

Comment and Context Matching

JERRY SUN and JIBANG WU, University of Virginia

General Terms: Information Retrieval

Additional Key Words and Phrases: topic models, text correspondence modeling, social media, user comments

1. INTRODUCTION

The live comment in Facebook Live and many other video broadcasting sites has gained huge popularity in recent years, as user-to-user interaction has become an important feature in video-browsing. However, the traditional news media hasn't done much to exploit its comments to promote user-experience, which, in fact, are very valuable especially among those controversial topics. In real life, because of its unstructured nature, the comments area seem to be a debate ground with various unsorted ideas. What we want to do is to restructure the comments into different topics based on the content of the news itself, so that user can immediately find the related comments while reading the news. This will give reader a sense of discussing the news with others while reading which changes the traditional on-line news browsing experience.

In our project, we propose a relational score between comments and each paragraphs based on algorithm developed by [Cai] and also filter out those unrelated spam comments. After receiving related comments, they will be presented in two major ways. Those significantly related comments will be inserted directly into the document, and others will be saved into sidebar, and show up when user clicks the related link in related paragraphs. In particular, this will be presented as a Chrome extension, and first targeted to specific news site, like ArcTechnica, or CNN. Later on, if time permits, we can scale this app to dynamically apply for different news sites(detail see 4).

2. BACKGROUND AND RELATED WORK

Recent studies have shown great interests in exploiting content of the user comments for various purposes[Park et al. 2011; Giannopoulos et al. 2012], but most of those only target to analyze the content of the site information or the sentiment triggered by the information, instead of directly serving the readers in their reading experience. Meanwhile, not much previous work further categorize the comments based on the content of the news itself.

3. SPECIFICATION

- (1) Automatic parsing comment during web browsing from different sites.¹
 - (a) Inspect news site HTML format
 - (b) Hard-code parsing comments content from particular HTML
 - (c) Crawl comments based on main page of the news as most sites only display a few top comments
 - (d) Study common pattern of news sites comments layout
 - (e) Dynamic parsing based on pattern search
- (2) Apply algorithm to filter and match the comment to certain context.
 - (a) Apply matching algorithm based on the study of [Cai ; Das et al. 2014]
 - (b) Detect and remove spam comment [Kant et al. 2012; Cai]

¹The project will start from hard-code parsing given sites, and then attempt to develop dynamically crawling algorithms

- (c) Use matching score to filter and rank comments
- (3) Restructured the web page to push the comment into static news. (For details see figure 1, 2 below)
 - (a) Rearrange comments based on relative passages based on algorithm
 - (b) Design UI pattern to insert top comments(only several of those) into original news page
 - (c) Design UI pattern to push all related comments if user is interested, i.e. clicks on a link, by pop-up sidebar
 - (d) Prototype details see Appendix A

4. MILESTONE

- (1) Algorithm realization
 - (a) Spam comment detection and removal
 - (b) Comment and context match
 - (c) Recommend comment pick
- (2) Toolkit implementation
 - (a) HTML injection with collection of comments in sidebar
 - (b) News content restructure to push recommended comments
- (3) Toolkit Scaling²
 - (a) Automatic comment crawling to adopt different website layout
 - (b) User setting and customization

REFERENCES

- Renqin Cai. Accounting for the Correspondence in Commented Data. (????).
- Mrinal Kanti Das, Trapit Bansal, and Chiranjib Bhattacharyya. 2014. Going Beyond Corr-LDA for Detecting Specific Comments on News & Blogs. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining (WSDM '14)*. ACM, New York, NY, USA, 483–492. DOI: <http://dx.doi.org/10.1145/2556195.2556231>
- Giorgos Giannopoulos, Ingmar Weber, Alejandro Jaimes, and Timos Sellis. 2012. Diversifying User Comments on News Articles. In *Proceedings of the 13th International Conference on Web Information Systems Engineering (WISE'12)*. Springer-Verlag, Berlin, Heidelberg, 100–113. DOI: http://dx.doi.org/10.1007/978-3-642-35063-4_8
- Ravi Kant, Srinivasan H Sengamedu, and Krishnan S Kumar. 2012. Comment spam detection by sequence mining. In *Proceedings of the fifth ACM international conference on Web search and data mining*. ACM, 183–192.
- Souneil Park, Minsam Ko, Jungwoo Kim, Ying Liu, and Junehwa Song. 2011. The Politics of Comments: Predicting Political Orientation of News Stories with Commenters' Sentiment Patterns. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work (CSCW '11)*. ACM, New York, NY, USA, 113–122. DOI: <http://dx.doi.org/10.1145/1958824.1958842>

² The Scaling part shall be optional for our main project based on time limit. We currently do not have a clear idea of how to implement automatic crawling. If the difficulty of this implementation exceeds our expectation, it can be omitted, and the project can focus on a number of major news sites.


A. APPENDIX

React has popularized functional programming in JavaScript. This has led to giant frameworks adopting the Component-based UI pattern that React uses. And now functional fever is spilling over into the web development ecosystem at large.

But the React team is far from relenting. They continue to dig deeper, discovering even more functional gems hidden in the legendary library.

Fig. 1. Original page from Medium

React has popularized functional programming in JavaScript. This has led to giant frameworks adopting the Component-based UI pattern that React uses. And now functional fever is spilling over into the web development ecosystem at large.

**Superman**
Wahoos

More


The JavaScript ecosystem is moving from "new framework of the week" to "new (faster) React clone of the week"...

9:33 AM · 13 Jan 2017

But the React team is far from relenting. They continue to dig deeper, discovering even more functional gems hidden in the legendary library.

Fig. 2. Insert matched comments between corresponding paragraphs

React has popularized functional programming in JavaScript. This has led to giant frameworks adopting the Component-based UI pattern that React uses. And now functional fever is spilling over into the web development ecosystem at large.

**superman**
@wahoo

More

The JavaScript ecosystem is moving from "new framework of the week" to "new (faster) React clone of the week"

9:33 AM · 13 Jan 2017

×

788 COMMENTS

Click here to read the best comments from the past week.

Share your thoughts.

All 788

AI Picks 8

Den

Palm Beach · 12 hours ago

Declaring state updates as pure functions makes it a breeze to test complex state transitions. Even no need for shallow rendering!

Reply · 1047 Recommend

patacant

Southern California · 12 hours ago

React is so awesome

Reply · 185 Recommend

Fig. 3. Button to toggle the sidebar with collection of comments