Creating Stories: Social Curation of Twitter Messages

Kevin Duñ [kevinduh], Akisato Kimura [akisato], Tsutomu Hirao, Katsuhiko Ishiguro [k_ishiguro], Tomoharu Iwata, Ching Maan Yu Yeung [albertauyeung]

What is Curation?

curate: /ˈkyoo,raɪt/ [verb, transitive]
- select, organize, and look after the items
- From Latin cura, meaning “to care”
[New Oxford American Dictionary]

Curation service offers ability to:
- bundle a set of contents from diverse sources
- reorganize them to give one’s own perspective
- publish the resulting story to consumers [Scobleizer: http://scobleizer.com]

Project Goals

1. To characterize the curation phenomenon
   How are curation services used today?
   What motivations curators to spend their time/effort?
   → first corpus analysis of microblog curation
   (The main content of the proceeding)
2. To develop assistive technologies for curators
   → a recommendation system that suggests new content given partially curated lists
   (Not included in the proceeding)

Corpus Analysis: Basic Statistics, Topic Distributions, & Curator Motivations

1. How large is a list?
   - median 40 tweets/list
   - 90% lists are under 250 tweets
2. How many users per list?
   - median 6 users/list
   - 90% lists are under 60 users
3. How often a list contain diverse sources or only tweets from curator?
   - bimodal distribution, esp. small lists
4. How diverse in terms of topics?
5. How diverse in terms of motivations?
   - Very diverse, no typical trends

GENERAL OBSERVATIONS:
- Lists are very diverse in terms of size, topic, and purpose (curator’s motivation for creating the list).
- Lists are rather large and elaborate.

USEFUL INSIGHT:
- Lists are supervised corpora for analyzing microblog messages.

Assisting Curators

Goal: Help curators discover useful, new content based on partial curation
Content discovery is time-consuming; let curators focus on manual filtering & presentation

System Architecture: Similar to web search
Tweets in partial list serve as queries; retrieve relevant tweets; aggregated tweets are ranked.

Data Collection

We collected curated lists of tweets from Togetter.com, a popular curation service in Japan (2009/09-2010/04)

Data Size:
- # Togetter Lists: 96k
- # Tweets: 10.2M
- # Distinct Users: 800k

Conclusions

Summary:
1. Our analysis found microblog curation to be a very diverse phenomenon in terms of size, topic, motivation.
2. Our prototype assistant helps curators discover new content; basic web search method gets reasonable results.

Open Questions:
1. Other areas of curation worth exploring:
   - Multimedia curation (e.g. storify, curated.by)
   - Discourse modeling & temporal dynamics in curated lists
   - Social interactions in collaborative curation
2. Improving the assistive system:
   - Incorporating prior research:
     User curation [Greene+ 2011], pre-filtering [Dan+ 2011], activity stream personalization [Guy+2011]
   - Multiple tweets in partial lists as multi-query retrieval