A Comparative Study of Diversity in Search Results of Search Engines

Malik M. Saad Missen¹, Dost Muhammad Khan², Muhammad Mughees Aqib³
¹, ², ³ Department of Computer Science and Information Technology, The Islamia University of Bahawalpur, Pakistan

ABSTRACT

The effectiveness of web information retrieval systems could be measured through many factors and one of the most important elements is the diversity of returned search results. In this paper, we evaluate the diversity of two very popular search engines (i.e. Google and Yahoo) for multi-faceted queries and compare them. We use entity-based approach for this purpose by taking into consideration the novelty of the entities. We evaluate the diversity of these search engines on behalf of 1) Common Entities, 2) Novelty, and 3) Grouping the documents. We use a manually collected data collection for our experiments. We find out that search results returned by Yahoo are more diverse than those of Google. Google results are found to be more coherent in their similarities as far as common entities are concerned.

Keywords: Entities; Diversity; Novelty; Search Engines

INTRODUCTION

Controversial topics or queries are debatable topics (such as “Cousin Marriages”, “Family planning”, etc.) that might have a huge number of relevant documents on the web with their sub-sets containing different opinions of a specific topic. Users using these types of queries as their search string are usually interested in getting different points-of-views opinions on some topic. For example, if a user enters a query, “Energy Crisis” as a search string of a search engine to know different points of views of people about this subject but as result user might get a set of documents exposing only one aspect of the given topic i.e. all returned documents discuss only one side of the search topic. In other words, the returned results are less diverse because the user has been deprived of other opinions about this topic. For such controversial queries, it is more important to present diversified documents higher than presenting a coherent set of similar documents. This diversity in search results can be measured and search information systems can be evaluated. Generally, diversity in text documents has been referred to: Diversity of sources (multiplicity of sources of texts and images), Diversity of resources (e.g. images, text), Diversity of topic, Diversity of viewpoint, Diversity of genre (e.g. blogs, news, comments), Diversity of language, Geographical/spatial diversity and temporal diversity¹.

This work is limited to evaluating the diversity of
viewpoint for a certain topic. We aim to compare the diversity of two very popular search engines for such controversial queries considering the diversity of viewpoints. We compute the diversity through the novelty approach. To be more specific and research-oriented, we aim to evaluate the diversity of search engines for a chosen set of controversial queries using entity-novelty approach. The basic idea behind choosing this approach is that if a document $D_1$ contains a different opinion on a certain topic than a document $D_2$ then it is likely to introduce some new entities in its content. It is assumed that both documents $D_1$ and $D_2$ are relevant to the given topic. The rest of the paper is organized as follows: Section 2 discusses the major work relevant to our work while we unveil our data collection in section 3. Section 4 describes our experiments in detail and we conclude this paper at the end of section 5.

RESEARCH METHODOLOGY

There is much search related to diversity and novelty of documents. They use different approaches for this purpose. So we divide our work into to two topics, i.e. Diversity and Novelty.

2.1 Diversity

There are following different researchers who work on diversity:

One of the works suggested the click entropy method\(^2\) for measuring the diversifying search results. The method minimizes the redundancy among different search results of the ambiguous queries. Another group of researchers performed analysis in the “Evaluation measures”\(^3\) in diversity search results of the various systems after comparing them. Their purpose was to develop an optimized system. Paramita and her colleagues tried to use test collections to promote diversity\(^4\) with their own test sets results. An Italian research team presented an Opinion diversification Model\(^5\) that discussed the relevance and irrelevance of documents on the basis of sentimental and semantic analysis. They categorized sentiments into three main categories i.e. Positive, Negative and Neutral about the query. The work by Agrawal and his colleagues\(^6\) focuses on an algorithm that compute relevancy of the documents on the web search. This algorithm exhibits the approximation of good search results. Algorithm defines the metrics for generalization of the solutions of ambiguous queries. Dang and other researchers proposed the idea of diversity by proportionality\(^7\). They considered a result list most diverse, with respect to some set of topics related to the query. Their technique iteratively determined, for each position in the result ranked list, the topic that best maintains the overall proportionality. There was a proposal for an online learning model\(^8\) and algorithm for learning rankings that balance relevance and diversity. In each step, the algorithm presents a ranking to the user. As feedback, the algorithm observes the set of documents the user reads in the presented ranking. Similarly, another group of researchers developed an algorithm\(^9\) for users to satisfy the user’s need. Algorithm works on multiple relevant documents using probabilistic approach to satisfy average user’s interest on subtopic.

2.2 Novelty

In this subsection, we highlight some work done on topic of Novelty. Leelanupab and colleagues\(^10\) evaluated a system which retrieves the topics. They set the parameter to fetch better novel topics. Carterette\(^11\) worked and concluded that if algorithm ranks the $k$ documents which provide the novelty and diversity then their worst case is NP-completeness problem. And he also concluded that in many cases the greedy algorithm provide best result. The work by Clarke and others\(^12\) maps the relationships between the components of information nuggets and relevance nuggets of information. Moghaddam and colleagues\(^13\) proposed a system for business point of view according to user’s behavior and similarity. The suggested approaches to opt the recommender system are item popularity to raise accuracy of novelty and diversity. The recommend system improved the accuracy of novelty and diversity satisfying the user’s needs in E-commerce. In the recommender system “Novelty” is referred quality of system avoid redundancy. Diversity referred set of items different from each other. Items showed “Novel” behavior with respect to other trade items.

3. Data Collection

We use 25 topics. These topics are controversial debatable issues of Pakistan. The debatable queries are that queries, which different people use different opinions.
These topics are: Match fixing scandal in Pakistan, Love marriage or arrange marriage in Pakistan, Second marriages in Pakistan, Should Pakistan declare India as best nation?, Bahawalpur province controversy, Kala bagh dam controversy, PTI, PPP or PML (n) which is good, Tahir ul Qadri long march controversy, Cousin marriages in Pakistan, Co-education in Pakistan, Family planning in Pakistan, Government job or business in Pakistan, Combine family in Pakistan, Budget distribution controversy of Pakistan, Democracy in Pakistan, Marriage in same cast in Pakistan, Metro bus controversy in Pakistan, Laptop scheme controversy in Pakistan, CNG vs. Petrol in Pakistan, Feminism in Pakistan, Privatization in Pakistan, Media Role controversy in Pakistan, Load shedding in Pakistan, Bureaucracy Good or Bad in Pakistan, School vs. Madrassa System in Pakistan, We use these topics as queries in search engines. We use at least 200 top ranked documents for each query from each search engine. So in total we collect 200 + 200 = 400 document for each query. And 200 * 25 = 5000, for each search engine. Thus we are working overall at least 200 * 25 * 2 = 10000 documents.

RESULTS & DISCUSSION

As we have discussed already that we focus on finding diversity in search results of Google and Yahoo search engines for topics with the expected diversified point of views (like Iran-Iraq war). For this purpose, we exploit the novelty of entities as computed from the returned search results of both search engines. For this research goal, we have focused on three research tasks listed below:

i. To evaluate the role of common entities between two documents to compute diversity.

ii. To evaluate the role of novelty between two documents to compute diversity.

iii. To evaluate the role of grouping the documents to compute diversity.

Task I

Generally, search results returned from a search engine could share the same set of entities (more or less) because these documents discuss the same topic. However, for a controversial topic like Iraq-Iran war, the set of entities to be discussed could be different because of multi-dimensionality of the topics. Therefore, we evaluate the diversity among search results of a search engine by considering the number of common or same entities between them. We assume that less the number of common entities, higher the diversity of search results and vice versa. We find average common entities of a topic by using following formula in Eq. (1):

$$\text{Average Common Entities} = \frac{\sum_{i=1}^{n} (d_i \rightarrow d(i+1))}{n}$$

(1)

Where d is document and n is a total number of documents on a specific topic. $d_i \rightarrow d(i+1)$, represents the common entities between document $d_i$ and document $d(i+1)$.

The result of this task is shown in Figure 1.

![Figure 1: Common Entities w.r.t. topics](image)

In this graph of Figure 1, X-axis shows the average common entities, Y-axis shows the controversial topics. The first bar is for Google and second bar is for Yahoo for the same topic. In the course of Evaluation, we concluded that for some topics Yahoo does provide a good deal of diversity in Search Results and in the same manner there are some topics for which the results showed up by Google are much more varied. But, as a matter of fact, the comparative analysis of the overall results depicts that diversity is a distinctive property of Google as it provides
one with a variety of responses for way more topics than any other Search engine -Conclusively, Google outshines Yahoo in regards of providing diversity with respect to common entities.

**Task II**

In the Previous Task, we evaluated the diversity on the basis of Common or same entities. Now we ought to evaluate the diversity among search results of a search engine by considering the novelty amongst them. We assume that higher the number of novel terms, higher the diversity of search results and vice versa. In order to calculate the novelty amongst any two documents (let's call them document A and document B) returned in search results, we use the following formula in Eq. (2):

\[ \text{Novelty} = \frac{\text{CE}(B-A)}{\text{CE}(A+B)} \]  

(2)

Where CE(B-A) is a number of unique entities of document B which are not in document A and CE (A+B) is a number of entities of document B which are already present in document A.

Average Novelty is calculated using the formula in Eq. (3).

\[ \text{Average Novelty} = \frac{\sum(\text{Novelty})}{n} \]  

(3)

Where n is a total number of documents on a topic. The result of this task is shown in Figure 2.

![Figure 2: Novelty w.r.t. topics](image)

In this graph of Figure 2, X-axis shows the average Novelty, Y-axis shows the controversial topics. The first bar is for Google and second bar is for Yahoo for the same topic. In the course of Evaluation, we concluded that for some topics Yahoo does provide a good deal of diversity in Search Results and in the same manner there are some topics for which the results showed up by Google are much more varied. But, as a matter of fact, the comparative analysis of the overall results depicts that diversity is a distinctive property of Yahoo as it provides one with a variety of responses for way more topics than any other Search engine -Conclusively, Yahoo outshines Google in regards of providing diversity with respect to novel entities.

**Task III**

For this particular task, we evaluate the diversity by the number of groups considering the number of common entities between documents for making a group. If this number happens to be high the diversity is low. For making groups of a particular document, we first calculate the average for that document, which is given in Eq. (4):

\[ \text{Average} = \frac{\text{total number of common entities for the topic}}{\text{number of documents which have a greater number of common entities for the particular document}} \]  

(4)

This average is our starting point for groups; documents, which have common entities below this average, are ignored. Then we specify a range for a group; which is calculated using Eq. (5):

\[ \text{Range} = \frac{\text{Average}}{2} \]  

(5)

So, for the first group, the common entities of other documents to the particular document are from average to average + range and for another group, the range of common entities would be the stop point of previous group + range. We built a table of both search engines for this task.

<table>
<thead>
<tr>
<th>Topics</th>
<th>Groups Formed</th>
<th>Group Selected (documents &gt; 2)</th>
<th>User Evaluation</th>
<th>Average Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bahawalpur Province Controversy</td>
<td>7</td>
<td>4</td>
<td>15 / 31 = 0.484</td>
<td>0.405</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 / 19 = 0.053</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 / 11 = 0.272</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 / 3 = 0.333</td>
<td></td>
</tr>
<tr>
<td>Bureaucracy Good Or Bad</td>
<td>9</td>
<td>5</td>
<td>9 / 18 = 0.5</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5 / 20 = 0.25</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4 / 8 = 0.667</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 / 3 = 0.333</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0 / 3 = 0</td>
<td></td>
</tr>
<tr>
<td>Budget Distribution</td>
<td>7</td>
<td>7</td>
<td>6 / 19 = 0.316</td>
<td>0.389</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>10 / 13 = 0.769</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4 / 13 = 0.308</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 / 8 = 0.25</td>
<td></td>
</tr>
<tr>
<td>Topic</td>
<td>Groups</td>
<td>Documents</td>
<td>Group Selected (Documents &gt; 2)</td>
<td>User Evaluation</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>--------</td>
<td>-----------</td>
<td>--------------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Match Fixing Scandal</td>
<td>7</td>
<td>4</td>
<td>24 / 26 = 0.923 22 / 26 = 0.88 4 / 4 = 1 3 / 3 = 1</td>
<td>0.951</td>
</tr>
<tr>
<td>Media Role Controversy In Pak</td>
<td>8</td>
<td>4</td>
<td>0 / 21 = 0.296 9 / 19 = 0.474 1 / 7 = 0.143 4 / 7 = 0.571</td>
<td>0.368</td>
</tr>
<tr>
<td>Metro Bus Controversy</td>
<td>7</td>
<td>5</td>
<td>13 / 17 = 0.765 3 / 7 = 0.426 1 / 3 = 0.333 0 / 3 = 0 3 / 4 = 0.75</td>
<td>0.455</td>
</tr>
<tr>
<td>Is Privatization Good or Bad</td>
<td>9</td>
<td>4</td>
<td>21 / 30 = 0.7 11 / 13 = 0.846 9 / 11 = 0.818 3 / 3 = 1</td>
<td>0.841</td>
</tr>
<tr>
<td>PTI, PPP or PML (N) Which Is Good</td>
<td>4</td>
<td>3</td>
<td>27 / 31 = 0.871 17 / 20 = 0.85 7 / 7 = 1</td>
<td>0.907</td>
</tr>
<tr>
<td>School vs. Madrassa System In Pak</td>
<td>5</td>
<td>4</td>
<td>22 / 34 = 0.647 7 / 12 = 0.563 2 / 3 = 0.667 3 / 3 = 1</td>
<td>0.724</td>
</tr>
<tr>
<td>Second marriages in Pak</td>
<td>6</td>
<td>4</td>
<td>6 / 14 = 0.428 0 / 3 = 0 0 / 3 = 0 0 / 3 = 0</td>
<td>0.107</td>
</tr>
<tr>
<td>Should We Declare India as Best Nation</td>
<td>8</td>
<td>5</td>
<td>2 / 24 = 0.083 0 / 13 = 0 0 / 6 = 0 0 / 6 = 0</td>
<td>0.017</td>
</tr>
<tr>
<td>Tahir ul Qadri Long March</td>
<td>6</td>
<td>3</td>
<td>33 / 43 = 0.767 16 / 19 = 0.842 9 / 11 = 0.818</td>
<td>0.809</td>
</tr>
<tr>
<td>Total Average Accuracy</td>
<td></td>
<td></td>
<td></td>
<td>0.521</td>
</tr>
</tbody>
</table>

Table 2: Grouping Information for Google
In the course of Evaluation, we conclude that for some topics Yahoo does provide a good deal of diversity in Search Results and in the same manner there are some topics for which the results showed up by Google are much more varied. But, as a matter of fact, the comparative analysis of the overall results depicts that diversity is a distinctive property of Yahoo as it provides one with a variety of responses for way more topics than any other Search engine - Conclusively, Yahoo outshines Google in regards of providing diversity with respect to grouping the documents.

Our research was aimed at comparing the diversity of viewpoint in search results of search engines, which for this matters were Google and Yahoo. For this purpose,
we determined diversity by novelty and using entity method. We specified the tasks for analyzing the novelty. The proposed tasks analyze the diversity by the following consideration: 1) Common Entities, 2) Novelty, and 3) Grouping the documents. In the task-1, we worked on common or same entities between two documents. In this task, higher the common entities mean lower the diversity. We evaluated that Google provides more diversity than Yahoo, regarding the number of common entities between documents. In the task-2, we worked on the novelty of two documents. In this task, higher the novelty means higher the diversity. We evaluated that Yahoo provides more diversity than Google, regarding the novelty between documents. In the last task, i.e. task-3, we work on grouping the documents. In this task, a number of groups mean more diversity. We evaluated that Yahoo provides more groups than Google, regarding the grouping the documents. Although it was observed from these tasks that both engines provide diversity after the evaluation and result, we conclude that the variety of search results showed up by Yahoo is way more than Google. Thus Yahoo has been proved as a better Search engine in regards to diversity in search results.

REFERENCES