A Diversity-Seeking Mobile News App Based on Difference Analysis of News Articles

Conference Paper · September 2015

DOI: 10.1007/978-3-319-22852-5_7

2 authors:

Keisuke Kiritoshi
Kyoto University

Qiang Ma
Kyoto University

A Diversity-Seeking Mobile News App Based on Difference Analysis of News Articles

Keisuke Kiritoshi (✉) and Qiang Ma (✉)

Kyoto University, Kyoto 606-8501, Japan
kiritoshi@db.soc.i.kyoto-u.ac.jp, qiang@i.kyoto-u.ac.jp

Abstract. To support the efficient gathering of diverse information about a news event, we propose a diversity-seeking mobile news app on smart mobile devices. At first, by extending our previous work, based on the entity-oriented news analysis, we propose three measures for searching and ranking news articles from perspectives of difference in opinions, difference in details, and difference in factor coverages. Then, by utilizing these measures, we develop a news app on mobile devices to help users to acquire diverse reports for improving news understanding. One of the notable features of our system is a context-aware re-ranking method for enhancing the diversity of news reports presented to users by considering the access history. The experimental results demonstrate the efficiency of our methods.

Keywords: News app · Diversity · Difference analysis · Context aware re-ranking · Crowdsource experiment

1 Introduction

In some sense, news is never free from bias due to the intentions of editors and sponsors. To helping users to understand news events, considering diversity of news articles is important [1,2,5,8]. With the spreading of smart phones and tablets, news apps, the news reading applications on smart mobile devices, have been widely used. However, a news app usually provides only one article per each topic. In addition, mobile search is more difficult than general web one due to the limitations of environment and devices. As a result, a user may lose the chance to obtain information from multi-viewpoints to avoid biased impression.

In this paper, to improve the users' experiences of reading news on smart mobile devices, we propose a novel news app with the function of helping users to seek diverse information on a news event.

At first, we propose three measures to search and rank news articles by extending our previous work [1]. In [1], based on a user survey, we have proposed four ranking measures, (*relatedness*, *diversity*, *difference in opinion*, and *difference of detailedness*). The experimental results described in [1] clarify that although the difference between articles is important, they should be related to each other at first. Based on this observation, we refine these ranking measures. The refined measures are summarized as follows.

Q. Chen et al. (Eds.): DEXA 2015, Part II, LNCS 9262, pp. 73–81, 2015.
DOI: 10.1007/978-3-319-22852-5_7

– *DC (Difference in Factor Coverage)* is the extent of how many different things are described in two news articles reporting the same event.
– *DO (Difference in Opinion)* is the extent of difference of subjective descriptions in two news articles reporting the same event.
– *DD (Difference in Details)* is the extent of difference of details of two articles reporting the same event.

By utilizing these measures, we develop a news system effectively providing different reports on same events to support users' news understanding on smart mobile devices. The major contributions of this paper can be as follows:

– We propose three entity-oriented measures for ranking news articles by extending our previous work (Sect. 3) to help us to search and rank news articles by focusing on the difference in news reports.
– We propose a diversity-seeking news app (Sect. 4). As one of the notable features of our system, we propose a context-aware re-ranking method for enhancing the diversity of news reports provided to users. (Sect. 4.3).
– We conducted crowdsource experiment to validate the effectiveness of our ranking measures (Sect. 5.1). The re-ranking method is validated by a simulation experiment (Sect. 5.2).

2 Related Works

To help users' better understanding of news articles, news browsing systems which visualize and highlight the differences between news articles have been proposed. Ogawa et al. [3] study the analysis of differences between news articles by focusing on named entities and they propose a stakeholder mining mechanism. They extract stakeholders who are mentioned and present a graph constructed based on description polarity. The target news of Ogawa et al. is text news, while Xu et al. study on stakeholder mining on multimedia news [4].

NewsCube [2] presents various aspects of a news event and presents these using an aspect viewer to facilitate understanding of the news. TVBanc [5] compares news articles based on a notion of topic structure. TVBanc gathers related news from various media and extracts pairs of topics and viewpoints to reveal the diversity and bias of news reports on a certain news event.

In contrast, we are studying on searching and ranking related news articles by estimating the difference between articles. Our method focuses on how to provide users diverse reports different from the ones users have already read. In addition, our system is the first attempt trying to effectively provide diverse news reports on smart mobile devices.

3 Entity-Oriented Ranking Measures of News Articles

In our previous work [1], the experimental results reveal that although the difference between articles is important to obtain diverse information, a news article reporting totally different event is not useful to help users' news understanding. Based on this observation, in this paper, we refine the definitions and propose three entity-oriented ranking measures: *DC (Difference in Factor Coverage)*, *DO (Difference in Opinion)*, and *DD (Difference in Details)*.

In this section, at first, we brief the method of extracting entities and entity-related descriptions from news articles (refer to [1] and [3] for the details). Then, we introduce the entity-oriented ranking measures.

3.1 Extraction of Entities and Entity-Related Descriptions

We use a language tool StanfordCoreNLP[1] to extract named entities and apply the method proposed by Ogawa et al. [3] to generate a tree structure (Ogawa et al. call it Relationship Structure). We consider descriptions on named entities as sets of sub-trees of relationship structure, the root of each being a verb and its descendants containing the target named entities.

Another important notion in our entity-oriented news analysis is core entity [1]. Intuitively, core entities in an event are the named entities mentioned frequently in the articles reporting that event.

3.2 Ranking Measures

As mentioned before, our entity-oriented ranking measures are defined based on comparing named entities and their descriptions in news articles.

DC (Difference in Factor Coverage): DC is the degree of how many different things are described in two news articles reporting the same event. We estimate *DC* from two aspects, (1) how many different factors mentioned in these articles, and (2) whether these articles are related to the same event or not. For aspect (1), we can simply compare the entities mentioned in articles. The more different entities two articles describe, the higher difference in factor coverage the two articles are. For aspect (2), we compare the entity mentioned in each article with core entity set of the given event. Let E_{core} be the set of core entities of a certain news event. *DC* between articles a and o, $dc(a, o)$ is calculated as follows:

$$dc(a, o) = rel_{eve}(E_{core}, a) \times div_{dif}(a, o) \tag{1}$$

$$div_{dif}(a, o) = |E_a - E_o| \tag{2}$$

$$rel_{eve}(E_{core}, a) = \frac{|E_a \cap E_{core}|}{|E_{core}|} \tag{3}$$

where, E_a and E_o are sets of named entities in mentioned by a and o respectively.

[1] http://nlp.stanford.edu/software/dependenciesmanual.pdf

DO (Difference in Opinion): DO denotes the different extent of opinions between two articles. We compare the description polarities (positive, negative, and neutral) on named entities in articles. If two news articles report the same entities while their polarities are different, we regard these articles are different from opinion. *DO* of two news articles a and o, $do(a, o)$ is calculated as follows.

$$do(a, o) = rel_{mut}(a, o) \times sup(a, o) \tag{4}$$

$$sup(a, o) = w_{do} \times \sum_{e \in \{E_a \cup E_o\}} |sup_a(e) - sup_o(e)| \tag{5}$$

$$rel_{mut}(a, o) = \frac{|E_a \cap E_o|}{|E_a \cup E_o|} \tag{6}$$

where, $sup_a(e)$ and $sup_o(e)$ are polarities of named entity e in articles a and o respectively. w_{do} is the weight of core entities and is calculated as follows:

$$w_{do} = \begin{cases} w_{core,do} & (e \in E_{core}) \\ 1 - w_{core,do} & (others) \end{cases} \tag{7}$$

DD (Difference in Details): DD denotes the different degree of details provided by two news articles. We compare named entity related descriptions in articles to estimate *DD* following *difference of detailedness* in our previous work. After extraction of named entity related descriptions, we apply LDA (Latent Dirichlet Allocation) [6] to detected topics from the description of named entities. Then, we compare topic coverage and topic word on each named entity between two news articles. *DD* includes weight w_{dd} calculated by the same as formula (7).

4 Diversity-Seeking Mobile News App

By utilizing the proposed ranking measures, we develop a system to help users to acquire diverse reports of news events on a smart mobile device.

Users may use mobile devices to check news when users have a little time during moving or waiting. However, due to the time limitation and features of mobile devices, it is not easy to find diverse reports on a certain news event even the users may be interested in the event. In addition, the typical news apps present only one article per each news event. These are reason why we focus on news app on smart mobile device. The system consists of a news server and news client and the main functions are explained Sects. 4.1–4.3.

4.1 News Server: Gathering and Analyzing News Articles

Our diversity-seeking news system targets English news articles. We gather top news articles and their related article by using Google News Realtime Coverage[2] (one of the Google News' functions to present articles related to top news

[2] https://support.google.com/news/answer/2602970

articles). After gathering news articles, the news server analyzes and ranks news articles to top news articles per each event with the method introduced in Sect. 3. Differences between news articles are quantified by *DC*, *DO*, and *DD*, respectively. Each top news article and its related articles with top rank of each measure will be delivered to news clients and then be presented. The server will carry out context-aware re-ranking every time a user history increases as Sect. 4.3.

4.2 News Client: Presenting News Articles

The news client of our diversity-seeking news system has two view modes. One is the *top news view* to list today's top news. The other is *article details view*, which presents the content of a news article with links to its three top-ranked related articles. Figure 1 illustrates the running examples of these two views.

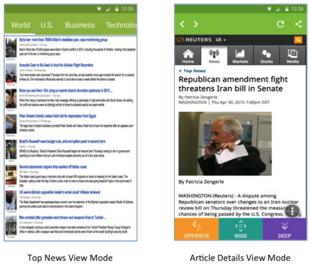

Fig. 1. View nodes of news client

Top news view shows top news articles gathered from Google News per each category. A user can click one article to view its details in *article details view*.

Article details view presents the details of a news article. There are three link buttons named "Opposite", "Wide", and "Deep", corresponding to the top ranked article of *DO*, *DC* and *DD*, respectively. By clicking these buttons, a user can access diverse reports on the same event as the current one.

4.3 Context-Aware Re-ranking

It is possible that the presented top-ranked different articles are the ones a user have read already and are similar to each other because the rank is decided by comparison with the current viewing article. In order to present more diverse information from different articles, we propose a context-aware re-ranking method. That is, the compare target to ranking does not only include the current article, but also includes the articles have been accessed before.

Suppose that A is the set of related news articles about a certain news event and $H (H \subset A)$ are the articles the user has read before. We update the ranking score $d(b, H)$ of article $b \in \{A - H\}$ regarding H as follow.

$$d(b, H) = \frac{1}{|H|} \sum_{h_i \in H} d(b, h_i) \quad (8)$$

where, $d()$ represent $dc()$, $do()$, and $dd()$.

5 Experiments

We carried out two experiments to evaluate our methods. One is for the ranking measures. The other is for the context-aware re-ranking method.

5.1 Crowdsource Experiment on Ranking

To evaluate our refined ranking measures by various people, we conducted a crowdsource experiment on the platform provided by CrowdFlower[3]. In the experiment, we gathered news articles with Google News US edition (Top stories and the Realtime Coverage) from June 17th 2014 to June 24th 2014. The number of news events is twenty and fourteen news articles in average were selected randomly per each topic. In each topic, we selected one article as the current article and the others as its related articles.

In our experiment, we asked workers to compare the pair of current article and one of its related articles. As a result, we have thirteen pair-comparisons per each news event in average. In each pair comparison, we asked workers score the related article according to the five grades evaluation system from three viewpoints, *DC*, *DO*, and *DD* respectively.

We compared ranking by the average scores assigned by crowdsource workers with that by our proposed ranking methods. nDCG (Normalized Discounted Cumulative Gain) [7] is the evaluation measure. We varied each parameter w_{do} (Formula 7), and w_{dd} (Sect. 3.2) and calculated the average of nDCG for the top k ranking results ($k = 6$) of the 20 topics (events). The nDCG results of each measure are shown in Table 1.

In Table 1, *DCB* and *DOB* denote the comparative method, which calculates *DC* and *DO* without considering relatedness between the target articles, respectively. These scores are calculated as $DCB = div_{dif}(a, o)$ and $DOB = sup(a, o)$.

In the nDCG results of *DD*, the highest evaluation value was 0.8964 and then $w_{dd} = 0.9$. These values are acceptably high. In contrary, it is hard to say which method is better for calculating *DC* and *DO*. One of the considerable reasons is that the articles used in the experiment are strongly related to each other. To confirm this assumption, we conduct a small additional experiment. We manually insert some noisy articles to a news event in our data set and test the ranks of these noisy articles. The ranking results with noisy articles are shown in Tables 2 and 3. These tables show the noisy articles are high-ranked by DCB and DOB while DC and DO decrease their ranks significantly. Therefore, we can say, although the proposed method cannot outperform previous method, it works well and is independent on the accuracy of gathering related news articles.

[3] http://www.crowdflower.com/

Table 1. nDCG of *DD*, *DC*, and *DO* ($k = 6$)

Parameter w_{dd}	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
DD	0.8934	0.8949	0.8953	0.8952	0.8945	0.8938	0.8938	0.8944	**0.8964**
Parameter w_f	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
DOB	0.9186	0.9204	0.9251	**0.9273**	0.9225	0.9229	0.9206	0.9222	0.9236
DO	0.9086	0.9076	0.9086	0.9088	0.9087	0.9130	0.9142	**0.9159**	0.9142
DCB	**0.9332**								
DC	**0.9242**	Note : DC and DCB have no parameter.							

Table 2. Ranking of *DC* with noisy articles

Article	DCB	DC
Noisy article1	3	5
Noisy article2	2	1
Noisy article3	1	8

Table 3. Ranking of *DO* with noisy articles

Article	DOB	DO
Noisy article1	2	18
Noisy article2	1	18
Noisy article3	6	18

5.2 Experiment on Context-Aware Re-Ranking Method

We conducted experiment to evaluate our context-aware re-ranking method by simulating the news reading sequences. We randomly selected fifteen news events that are used in our crowdsource experiment. The simulation based experiment is carried out by repeating the following steps per each news event.

1. Let a and a_b be the current news article. Let A be the related article set of a certain news event. Let accessed articles $H = a$.
2. Rank the articles in A respectively from three aspects: *DC, DO* and *DD*. Store these ranks in R_c, R_o, and R_d, respectively.
3. Perform the context-aware re-ranking method to create ranks of articles in A respectively from three aspects.
4. Randomly choose one aspect x with the limitation that each aspect should not be selected more than twice. Choose the top-ranked article t from the chosen aspect x. Also, choose the top-ranked article t_b from R_x.
5. One evaluator reads article t and t_b, and then evaluates how much difference information t and t_b have comparing H in five grades, respectively.
6. $a \leftarrow t$. $a_b \leftarrow t_b$. $A = A - t$, $H = H \cup t$. $R_c = R_c - t_b, R_o = R_o - t_b, R_d = R_d - t_b$.
7. If A is not empty, go to step 3; else stop.

Where, a and t are used for our context-aware re-ranking method while a_b and t_b are used for the baseline method.

The five grades evaluation has been conducted to estimate how much different information can be obtained from the new article comparing with the articles have been read before from three viewpoints in user survey [1]: relevance, opinion, and the amount of additional information. The evaluator scored articles

Table 4. Experimental results of re-ranking method

Event	Event 1	Event 2	Event 3	Event 4	Event 5	Event 6
Baseline	3.00	2.67	3.67	3.83	2.17	3.00
Re-ranking	3.50	3.33	3.67	3.83	2.17	3.17

comprehensively from these three viewpoints. There are six news events' results of baseline and our re-ranking methods are different. We calculated the average of user scores of these six events.[4] The results are shown in Table 4.

Among the six events, the user scores of re-ranking method are greater than or equal to those of baseline. This result reveals the availability of our re-ranking method to enhance users to read more diverse information.

6 Conclusion

In this paper, we propose three entity-oriented ranking measures to support users obtaining diverse reports on the same news event. As one application of these ranking measures, we develop a diversity-seeking news app on smart mobile devices. A context-aware re-ranking method is also proposed to provide more diverse information. We conducted crowdsource experiment to validate our entity-oriented ranking measures. The context-aware re-ranking method is validated by a simulation-based analysis.

As one of the important future work, we need conduct user study of our news app and re-ranking method. In addition, we plan to carry out experiments to investigate the change of user's media literacy by using our system.

Acknowledgement. This work is partly supported by KAKENHI(No. 25700033) and SCAT Research Funding.

References

1. Kiritoshi, K., Ma, Q.: Named entity oriented related news ranking. In: Decker, H., Lhotská, L., Link, S., Spies, M., Wagner, R.R. (eds.) DEXA 2014, Part II. LNCS, vol. 8645, pp. 82–96. Springer, Heidelberg (2014)
2. Park, S., Kang, S., Song, S.C.J.: Newscube: delivering multiple aspects of news to mitigate media bias. In: CHI 2009, pp. 443–452 (2009)
3. Ogawa, T., Ma, Q., Yoshikawa, M.: News bias analysis based on stakeholder mining. IEICE Trans. **94–D**, 578–586 (2011)
4. Xu, L., Ma, Q., Yoshikawa, M.: A cross-media method of stakeholder extraction for news contents analysis. In: Chen, L., Tang, C., Yang, J., Gao, Y. (eds.) WAIM 2010. LNCS, vol. 6184, pp. 232–237. Springer, Heidelberg (2010)

[4] Notice the other nine news events have same scores.

5. Ma, Q., Yoshikawa, M.: Topic and viewpoint extraction for diversity and bias analysis of news contents. In: Li, Q., Feng, L., Pei, J., Wang, S.X., Zhou, X., Zhu, Q.-M. (eds.) APWeb/WAIM 2009. LNCS, vol. 5446, pp. 150–161. Springer, Heidelberg (2009)
6. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. **3**, 993 (2003)
7. Jarvelin, K., Kekalainen, J.: Evaluation, cumulated gain-based, of IR techniques. ACM Trans. Inf. Syst. **20**, 422–446 (2002)
8. Ma, Q., Nadamoto, A., Tanaka, K.: Complementary information retrieval for cross-media news content. Inf. Syst. **31**(7), 659–678 (2006)