Image Context Discovery from Socially Curated Contents

Akisato Kimura*, Katsuhiko Ishiguro, Makoto Yamada
Communication Science Labs, NTT Corporation
Keihanna Science City, Japan

Alejandro Marcos Alvarez†
Systems and Modeling Research Unit, University of Liege, Liege, Belgium

Kaori Kataoka, Kazuhiko Murasaki
Media Inteligence Labs, NTT Corporation
Yokosuka, Kanagawa, Japan

ABSTRACT

This paper proposes a novel method of discovering a set of image *contents* sharing a specific *context* (attributes or implicit meaning) with the help of image collections obtained from social curation platforms. Socially curated contents are promising to analyze various kinds of multimedia information, since they are manually filtered and organized based on specific individual preferences, interests or perspectives. Our proposed method fully exploits the process of social curation: (1) How image contents are manually grouped together by users, and (2) how image contents are distributed in the platform. Our method reveals the fact that image contents with a specific context are naturally grouped together and every image content includes really various contexts that cannot necessarily be verbalized by texts.

Categories and Subject Descriptors

J.3.1 [**Human-centered computing**]: Collaborative and social computing—*Collaborative and social computing theory, concept and paradigms*

Keywords

Social curation, contexts, graph clustering, classification

1. INTRODUCTION

One of the ultimate goals in the field of computer vision and pattern recognition is to learn the structure of our visual world with the help of large amount of real data such as images, videos and texts. This problem is known to be challenging but fundamentally important. High performance comparable with humans seems to be extremely difficult to achieve due to variations in appearances, especially when dealing with contexts that are often difficult to verbalize.

*The corresponding author. E-mail: akisato@ieee.org
†He contributed to this work during his internship at NTT CS Labs.

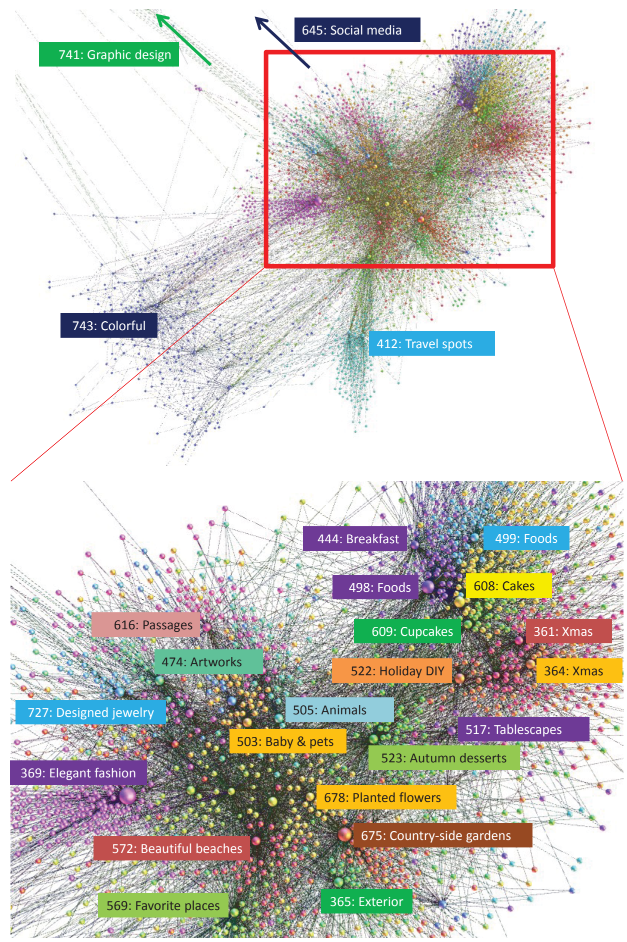

Figure 1: Image context mapping on 2D graph structures obtained by our proposed method (See Sec. 4 for the detail).

This background would direct researchers to the use of social network services (SNSs) as a promising resource, since they contain huge amount of side information useful for understanding contexts of image contents. Such side information includes user profiles, geo locations and user networks. The major interest of the current SNS developers and users are rapidly shifting from conventional text-based (micro)blogs such as Twitter and Facebook to multimedia contents such as Flickr, Snapchat, MySpace and Tumblr. However, the ability to analyze and exploit these unorganized multimedia data still remain inadequate, even with state-of-the-art media processing techniques.

This paper focuses on another emerging trend called *social curation*, a human-in-the-loop alternative to automatic algorithms for social media analysis [15]. Social curation can be defined as a spontaneous human process of remixing social media contents for the purpose of further consumption. What characterizes social curation is definitely the *manual efforts* involved in organizing social media contents, which indicates that curated contents are information sources potentially richer than automatic summaries generated by algorithms. Also, curated contents would provide latent perspectives and contexts that are not explicitly presented in the original resources. Following this trend, several studies have tried to examine and exploit social curation data [7, 9].

The main claim in this paper is that social curation data has a great potential as corpora to discover, visualize and retrieve contexts of image contents. In particular, this paper focuses on Pinterest, one of the most emerging social curation platforms dealing with image contentsPinterest has favorable and unique characteristics for mining image contexts.

(1) Focused contexts: There are so many groups (called *"boards"* in Pinterest) of image contents that are all manually collected, selected and maintained, so that board contributors and consumers can easily and quickly find images they want [10]. Therefore, most of the images on a specific board share the same context board contributors keep in mind.

(2) Content-centric networks: Contents and boards form networks in Pinterest, whereas users and user groups are the basis in conventional SNSs. Users may have several boards, follow boards of other users, collect contents from these following boards, and put them on their own boards. As a result, contents are distributed from boards to boards, which constitutes a diffusion network. This is very different from other SNSs that constitute user-centric networks in nature.

The above two characteristics readily imply that two boards sharing a lot of image contents often share some specific context. Our method makes the full use of this insight, which enables us to reveal various kinds of contexts engraved in every image content. When doing this, content diffusion takes an important role: Groups of boards sharing specific contexts can be discovered from content diffusion as shown in Figure 1, even though we know nothing about visual contents of images (See Sec. 4 in detail). With the additional use of image content information, we can further augment the benefits to unseen images, which enables us to visualize image contexts more clearly and to build a simple context-aware image-to-image retrieval [12].

2. RELATED WORK

Finding and understanding image contexts is quite a difficult task since it has not been well formulated yet, although several stimulating attempts have been executed so far (collaborative filtering with graphical models [5, 6], indexing representative images [19] and predicting importance of images [3]).

Meanwhile, image content analysis powered by social network information has become a hot topic in recent years, since a huge amount of side information such as associated texts, geo-tags and underlying user networks is freely available. A typical resource for this purpose is Flickr, the most famous image sharing platform. Images taken from Flickr have been broadly utilized in various researches such as preference recommendation [22, 18], community detection [13] and learning to rank for images [17]. In particular, Qi et al. [14] presented a method of embedding both content similarities and content-tag links in SNSs. Recently, other types of

Figure 2: Example of pinned image (See http://pinterest.com/pin/22166223139059126/).

SNSs have also been utilized for content recommendation with the help of content diffusion [1] and image annotation [20].

Our major contribution against the above researches is to focus on the utilization of social curation data for image context discovery, rather than usual social networks.

Social curation services have grown in popularity in recent years [15], however, few academic researches have been performed so far (see [11]). Hall and Zarro [8, 21] investigated the potential of Pinterest as a social curation platform from a sociological aspect. Duh et al. [7] first analyzed several characteristics of data collections taken from social curation services. Ishiguro et al. [9] estimated view counts of image contents as an indicator of their interestingness with the help of socially curated collections of microblogs.

To the best of our knowledge, our work is the first trial to explore image contexts from social curation data.

3. DEPICTING CONTENT DIFFUSION

3.1 Corpus Analysis

First we briefly describe the structure of Pinterest. An example of a pinned image is depicted in Figure 2. Every user has several boards and can *"pin"* (verb, meaning link and publish) an image to a specific *"board"* (a group of pinned images). Two major ways to pin an image are served in Pinterest: (1) Pin an image, namely upload a link to the image and publish it on her own board, and (2) *"repin"* an image, namely create a new link to an existing pin to her own board, quite similar to a retweet in Twitter. Typical users try to associate every board to a specific context comprising objects, scenes, feelings, events etc., so that they and their followers can easily find images they want [10]. The process of pinning shown above is indeed a sort of social curation.

We collected pinned images from Pinterest. The list of collected pins can be seen in the supplemental material (pinIdList.txt). The number of pins is 1063832, including 142385 original pins and 921447 repins, the number of unique images is 776860, and the total number of boards is 188818. As shown in Figure 2, every pin contains several useful information, such as (1) the pinned board

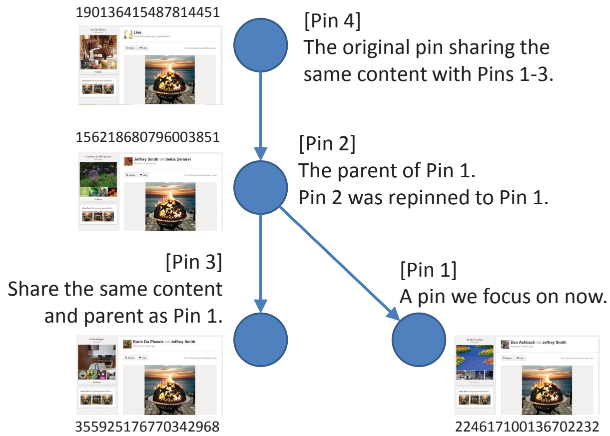

Figure 3: A simple example of pin diffusion graph

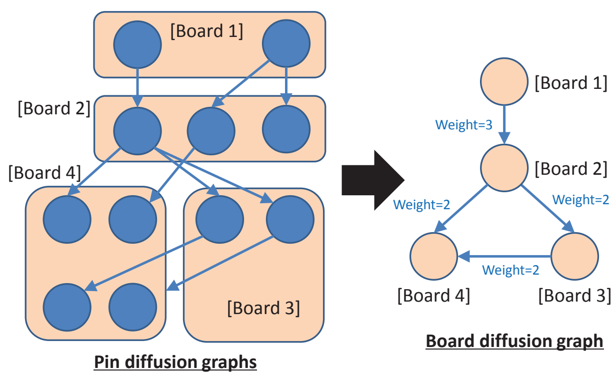

Figure 4: A simple example of board diffusion graph

name (red), (2) the board name and holder from which the image is repinned (blue), the original pinner and pinned board (green), and (4) the name of boards to which the pin is repinned (purple).

3.2 Building a Board Diffusion Network

This section describes a procedure for building image content diffusion graphs from the collected data. We illustrate one simple example in Figures 3 and 4. This structure takes an important role in revealing image contexts.

Let us focus on Pin 1 shown in Figure 3. From the collected data, Pin 2 can be identified as the parent of (the pin that was repinned to) Pin 1. Here, we create two nodes (pin nodes) of Pins 1 & 2 and one directed edge (a pin edge) from Pin 2 to Pin 1. The pin edge represents a route of content diffusion. In a similar manner, the route of diffusion related to Pin 1 can be identified. This results in yielding a tree-structured directed graph as shown in Figure 3, which we call a *pin diffusion graph*. Investigating all the pins in this way, a lot of pin diffusion graphs can be obtained.

Note that every board can be expressed as a disjoint set of pins. Thus, an induced graph can be created from those pin diffusion graphs by integrating all the pin nodes in the same board into a new single node (a board node), as shown in Figure 4. Through this procedure, pin edges are also integrated into a new single edge (a board edge) with weight the same as the number of pin edges. We call this induced graph a *board diffusion graph*.

Table 1: Board names in major 10 clusters

1	christmas-cookies-and-desserts, cooooooookies, natural-christmas, salads-salads, recipes, christmas-cuisine
2	holidays-christmas-decor-new-year, christmas-joy, cards-and-printables, vintage-illustrations, crafting-christmas
3	christmas-trees, christmas-decor, interior-exterior-design, holiday-seasonal-decorating-ideas, white-christmas
4	favorite-crafts, gift-wrap, christmas-ideas, gifts-to-go, christmas-quilts-crafts, snowmen-and-santa-claus
5	fashion-style, leggings-tights-and-socks, red-black, colored-redwinesspectrum, the-clothes-i-wish-i-owned
6	dresses-let-s-play-dress-up, classy-and-elegant, gorgeous-fashion, fall-winder-2012-trends, 2012-2013-high-fashion
7	silver-black-glitter-sparkles, sparkle-and-shine, silver-chrome-and-shine, bling, my-imaginary-closet
8	book-covers, just-kids-other-books, covers-we-love, book-cover-inspirations, judge-books-by-their-covers
9	type-i-like, packaging-and-identity, logos-and-icons, magazines, packaging, books-worth-reading, typography
10	preschool, homeschool-ideas, auntie-jenna-s-board-of-kid-crafts, creative-spaces, fun-for-kids, toy-lust-wish-list

We created a board diffusion graph from the collection and derive a 2-D node allocation with a standard spring model for visualization. The number of board nodes is 188818, and the number of board edges is 319591. The obtained graph with GML format can be seen in the supplemental material (boardDiffusionGraph.gml).

4. DISCOVERING IMAGE CONTEXTS

Once a board diffusion graph is obtained, groups of image contents sharing a specific context can be discovered without any help of image content information. Let us remember that boards sharing a lot of pinned image contents would share some specific context, as described in Sec. 1. From this viewpoint, finding board groups sharing similar contexts can be formulated by a problem of finding an optimal set of clusters that maximizes modularity. We adopt Louvain method [4, 16], one of the standard and scalable techniques for modularity-based graph clustering. However, its direct application to a board diffusion graph often yields several huge clusters due to the nature of board diffusion graphs, which makes it difficult to analyze the results. To mitigate the problem, our method performs adaptive hierarchical clustering: clusters with a number of nodes larger than a threshold are separated again into sub-clusters. This produces clusters with moderate sizes.

A part of the clustering result is shown in Table 1 that contains major 10 clusters and several board names as their members. The entire clustering result can be seen in the supplemental material (clusteringResultsHierarchical.csv). The result indicates that board names of every cluster implicitly describe a specific context of the cluster. It also indicates that hierarchical clustering effectively produces a layered structure of contexts. For example, abstract contexts related to Christmas (clusters 1-4) are separated into more specific contexts such as Christmas dinners (cluster 1), decorations (cluster 2), Christmas trees (cluster 3) and gifts (cluster 4).

A visualization of a board diffusion graph according to the clustering result with Gephi [2] can be seen in Figure 1 (Page 1), where the size of each node reflects its degree and nodes sharing the same color belong to the same cluster. The colored graph with GML format is in the supplemental material (clusteredBoardDiffusionGraph.gml). For simplicity, we selected only the clusters with more than 1000 nodes and more than 2 degrees in this visualization. This visualization implies that (1) every cluster has a specific context, (2) similar contexts (ex. cakes and foods, gardens and exterior) are

Table 2: Classification accuracy (in %) averaged over 10 runs. Column: Type of features. Row: Type of classifiers.

		User	Board	Cluster
Linear regression		87.09	98.94	97.36
kNN	$k = 1$	85.46	98.70	97.13
	$k = 5$	88.08	98.88	97.64
	$k = 10$	88.78	98.84	97.62

close to each other and (3) contexts that tend to appear together (ex. holiday DIY, cakes and Xmas) are also close to each other.

5. PRELIMINARY VALIDATIONS

This section validates our hypothesis that every board in a specific board cluster share a similar context, which would indicate the utility of image contexts founded by our method as a corpus of context-aware image retrieval.

The task in the validation is image classification according to the following setup: We selected about $380,000$ pins and associated $200,000$ unique images with 10 classes (architecture, fashion, cupcake, animals, chocolate, flowers, blue, sea, Christmas and green) from the whole collection shown in Sec. 3. Class label for each image was determined by the names of boards the image belongs to.

We utilized three types of features for classification: user, board and cluster features. (1) User feature $\boldsymbol{x}_i^u$ (192 dim.): If user j pinned or repinned image i, then the j-th element is set to 1, otherwise 0. (2) Board feature $\boldsymbol{x}_i^b$ (963 dim.): If image i was pinned to board l, then the l-th element is set to 1, otherwise 0. (3) Board cluster feature $\boldsymbol{x}_i^c$ (570 dim.): The m-th element is given by $n_{i,m}/n_i$, where $n_{i,m}$ is the number of times image i is (re)pinned to a board contained in cluster m, and n_i is the sum of $n_{i,m}$ over all m. Note that single image might be pinned or repinned to multiple boards by multiple users. From the above setup, board features should yield the highest classification performance (but not necessarily 100%).

The classification was performed with linear regression and k-NNs. The features were compared based on the classification accuracy averaged over 10 runs, where each run randomly selected $150,000$ images for learning and $50,000$ images for testing.

Table 2 summarizes the classification performance for three feature types (user, board and board cluster) and four classifiers (linear regression and k-NNs with $k = 1, 5, 10$). The table indicates that cluster features achieved performance comparable to board features and significantly higher than naive user features for all the classifiers. We observe that board clusters discovered by our method appropriately captured image contexts.

6. CONCLUDING REMARKS

This paper presented that social curation data has a great potential as corpora to discover, visualize and retrieve various kinds of contexts engraved in image contents. Especially, we revealed image contexts from a million image collection with the full use of the nature of a social curation platform and a content-centric network. Although we have exploited Pinterest as corpora in this paper, our method does not rely on any specific platforms if they have a structure with content diffusion and context-specific grouping.

Until then, we did not matter image contents. We have enjoyed the benefits of content diffusion on a social curation platform to reveal image contexts. However, our method shown in this paper considered only pinned images, which implies that any other images have been still out of focus. With the use of another method we have developed [12], we take the content into account to augment the benefits to much broader areas.

Since the image collection we have utilized was clean and well-organized and it had several favorable characteristics, fundamental techniques with well-designed features were sufficient for our purpose. However, we should develop more sophisticated techniques for further improvements.

Acknowledgments

We would like to thank Dr. Junji Yamato and Mr. Koh Takeuchi of NTT Communication Science Labs for their fruitful discussions.

7. REFERENCES

[1] Baluja et al. Video suggestion and discovery for YouTube: Taking random walks through the view graph. In *Proc. WWW*, 2008.

[2] Bastian et al. Gephi: An open source software for exploring and manipulating networks. In *Proc. ICWSM*, 2009.

[3] Berg et al. Understanding and predicting importance in images. In Proc. *CVPR*, 2012.

[4] Blondel et al. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics*, 2008.

[5] Boutemedjet et al. A graphical model for context-aware visual content recommendation. *IEEE Trans. MM*, 2008.

[6] De Pessemier et al. Context aware recommendations for user-generated content on a social network site. In *Proc. EuroTV*, 2009.

[7] Duh et al. Creating stories: Social curation of Twitter messages. In *Proc. ICWSM*, 2012.

[8] C. Hall et al. Social curation on the website Pinterest.com. In *Proc. ASIST*, 2012.

[9] Ishiguro et al. Towards automatic image understanding and mining via social curation. In *Proc. ICDM*, 2012.

[10] Jess Swiderski. *Pinterest for Business*. Que, 2012.

[11] Kimura. Social curation as corpora for large-scale multimedia content analysis. presented in *ICMR*, http://bit.ly/ICMR2013Curation, 2013.

[12] Marcos Alvarez et al. Exploiting socially-generated side information to improve dimensionality reduction. In *Proc. IWSAM*, 2013.

[13] Negoescu et al. Modeling Flickr communities through probabilistic topic-based analysis. *IEEE Trans. MM*, 2010.

[14] Qi et al. Exploring context and content links in social media: A latent space method. *IEEE Trans. PAMI*, 2012.

[15] Rosenbaum. Why content curation is here to stay. http://on.mash.to/ZAayZo, 2010.

[16] Shiokawa et al. Fast algorithm for modularity-based graph clustering. In *Proc. AAAI*, 2013.

[17] Trevisiol et al. Image ranking based on user browsing behavior. In *Proc. SIGIR*, 2012.

[18] Wang et al. Recommending Flickr groups with social topic model. *Information Retrieval*, 2012.

[19] Wang et al. Towards indexing representative images on the Web. In *Proc. ACMMM*, 2012.

[20] Xu et al. Towards annotating media contents through social diffusion analysis. In *Proc. ICDM*, 2012.

[21] Zarro et al. Wedding dresses and wanted criminals: Pinterest.com as an infrastructure for repository building. In *Proc. ICWSM*, 2013.

[22] Zheng et al. Flickr group recommendation based on tensor decomposition. In *Proc. SIGIR*, 2010.