is shown in Table. 5. In Fig. 5, we show an example result where the actual tags and the recommended tags are listed. True positive words are highlighted in blue, false positives are in red, and true negatives are in green. The four true positives are obvious features. Some tags are inherently difficult to learn, such as the brand name ”Forever-21” and ”Topshop”. Textures like silk is also hard to be recognized, whether from social features or image content features.

Features	Hamming score
Social + Social* + Content (without Tags)	0.1357

Table 5: *Tag recommendation result.*

Predicted tags:	**Ground truth tags:**
shoes	shoes
vintage	vintage
white	white
dress	dress
everyday	topshop
casual	necklace
thrifted	suede
forever-21	silk
romantic	pearl
denim	white dress
socks	purple

Figure 5: *An example of tag recommendation results. Blue words are true positives. Red words are false positives. Green words are true negatives.*

8 Conclusion

Images posted on social networks usually come with text and user information. By leveraging these metadata, some image prediction tasks can be improved. We used the Chictopia dataset [4] from the fashion clothing website to predict image popularity. We utilized regression and classification models to predict image popularity, with features from text, metadata, and image content. Combining these features, we achieve better prediction popularity results, compared to using only image features. We also experimented with tag recommendation, which is a multi-class and multi-label classification task. With One-Vs-the-Rest algorithm, we found that network features are better correlated to tags.

References

[1] A. D. Sarma A. Khosla and R. Hamid. What makes an image popular? In *WWW*, 2014.

[2] P. Isola A. Torralba A. Khosla, J. Xiao and A. Oliva. Image memorability and visual inception. In *SIGGRAPH Asia 2012 Technical Briefs*, 2012.

[3] J. Leskovec J. McAuley. Image labeling on a network: using social-network metadata for image classification. In *ECCV*, 2012.

[4] Luis E Ortiz Kota Yamaguchi, Tamara L Berg. Chic or social: Visual popularity analysis in online fashion networks. In *ACM Multimedia*, 2014.

[5] D. Parikh and K. Grauman. Relative attributes. In *ICCV*, 2011.

[6] van Zwol Sigurbjörnsson, B. Flickr tag recommendation based on collective knowledge. In *WWW*, 2008.

[7] Zickler T. Darrell T. Stone, Z. Autotagging facebook: Social network context. In *CVPR Workshop on Internet Vision*, 2008.

[8] Sirion Vittayakorn, Kota Yamaguchi, Alexander C. Berg, and Tamara L. Berg. Runway to realway: Visual analysis of fashion. In *WACV*, 2015.

[9] K. Yamaguchi, M.H. Kiapour, and T.L. Berg. Paper doll parsing: Retrieving similar styles to parse clothing items. In *ICCV*, 2013.