IEICE TRANSACTIONS

on Information and Systems

VOL. E97-D NO. 6
JUNE 2014

A PUBLICATION OF THE INFORMATION AND SYSTEMS SOCIETY

The Institute of Electronics, Information and Communication Engineers
Kikai-Shinko-Kaikan Bldg., 5-8, Shibakoen 3 chome, Minato-ku, TOKYO, 105-0011 JAPAN

PAPER

Creating Stories from Socially Curated Microblog Messages*

Akisato KIMURA[†a)], ***Senior Member***, **Kevin DUH**[††b)], **Tsutomu HIRAO**[†], ***Nonmembers***, **Katsuhiko ISHIGURO**[†], **Tomoharu IWATA**[†], ***Members***, ***and*** **Albert AU YEUNG**[†††], ***Nonmember***

SUMMARY Social media such as microblogs have become so pervasive such that it is now possible to use them as sensors for real-world events and memes. While much recent research has focused on developing automatic methods for filtering and summarizing these data streams, we explore a different trend called social curation. In contrast to automatic methods, social curation is characterized as a human-in-the-loop and sometimes crowd-sourced mechanism for exploiting social media as sensors. Although social curation web services like Togetter, Naver Matome and Storify are gaining popularity, little academic research has studied the phenomenon. In this paper, our goal is to investigate the phenomenon and potential of this new field of social curation. First, we perform an in-depth analysis of a large corpus of curated microblog data. We seek to understand why and how people participate in this laborious curation process. We then explore new ways in which information retrieval and machine learning technologies can be used to assist curators. In particular, we propose a novel method based on a learning-to-rank framework that increases the curator's productivity and breadth of perspective by suggesting which novel microblogs should be added to the curated content.

key words: *social curation, microblogging, learning to rank*

1. Introduction

We are entering the age of ubiquitous social media. User-generated content such as microblogs have become so pervasive such that it is now feasible to exploit them as sensors for real-world events and memes. As such, an active research area is the development of new algorithms for social media analysis. Examples include topic detection [2], event summarization [3], [4], and automatic filtering [5], [6] in microblogs. We imagine these algorithmic advances will provide efficient ways to discover and summarize events of interest from large streams of social media.

This paper focuses on a different trend: a recent phenomenon called *social curation* is emerging as a *manual* human-driven alternative to *automatic* algorithms for social media analysis [7], [8]. Social curation, or sometimes called content curation, can be defined as the human process of remixing social media contents for the purpose of further consumption. At the most basic level, a curation service offers the ability to (1) bundle a collection of content from diverse sources, (2) re-organize them to give ones own perspective, and (3) publish the resulting story to consumers [9]. See Fig. 1 for a schematic example.

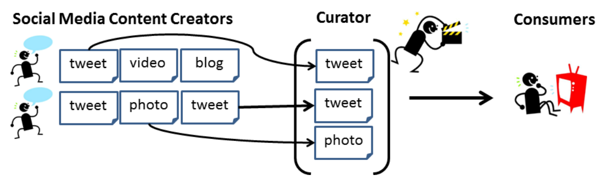


Fig. 1 Curation process: manual filtering and re-organization of social media for further consumption.

What characterizes social curation is the manual effort involved in organizing social media content. This human-in-the-loop means the curated content is a potentially richer source of information than automatic summaries and stories generated by algorithms. Specifically, curated content may give additional perspectives that are not present in the original sources; they may also be open-ended and evolve according to community interaction. Due to this excitement, social curation services such as Curated.by, Pearltrees, Storify, Scoop.it, Togetter and Naver Matome have grown in popularity in recent years.**

Let us take for example the reporting of a major event, such as the Arab Spring. Hundreds to thousands of local people are in the field, tweeting their observations, uploading photos and videos, and blogging their opinions – creating torrents of content. It takes a curator-reporter to weave these disparate tweets and photos into a coherent, meaningful story. This kind of personalized perspective adds value to social media, and provides something different from, for example, Google News' automatic summaries aggregated from major news publishers.

As another example, consider the diary of a group of friends on vacation in Tahiti. They tweet on Twitter, post on Facebook, and upload photos on Flickr. Further, other friends from their social networks (who were not so lucky as to get a vacation) retweet, like, and comment on their social media – creating threads of conversations throughout

Manuscript received September 13, 2013.
Manuscript revised January 21, 2014.

†The authors are with NTT Communication Science Laboratories, NTT Corporation, Kyoto-fu, 619–0237 Japan.
††The author is with the Graduate School of Information Science, Nara Institute of Science and Technology, Ikoma-shi, 630–0192 Japan.
†††The author is with Axon Labs Limited, Hong Kong.

*A part of the material in this paper has been presented in International AAAI Conference on Weblogs and Social Media (ICWSM2012) [1].

a) E-mail: akisato@ieee.org
b) E-mail: kevinduh@is.naist.jp

DOI: 10.1587/transinf.E97.D.1557

**http://curated.by, http://pearltrees.com, http://storify.com, http://scoop.it, http://togetter.com, http://matome.naver.jp the entire trip. After the trip, wouldn't it be nice to collect these memories in one central location, creating a social diary for future enjoyment?

These are real usage cases of social curation. And many other creative uses are imaginable. The goal of this work is to explore this emerging phenomenon. In particular, we seek to answer two major questions:

1. How are social curation services used today? What motivates curators to spend their time and effort?
2. How can we assist curators so that the manual effort is more natural and the resulting story is better?

In the following, we first present an analysis of a large corpus of social curation data. Based on this analysis, we then propose a novel method for assisting curators: given a partially curated story, it suggests a list of new microblog content (i.e. tweets) that might be valuable to include based on a learning-to-rank framework.

2. Related Work

As a new kind of web service, curation has been actively developed and debated in the popular press and blogosphere. For a summary of these discussions, see e.g. a curated website about curation in Scoop.it†. As a research topic, however, curation has been relatively unexplored [10].

The only previous research we are aware of was presented by Greene et al [11], which proposes to curate lists of *Twitter users* (as opposed to lists of *Twitter messages* studied here). The motivation is to monitor breaking news by curating a list of reliable and informative citizen-reporters. They propose an effective system to rank Twitter users based on multiple views of social network information. The impact of curating user lists is very different from directly curating Twitter messages, however. The former allows one to follow a dynamic stream of filtered news, while the latter is more akin to creating a story, snapshotted at curation time. We believe both kinds of curation each have their uses, and the methods proposed here may be helpful as features in [11] and vice versa.

In principle, social media research involving user recommendation [12]–[14], microblog filtering [5], [6], topic modeling [15], [16], event summarization [3], [4], and activity stream personalization [17] may all be beneficial for assisting curators. Microblog ranking and recommendation based on content freshness [18] and account authority [19] would be promising for finding seeds of stories that attract much attention of consumers. We think incorporation of these specialized techniques for different kinds of curated lists would lead to improved performances and rich user experiences. After all, as our analysis showed, usage scenarios are quite diverse so an assistive system is better off incorporating multiple techniques.

Finally, we note several influential works have performed in-depth analysis of the microblog landscape [20]–[22]. It would be interesting to connect these findings to the analysis of curators here. As [23] eloquently argues, microblogs have already evolved beyond the original purpose of letting users say "What are you doing?"; now they enable conversations, collaborations, and much more.

†http://www.scoop.it/t/social-media-content-curation

3. Formal Definitions

Before beginning, we formally define what we mean by *curation* here since this buzzword is used quite liberally in the popular press and blogosphere to describe many things.

In our world, there are *content creators*, *content consumers*, and *curators*. Content creators generate new nuggets of digital artifacts, such as tweets, blog posts, or uploaded photos. We define a curator as one who collects and organizes existing content into a larger unit. For example, a curator does not generate new tweets per se, but instead organizes a list of tweets from others. Consumers subscribe either to content creators directly or to curators.

Curation can be either an individual or collaborative process. We use the term social curation to mean social media curation, i.e. the curation of any social media content. Some pundits use social curation in its more restrictive sense to mean only the collaborative process of curation, but here we do not make this distinction.

4. Corpus Analysis

In order to understand social curation as it is happening today, we present an analysis of a large corpus of curation data.

4.1 Data Collection

In this study, we focus on the social curation of microblogs. We collected data from Togetter, which is quite a popular curation service in Japan, and it cites 4 million unique user-views per month in 2011. The Togetter curation data is in the form of *lists* of Twitter messages. An English example of a list can be seen in Fig. 2 (naturally, the majority of tweets are in Japanese). A list of tweets corresponds to what we called a *story*, representing a manually filtered and organized bundle.

Lists in Togetter draw on Twitter as its source. They may be created individually in private or collaboratively in public as determined by the initial curator. In the Togetter curation interface, the curator begins the list curation process by looking through his Twitter timeline (tweets from users that he or she follows), or directly searching tweets via relevant words or hashtags. The curator can drag-and-drop these tweets into a list, reorder them freely, and also add annotations such as list header and in-place comments.

A total of around 96,000 Togetter lists were collected from the period September 2009 - April 2010. This corresponds to a total of 10.2 million tweets from 800 thousand distinct Twitter users.

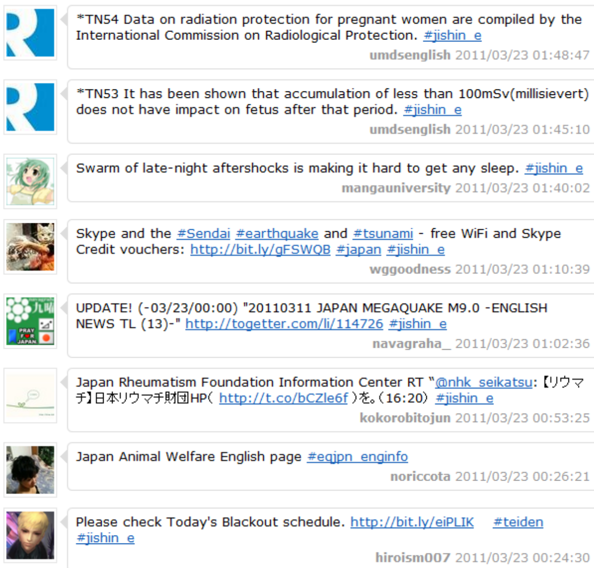


Fig. 2 An example of a list in Togetter. The purpose of the list is to curate up-to-date information about the 2011 Earthquake in Japan and its aftermath. As seen here, informative tweets from various sources are all collected together in one place. (Full list at http://togetter.com/li/112934)

4.2 Summary Statistics

We first provide some summary statistics to get a feel for the curation data. We are interested in basic questions such as:

1. How large is a list?
2. How many Twitter users are involved in a list?
3. How often does a list contain diverse sources vs. only tweets from the curator himself?

What are the answers you would expect? Some of the statistics were surprising to us:

1. The median size of a list is 40 tweets, and 90% of all lists have under 250 tweets. A scatter plot is shown in Fig. 3.
2. The median number of users per list is 6, and 90% of all lists have under 60 users. A scatter plot is shown in Fig. 4.
3. There is a bi-modal distribution, separating lists that consists of mainly self-tweets and diverse sources (Fig. 5).

Figures 3 and 4 are typical skewed distribution that are often observed in social media datasets. Nevertheless, what surprised us was the relatively large size of lists and number of users. A list of 40 tweets must take considerable effort to curate. Similarly, lists drawing from 60 distinct users' tweets appear difficult to gather: in these larger lists there must be much collaborative curation going on.

Figure 5 presents an interesting finding. Here we first separate the lists by size (i.e. number of tweets in a list). Then for each subset, we compute the percentage of self-tweets, defined as the fraction of tweets in the list written by the list curator. We observe an interesting bi-modal distribution in particular for the subset of small lists (under 30 tweets): a large fraction of lists in this category have either low self-tweet rate (less than 0.2) or 100% self-tweet rate, and few lists in-between.

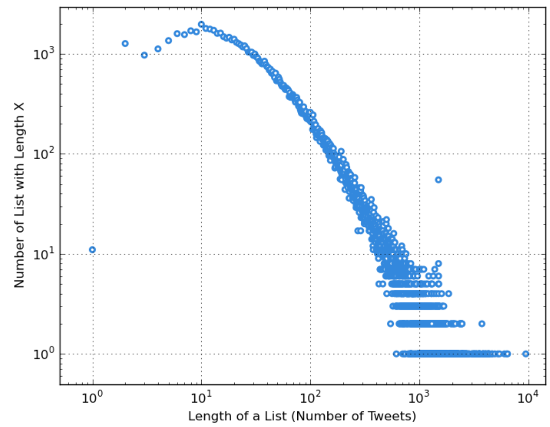


Fig. 3 Distribution of List Size by number of tweets.

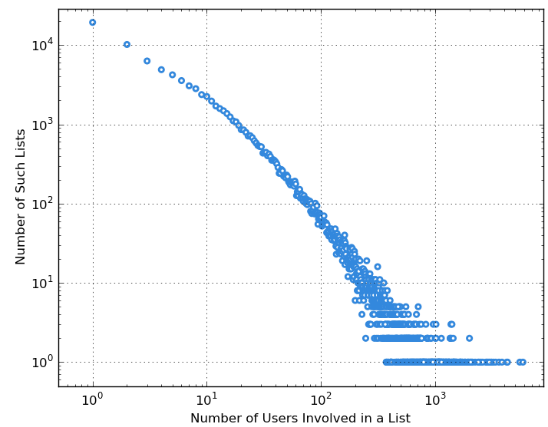


Fig. 4 Distribution of number of users involved in a list.

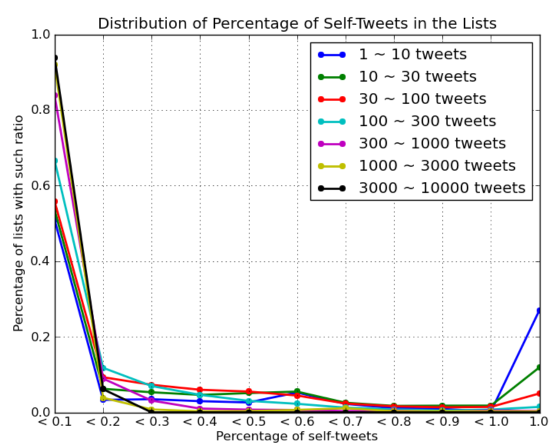


Fig. 5 Percentage of self-tweets. The lines represent lists of different sizes.

This suggests there is considerable diversity in how and

Table 1 Categories in the curated lists: Category labels are translated into English from Japanese.

Category Label	Percent of Data	Median #tweets/list	Median #users/list	Average #tweets/list	Average #users/list
Entertainment & Hobby	17%	45	7	100	28
Talk & Discussions	12%	36	4	87	16
Society, Politics & Economics	10%	36	5	102	26
Jokes	6%	37	9	94	27
News	6%	50	9	131	51
Science, Tech, & Computers	6%	47	8	138	37
How-To guides	3%	33	7	88	30
Unlabeled	40%	39	6	102	27
OVERALL	100%	40	6	106	28

why people use curation services. For example, one can imagine self-tweet lists as sort of personal bookmark folder while lists with diverse sources represent conversations and collaborative efforts. We turn to this question next.

4.3 Understanding Curator Motivations

Social curation appears to be a varied phenomenon: curators have different motivations for creating lists, and novel usage scenarios of curated lists are still being explored. We therefore think it would be insightful to investigate this diversity.

First, we ask the question: *What are the topics being curated?* Togetter curators usually categorize their lists with a predefined category label, and we analyze these distributions. Table 1 show the distribution of categories as well as length size statistics per category. As seen, a large fraction of lists (17 percent) talk about `Entertainment & Hobbies`, including music, sports, game, and anime. Serious topics (`Society, Politics, & Economics`) are also well-represented, consisted of 10 percent of the data. Lists about `News` are generally the largest, while lists labeled `Talk & Discussions` are shorter with fewer users. The correlation between number of tweets and number of users per tweet is high, with Pearson's coefficient $\rho = 0.729$. While there are differences among categories, it generally appears that all categories of conversation in Twitter are also curated in Togetter.

Our second question directly address the issue of curator motivation, asking: *What is the intended purpose of the list?* Since we do not have any prior knowledge of potential curator motivations, we performed this analysis via manual annotation. To do so, we randomly sampled and read through 435 lists. The annotators (the authors of this paper) would read each list and attempt to label it with its *intended purpose*. We started with a small set of intended purpose labels and through various annotation rounds gradually settled on a fixed set of 7 labels that encompasses most cases. Inter-annotator agreement is performed to check that the intended purpose labels can be agreed upon reasonably[†]. The final set of intended purpose labels and their frequencies in the annotation are:

- **Recording a Conversation** (19%): One of the most popular motivation for curating a list is to record a multi-party conversation on Twitter. Twitter conversations happen dynamically with its @reply and retweet features, but these are not suitable for browsing the conversation at a later time. Thus curators are motivated to manually format these conversations into an easily readable list.
- **Writing a long article via Tweets** (19%): The 140-character limit of Twitter does not prevent users from doing a soliloquy, writing a long article as a continuous series of tweets. Thus another popular use of curation is to present these tweets as they were originally intended, as a full article. The curator may or may not be the tweet author: both cases were observed in practice.
- **Summarizing an Event** (18%): A growing phenomenon with microblogs is the blending of conversations in physical and digital space. In particular, #hashtags are often used to connect conversations among participants of the same physical event (e.g. #wsdm tag on tweets related to the WSDM conference). While one could easily collect these tweets with keyword search, these curated lists represent a kind of final report summarizing the event.
- **Gathering Complex Info and Problem-solving** (16%): A curator may post a question and collect all the answers in a list. Or one may engage in a group brain-storming session. Also, one may be doing citizen-reporting as mentioned in the Introduction. This category is more difficult to pin down, but generally it involves figuring out some complex issues, leading to lists that are carefully curated and iteratively updated.
- **Just Playing** (14%): Human beings are fond of playing, and an undeniable apsect of social media is that it is a brave new playground. We have discovered many entertaining uses of social curation in practice, such as playing multi-player word games, jotting down the first random thought at time 23:59, and many others that are perhaps fun for the involved parties but unintelligible otherwise.
- **Diary** (9%): These lists contain individually or group-curated Twitter updates of ones day.
- **TV/Radio Show Transcript** (4%): This is a somewhat surprising use that caught us by surprise, and may be peculiar to a sub-population of the Togetter community.

[†]Average Cohen's kappa for 3 annotators on 7 labels is 0.42. This is not high but is reasonable as a first annotation procedure.

In summary, we have found that the usage scenario for social curation can be very diverse, encompassing various topics and intended purposes. As any good technology platform ought to do, social curation does not presuppose any usage scenario and the curators can be left to explore and evolve on their own.

5. Assisting Curators

Having analyzed some characteristics of curation as it is performed today, we also hope to build a method that will improve the experience for curators. Motivated by the observation that lists can be large and draw from diverse sources, we propose a method that helps curators discover useful tweets to include into a list. We have to note that the corpus analysis presented in the previous section serves as a basis of our proposed method, as presented in the following.

5.1 Problem Formulation and Proposed Method

We frame the problem as *tweet discovery based on partially curated lists*. First, we assume that a partially curated list is available. Namely, the list contains some initial tweets but is not yet entirely complete. For example, these may be lists in the process of being created, or lists that are occasionally updated by curators in different sittings.

The goal of our method is to suggest new tweets that would benefit the story if added to the initial partially curated list. The general architecture is shown in Fig. 6. It works as follows:

1. Given the set of seed tweets $S = \{s_i\}_{i=1,...,N_s}$ in the partially curated list, identify all the corresponding Twitter authors.
2. Retrieve the timeline of the curator and of all authors. This generates a set of candidate tweets $T = \{t_j\}_{j=1,...,N_t}$ that could potentially be added to the partially curated list.[†]
3. Rank candidate tweets in T and return a ranking sorted by relevance to seed tweets S to the curator.
4. The curator completes his curation work by surveying top-ranked candidate tweets or by other means as desired.

This architecture is similar to web search if we consider seed tweets as queries and candidate tweets as web pages to be retrieved and ranked. One difference is that we have multiple "queries" $S = \{s_i\}$ as opposed to one. Also, the operations used to retrieve candidates as well as the features used for ranking are necessarily different from web search. For example, we would need to take into account time information with respect to seed tweets when crawling user timelines. Nevertheless, we will be able to borrow techniques from web search to solve this problem.

[†]Optionally (not shown in Fig. 6 and not tried here), we might also include candidate tweets found by keyword search, using automatically identified keywords from seed tweets S.

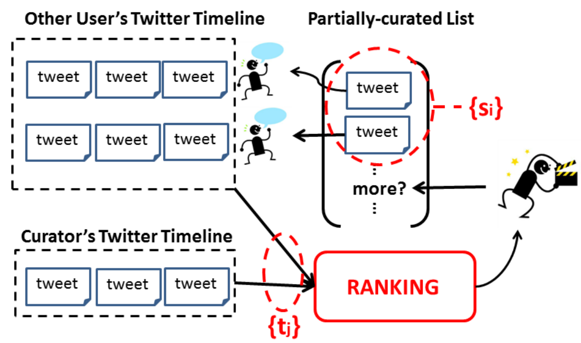


Fig. 6 Architecture of our proposed method: Given a partially curated list $S = \{s_i\}$, we retrieve candidate tweets $T = \{t_j\}$ from all relevant timelines, and return a ranked list to the curator.

5.2 Implementation Details

We next describe in detail some components of our method, in particular how we propose to rank candidate tweets T so that top rank tweets will be relevant with respect to seed tweets S. We will adopt a learning-to-rank approach [24], and use SVMrank [25], one of the state-of-the-art techniques.

First, we need to prepare a training set consisting of tuples $S = \{s_i\}$ and $T^{(L)} = \{(t_j, l_j)\}$ where l_j are labels indicating whether t_j is relevant for completing the seed tweets $\{s_i\}$. Then, we extract feature vectors for each candidate tweet t_j and run the learning algorithm, SVMrank. This generates a set of weights for each feature, which allows us to rank any new list of candidate tweets.

For each candidate tweet t_j, we define the following three categories of features:

- **Word similarity:** Word similariy is introduced to track and trace conversations with quotations and/or no hashtags. Let $\boldsymbol{t}_j$ be a vector-space representation of a candidate tweet t_j. Similarly, $\boldsymbol{s}_i$ is a vector-space representation of a seed tweet s_i. Then, we can define a word similarity feature f_j to be the sum of cosine distances between a candidate tweet vector $\boldsymbol{t}_j$ and all the seed tweet vectors $\boldsymbol{s}_i$ $(i = 1, 2, \ldots, N_s)$

$$f_j = \sum_{i=1}^{N_s} \cos(\boldsymbol{t}_j, \boldsymbol{s}_i). \tag{1}$$

Three different vector-space representations [26] are employed, leading to three different features, and as the fourth feature, we use the sum of BM25 scores [27], a very effective retrieval function in the information retrieval literature:

1. Term frequency (TF)

$$\mathrm{TF} = 1 + \log(c(w)),$$

where $c(w)$ is the count of a word w in the tweet.[††]

[††]We use an automatic word segmenter, Mecab, to split the Japanese sentences into words. http://mecab.sourceforge.net

2. TF-IDF

$$\text{TF-IDF} = 1 + \log(c(w)) \cdot \log \frac{N}{df(w)},$$

where N is number of tweets in list and $df(w)$ is the number of tweets containing word w.

3. Term occurrence: 1 if $c(w) > 0$, and 0 otherwise.
4. SumBM25:

$$f_j = \sum_i \text{BM25}(\boldsymbol{t}_j, \boldsymbol{s}_i),$$

$$\text{BM25}(\boldsymbol{t}, \boldsymbol{s}) = \sum_{m:s_m>0} \frac{3t_m}{2(0.25 + 0.75 \cdot L(\boldsymbol{t})/\overline{L}) + t_m} \cdot \log \frac{N - df(w_m) + 0.5}{df(w_m) + 0.5},$$

where w_m is the m-th word in the vector space representation, $L(\boldsymbol{t})$ is the length of document $\boldsymbol{t}$, and $\overline{L}$ is the average document length.

We refer to these features above as `Word-TF`, `Word-TFIDF`, `Word-OCC`, and `Word-BM25`, respectively.

- **Hashtag similarity:** We define other four features analogously to the word similarity features, except that we restrict the vector space representation to include only words appearing in hashtags. This is motivated by the lists intended to event summarizations that often use hashtags during curation. Naturally these are sparser vectors, but indicate more explicit information. Accordingly, we refer to these features as `Hash-TF`, `Hash-TFIDF`, `Hash-OCC`, and `Hash-BM25`.
- **Meta-information:** We also define four binary features based on meta-information of the social network or the partially curated list. These are developed to form lists with the following 3 intentions: "Record conversations", "write long articles", and "gather complex info".

1. Is the candidate tweet t_j by the same author as any of the seed tweets S?
2. Does the candidate tweet t_j and any of the seed tweets S contain the same @mention?
3. Does the candidate tweet t_j and any of the seed tweets S contain the same HTTP links?
4. Does the @mention in the candidate tweet t_j, if it exists, match the author of any of the seed tweets S?

We refer to these above features as `META-1`, `META-2`, `META-3`, `META-4`, respectively.

Note that for many of these 12 features we basically treat the "set of queries" $S = \{s_i\}$ independently, and sum up their scores to generate features. This is a simple solution to handle query sets of various sizes. However, it ignores inter-relationships among tweets. As future work, we imagine trying more sophisticated methods such as graph-based ranking [28], where $S = \{s_i\}$ and $T = \{t_j\}$ form vertices, and inter-relationships among tweets are captured as edges between vertices.

Table 2 Data characteristics for the task: Tweet discovery for assisting curators.

	#Lists	#TimelineTweets
Training	5000	1.3 million
Validation	1500	400 thousand
Test	2000	580 thousand

5.3 Experiment Setup and Results

To evaluate our proposed method, we did another crawl of Togetter in the period July-October 2011, resulting in a total of 29,000 lists. Concurrently we retrieved Twitter timelines of all authors involved in these lists at the same period. It was not possible to get sufficient timeline data for all authors; we filtered out lists where number of candidate tweets T is less than 5, ending up with a dataset of approximately 8500 lists. This dataset was randomly divided into training, validation, and test splits as shown in Table 2.

To simulate partially curated lists, we randomly chose 20% of tweets in the list as seed tweets $S_k = \{s_{i,k}\}$, where k is the index of the list. The remaining 80% are considered as gold reference tweets $R_k = \{r_{i,k}\}$ that we would like to discover. In particular, the candidate tweet set $T_k = \{t_{j,k}\}$ contains all tweets collected from author and curator timelines and is a superset of these gold reference tweets R_k. So, the task is to rank these gold reference tweets R_k above the other irrelevant tweets $T_k - R_k$ using the seed tweet set S_k as queries. In this dataset, the median number of seed tweets S_k is 12, and the median number of candidate tweets T_k is 192. Around 30% of these candidate tweets are gold reference R_k that we hope the assistive method will discover.

We trained our ranker on the training data, tuned for the best SVMrank parameters (i.e. accuracy-regularization tradeoff parameters) on the validation set, and present evaluation results on the test set. As an evaluation metric, we use Mean Average Precision (MAP), a standard metric for evaluating ranked results. Averaged precision (AP) for the k-th list is defined as:

$$\text{AP}_k = \frac{1}{|R_k|} \sum_{j:t_{j,k} \in R_k} \mathit{precision}(j),$$

where $\mathit{precision}(j)$ is the percentage of relevant tweets up to rank of $t_{j,k}$. MAP is defined as the mean of these AP over the entire test data.

Figure 7 summarizes the result for our machine learned ranker, compared to baselines of ranking with individual features (which is unsupervised). Since this is a new task, we also report results using random prediction for reference.

We observe that the proposed method, SVMrank with all the features, is the best performer, with MAP = 0.857. Some individual features also do reasonably well: Word-TFIDF achieves MAP = 0.825 and Word-BM25 gets

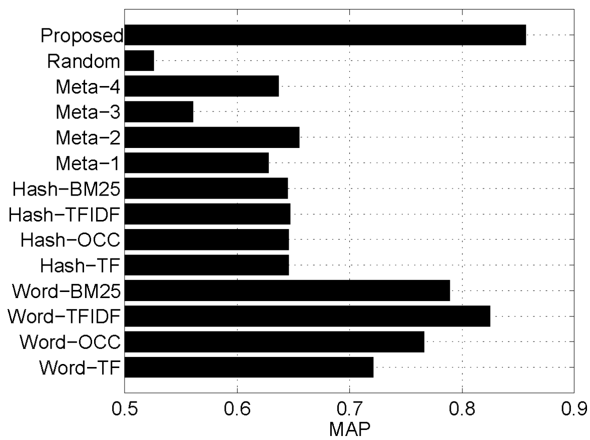


Fig. 7 MAP result comparisons.

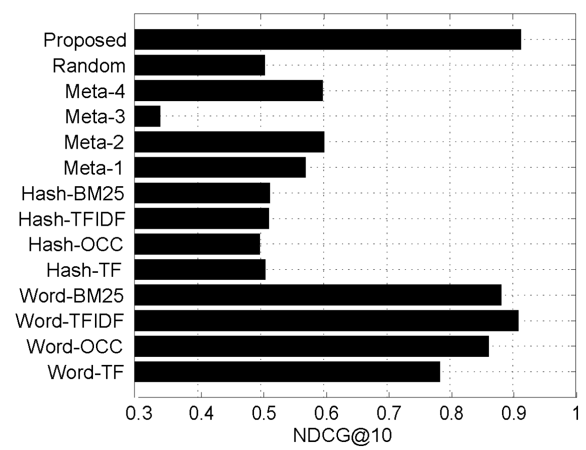


Fig. 8 NDCG@10 result comparisons.

MAP = 0.789. Nonetheless, the proposed method outperforms these by statistically significant margins (under T-test with 0.05 level). We also measured similar trends using another evaluation NDCG [29], with the proposed method achieving NDCG@10 = 0.895, as shown in Fig. 8. In the proposed method, the top features that received most weight from the learned model are, in order: Word-TFIDF, Meta-3, Word-BM25, Hash-TFIDF, Meta-2. This suggests that features that exploit tweet structure (e.g. Hash, Meta) are quite complementary to word-based similarities. These relative high evaluation scores are promising. They suggest that tweet discovery for assisting curation is a well-formulated problem that could be meaningfully tackled by current techniques.

Finally, there is more that could be done to improve the method, though: some partially curated lists appear much harder than others, as shown in Fig. 9. The proposed method achieved MAP > 0.86 for about half of lists in the test set, but there is a long tail of difficult lists. As future work, we will try to see if these hard examples correlate with any of the diverse intended purpose or topics described previously. We will also need subjective evaluations to understand how it impacts the overall curation experience. As the first step

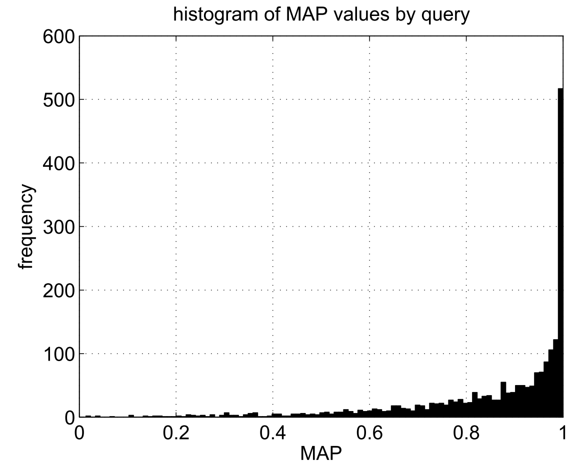


Fig. 9 Histogram of MAP results by list.

for this purpose, we built a demo system to interactively create a story from tweets with our proposed method. Figure 10 shows screenshots describing our demo system (see http://www.brl.ntt.co.jp/people/akisato/socialweb1.html for the detail). We hope this will bring some extensive knowledge and findings for assisting curators.

6. Conclusion

In this paper, we first presented a analysis of the social curation phenomenon. We asked the questions: (1) *What are the characteristics of curated lists?* (2) *Why are curators motivated to perform this manual-intensive endeavor?* We found that curated lists of tweets today can be quite elaborate, and encompass a wide range of topics. Curator motivations are quite diverse, ranging from conversation records and personal diaries to collaborative gathering of complex information.

We found that curators today are a very diverse group, with a range of styles, motivations, and usage scenarios.

As the second contribution, we have introduced a method for assisting curators, helping them discover relevant tweets to curate. The promising experimental results suggest that much future research can be done in this space. In particular, we are interested in two new directions.

- *What are the characteristics of multimedia curation data?* While we focused solely on curation of Twitter messages here, many curators also integrate microblogs with photos, videos, and links to various kinds of rich media (cf. Storify, Scoop.it and Naver Matome[†]). Our proposed framework would be applicable to such multimedia curation services, but several minor modifications might be necessary, including the way of extracting features for suggesting articles and multimedia contents. Further, these need not be confined to lists, but can also organized as threads, magazine layouts, etc.

[†]http://storify.com, http://www.scoop.it, http://matome.naver.jp

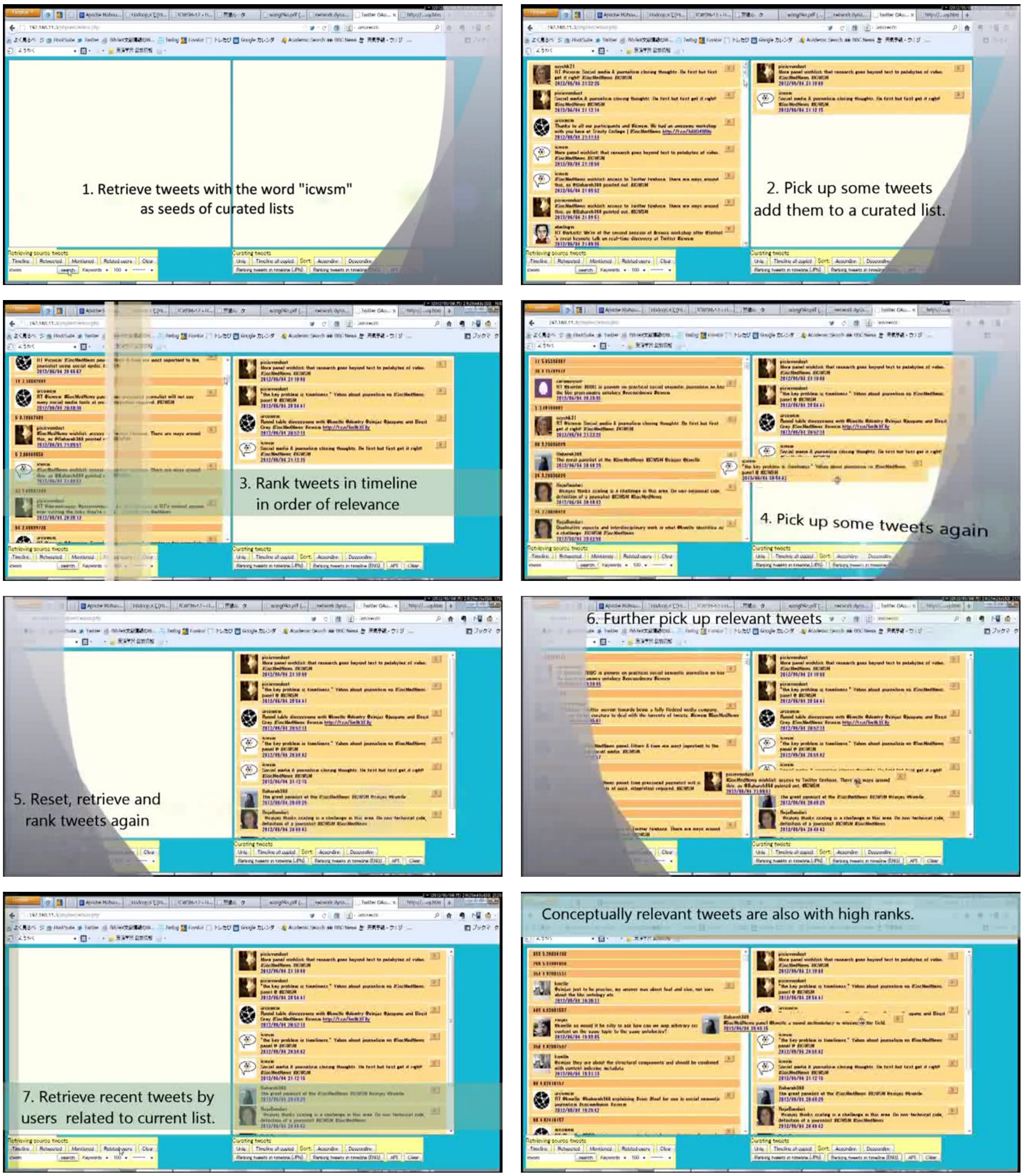


Fig. 10 Screenshots ot the demo system to interactively creates a story from Twitter messages with our proposed method. First, it retrieves and aggregates tweets from the timeline with the word "ICWSM" (1), and selects some of them as seeds manually to form a partially curated list (2). Given a partially curated list, it ranks tweets in the timeline according to the relevance to the list (3), and picks up some tweets manually to further augment the curated list (4).

- *Are there other ways to assist curators?* When creating a story with our method, we assume that a partially curated list is available as a seed of the story. That might be laborious especially for non-expert curators. First story detection methods [2], [30]–[32] might help them obtain promising seeds of stories without any effort. Meanwhile, curation is a manual process by definition, so an automatic system should not help too much. That would limit the curator's creativity. However, as evidenced by the constant update of new features introduced by popular curation services, we have not yet established a fixed way to do curation. One possibility is tools to help collaborative curation.

The spread of social media has empowered anyone to become a content creator. We are now witnessing a new trend, where anyone can also be a curator, manually filtering and organizing the torrent of social media. They do not create content, but by adding their own perspectives to existing content, they provide new value to social media consumers. It would be exciting to see how this trend evolves in the coming years.

Acknowledgment

The authors would like to thank Mr. Hirofumi Fujimoto for all his help as to this project, especially collecting a large corpus of curation and microblog data and building the demo system.

References

[1] K. Duh, T. Hirao, A. Kimura, K. Ishiguro, T. Iwata, and C. Au Yeung, "Creating stories: Social curation of Twitter messages," Proc. International AAAI Conference on Weblogs and Social Media (ICWSM), pp.447–450, 2012.

[2] Q. Diao, J. Jiang, F. Zhu, and E.P. Lim, "Finding bursty topics from microblogs," Proc. Annual Meeting of the Association for Computational Linguistics (ACL), pp.536–544, July 2012.

[3] H. Becker, M. Naaman, and L. Gravano, "Beyond trending topics: Real-world event identification on Twitter," Proc. International AAAI Conference on Weblogs and Social Media (ICWSM), pp.438–441, 2011.

[4] D. Chakrabarti and K. Punera, "Event summarization using tweets," Proc. International AAAI Conference on Weblogs and Social Media (ICWSM), pp.66–73, 2011.

[5] O. Dan, J. Feng, and B. Davison, "Filtering microblogging messages for social TV," Proc. International Conference Companion on World Wide Web (WWW), pp.197–200, 2011.

[6] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas, "Short text classification in Twitter to improve information filtering," Proc. International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp.841–842, 2010.

[7] S. Rosenbaum, "Why content curation is here to stay." http://mashable.com/2010/05/03/content-curation-creation/, May 2010.

[8] M. Ingram, "The future of media: Storify and the curatorial instinct." http://gigaom.com/2011/04/25/the-future-of-media-storify-and-the-curatorial-instinct/, April 2011.

[9] R. Scoble, "The seven needs of real time curators." http://scobleizer.com/2010/03/27/the-seven-needs-of-real-time-curators/, March 2010.

[10] A. Kimura, "Social curation as corpora for large-scale multimedia content analysis," Proc. ACM International Conference on Multimedia Retrieval, April 2013. http://bit.ly/ICMR2013Curation

[11] D. Greene, F. Reid, P. Cunningham, and G. Sheridan, "Supporting the curation of Twitter user lists," Proc. NIPS Workshop on Computational Social Science and the Wisdom of Crowds, 2011.

[12] J. Hannon, M. Bennett, and B. Smyth, "Recommending Twitter users to follow using content and collaborative filtering approaches," Proc. ACM Conference on Recommender Systems (RecSys), pp.199–206, 2010.

[13] J. Weng, E.P. Lim, J. Jiang, and Q. He, "TwitterRank: Finding topic-sensitive influential Twitterers," Proc. ACM International Conference on Web Search and Data Mining (WSDM), pp.261–270, 2010.

[14] J. Chen, R. Nairn, L. Nelson, M. Bernstein, and E. Chi, "Short and Tweet: Experiments on recommending content from information streams," Proc. International Conference on Human Factors in Computing Systems (CHI), pp.1185–1194, 2010.

[15] L. Hong and B. Davidson, "Empirical study of topic modeling in Twitter," Proc. KDD Workshop on Social Media Analytics, 2010.

[16] D. Ramage, S. Dumais, and D. Liebling, "Characterizing microblogs with topic models," Proc. International AAAI Conference on Weblogs and Social Media (ICWSM), pp.130–137, 2010.

[17] I. Guy, I. Ronen, and A. Raviv, "Personalized activity streams: Sifting through the "River of News"," Proc. ACM Conference on Recommender Systems (RecSys), pp.181–188, 2011.

[18] Y. Duan, L. Jiang, T. Qin, M. Zhou, and H.Y. Shum, "An empirical study on learning to rank of tweets," Proc. International Conference on Computational Linguistics (COLING), pp.295–303, Aug. 2010.

[19] F. Cheng, X. Zhang, B. He, T. Luo, and W. Wang, "A survey of learning to rank for real-time Twitter search," Proc. International Conference on Pervasive Computing and Applications (ICPCA), pp.150–164, 2012.

[20] B. Jansen, M. Zhang, K. Sobel, and A. Chowdury, "Twitter power: Tweets as electronic word of mouth," J. American Society for Information Science and Technology, vol.60, no.11, pp.2169–2188, Nov. 2009.

[21] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?," Proc. International Conference Companion on World Wide Web (WWW), pp.591–600, 2010.

[22] A. Java, X. Song, T. Finin, and B. Tseng, "Why we twitter: Understanding microblogging usage and communities," Proc. WEBKDD and SNA-KDD Workshop on Web Mining and Social Network Analysis, pp.56–65, 2007.

[23] C. Honeycutt and S. Herring, "Beyond microblogging: Conversation and collaboration via Twitter," Proc. Hawaii International Conference on System Sciences (HICSS), pp.1–10, 2009.

[24] C. He, C. Wang, Y.X. Zhong, and R.F. Li, "A survey on learning to rank," Proc. International Conference on Machine Learning and Cybernetics (ICMLC), pp.1734–1739, 2008.

[25] T. Joachims, "Training linear SVMs in linear time," ACM International Conference on Knowledge Discovery and Data Mining (KDD), pp.217–226, 2006.

[26] C. Manning, P. Raghavan, and H. Schütze, Introduction to Information Retrieval, Cambridge University Press, 2008.

[27] S. Robertson and H. Zaragoza, "The probabilistic relevance framework: BM25 and beyond," Foundations and Trends in Information Retrieval, vol.3, no.4, pp.333–389, April 2009.

[28] J. Weston, R. Kuang, C. Leslie, and W. Noble, "Protein ranking by semi-supervised network propagation," BMC Bioinformatics, vol.7, no.S-1, 2006.

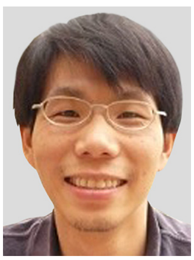
[29] K. Jarvelin and J. Kekalainen, "IR evaluation methods for retrieving highly relevant documents," Proc. International ACM Conference on Research and Development in Information Retrieval (SIGIR), pp.41–48, 2000.

[30] S. Petrović, M. Osborne, and V. Lavrenko, "Streaming first story detection with application to Twitter," Proc. Conference of the North American Chapter of the Association for Computational Linguistics, pp.181–189, 2010.

[31] S. Petrović, M. Osborne, and V. Lavrenko, "Using paraphrases for improving first story detection in news and twitter," Proc. Conference of the North American Chapter of the Association for Computational Linguistics, pp.338–346, 2012.

[32] M. Osborne, S. Petrović, R. McCreadie, C. Macdonald, and I. Ounis, "Bieber no more: First story detection using twitter and wikipedia," Proc. SIGIR Workshop on Time-aware Information Access, 2012.

Akisato Kimura received his B.E., M.E. and D.E. degrees in Communications and Integrated Systems from Tokyo Institute of Technology, Japan in 1998, 2000 and 2007, respectively. Since 2000, he has been with NTT Communication Science Laboratories, NTT Corporation, where he is currently a senior research scientist in Media Information Laboratory. He has been engaged in multimedia content identification, computational models of human visual attention, automatic image/audio/video annotation, and cross-media mining. His research interests include pattern recognition, computer vision, image/video processing, human visual perception, statistical signal processing, machine learning and social media.



Kevin Duh received his B.S. in 2003 from Rice University and Ph.D. in 2009 from the University of Washington, both in Electrical Engineering. From 2009 to 2012, he was a Research Associate at the NTT Communication Science Laboratories, NTT Corporation. He is currently an Assistant Professor at the Nara Institute of Science and Technology (NAIST), Graduate School of Information Science. His research interests include natural language processing, machine learning, and information retrieval.

Tsutomu Hirao received the B.E. from Kansai University in 1995, M.E. and Ph.D. in Engineering from Nara Institute of Science and Technology in 1997 and 2002, respectively. He is currently with NTT Communication Science Laboratories. His current research interests include Natural Language Processing and Machine Learning.

Katsuhiko Ishiguro has been a researcher at NTT Communication Science Laboratories, NTT Corporation, Kyoto since 2006. He received B.E. and M.Info. degrees from the University of Tokyo, Japan in 2000 and 2004, respectively, and the Ph.D. degree from the University of Tsukuba, Ibaraki, Japan in 2010. His research interests include multimedia data modeling with Bayesian approaches, probabilistic models for data mining, and time series analysis. He is a member of the IEEE, IEICE and IPSJ.

Tomoharu Iwata received the B.S. degree in environmental information from Keio University in 2001, the M.S. degree in arts and sciences from the University of Tokyo in 2003, and the Ph.D. degree in informatics from Kyoto University in 2008. He is currently a research scientist, distinguished researcher at Learning and Intelligent Systems Research Group of NTT Communication Science Laboratories, Kyoto, Japan. His research interests include data mining, machine learning, information visualization, and recommender systems.

Albert Au Yeung received his B.E. in Information Engineering and M.Phil in Computer Science and Engineering from the Chinese University of Hong Kong in 2004 and 2006. He received his Ph.D in Electronics and Computer Science from the University of Southampton in 2009. He was a research associate at the NTT Communication Science Laboratories. He is currently with Ayon Labs Limited, Hong Kong. His research interests include social media, social network analysis, user modeling, and recommendation systems.