Spatial Diversity, Do Users Appreciate It?

Jiayu Tang
Department of Information Studies
University of Sheffield, UK
j.tang@shef.ac.uk

Mark Sanderson
Department of Information Studies
University of Sheffield, UK
m.sanderson@shef.ac.uk

ABSTRACT

Spatial diversity is a relatively new branch of research in the context of spatial information retrieval. It tries to answer user's query with results that are not only relevant but also spatially diversified so that they are from many different locations. Although the assumption that spatially diversified results may meet users' needs better seems reasonable, there has been little hard evidence in the literature indicating so. In this paper, we will show our follow-up work on the novel approach to investigating user preference on spatial diversity by using Amazon Mechanical Turk.

Categories and Subject Descriptors

H.1.2 [**Information Systems**]: Models and Principles—*Human factors*; G.3 [**Mathematics of Computing**]: Probability and Statistics—*Experimental Design*

General Terms

Experimentation, Algorithms

Keywords

Amazon Mechanical Turk, User Study, Spatial Diversity

1. INTRODUCTION

Spatial diversity, as a special kind of diversity in spatial information retrieval, aims to bring documents that are not only relevant but also diversified spatially to the user. In the work of [4], the phrase "high spreading" is used to express the notion of spatial diversity. Paramita *et al.* [1], on the other hand, described spatial diversity as "a way to present diverse results to users by presenting documents from as many different locations as possible". [3] defined spatial diversity as "a measure of location coverage". The more locations that are covered and more intense the coverage is, the better spatial diversity a list of documents achieves.

GIR'10, 18-19th Feb. 2010, Zurich, Switzerland

Table 1: Sample Topics

ID	Query
1	Botanical Garden
13	Lakes with swimming birds
15	Footpath in North York Moors
22	Places for Climbing in Peak District
29	Ruins of Roman Wall

2. DATA-SET

We have used an image collection called the Geograph data-set for our diversity experiment. The data-set, which contains 759,638 images, are part of the Geograph British Isles project[1] which aims to collect geographically representative photographs for every square kilometre of Great Britain and Ireland. The National Grid reference system invented by Ordnance Survey (OS) was used to divide the map into grids at three granularity levels, 100km by 100km, 10km by 10km and 1km by 1km. Thus, each image has a reference number indicating from which cell it comes.

We indexed the text fields of each image with the open source search engine Lucene, and generated a document list for one of 30 topics (See Table 1 for some samples). The document lists were then re-ranked by one of the two spatial diversity algorithms, named as GM and SC-1 [2]. Both the algorithms work by re-ranking the standard search results so that documents from different locations are promoted to the top of the list, while trying to maintain the precision.

3. SETUP OF USER STUDY

In order to investigate if users do favor spatially diversified results over standard results, we ran a user preference study in Amazon Mechanical Turk (MTurk). The idea is simple - by showing two maps (one is the standard and the other is the spatially diversified) in which search results are depicted, we ask the MTurk users to decide which one they prefer based on the corresponding query.

Specifically, we compared two spatial diversity algorithms with the standard search. For each of the 30 topics, we generated two maps. One map was plotted with the top 10 results from the standard search, and the other was plotted with the top 10 results from one of the spatial diversity algorithms. Search results were plotted on the locations where the photos were taken. The users were given three options: 1) the left map is better; 2) no preference/similar; 3) the right map is better. We call this a preference judge-

[1] www.geograph.org.uk

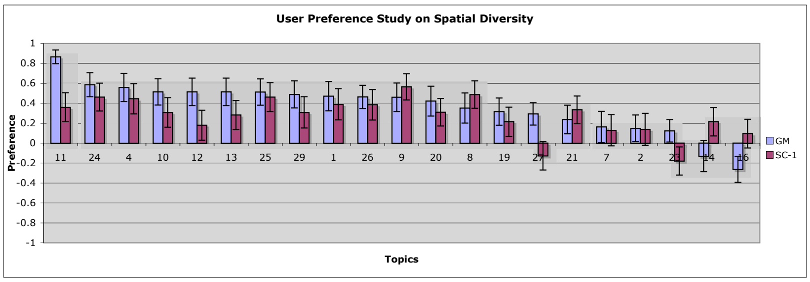

Figure 1: User preference on spatial diversity with the Geograph Data-set. Preference value equals to 1 if the users prefer spatially diversified results, -1 if they prefer the standard results.

ment unit (PJU). It should be noted that since documents are represented as red circles rather than the actual images, this experiment is only concerned with the spatial dimension, i.e. spatial diversity and spatial relevance. In other words, content of the images are not considered.

We have used a combination of several methods to control the quality of judgements. Firstly, we required the users to have a minimal approval rate of 90%. The approval rate indicates how much of all the work done by the user has been accepted. Secondly, we randomly inserted some "trap" PJUs into each HIT (Human Intelligence Task). A HIT is a piece of work submitted to MTurk to be completed by the users. A "trap" is a manually constructed PJU in which the preference of two maps is so obvious that every genuine user should be able to tell easily and correctly. We only used submissions from users who correctly answered all the trap PJUs of a HIT. 30 topics were grouped into 4 HITs with 7, 7, 8 and 8 topics respectively. Trap PJUs were then injected into each HIT to make the topic number per HIT to be 10. For each HIT, 50 different users were requested.

In order to conduct a fair experiment, we have randomized not only the order of topics, but also the order of the two maps in each PJU. Besides, to void biasing the users' judgement, we did not give any mention of the notion or the word "diversity" in the task description.

4. RESULTS

After filtering out the work that failed to pass all the traps from a particular HIT, we have accepted an average of 38 judgements per topic (about 76% of the 50). We noticed that diversity algorithms are unfavored on some topics with geographical constraints as too many spatially irrelevant documents were promoted. For example, in topic "Footpath in North York Moors", documents from outside of "North York Moors" were promoted by diversity algorithms because they are far away from previous documents. This is however more of a problem with the search engine than the diversity algorithms, because the search engine used in this work was unable to tell such documents were spatially irrelevant. Figure 1 shows the average user preference for 21 topics, after removing the 9 geographically constrained ones.

Since each Geograph image has a reference number indicating to which cell of the National Grid reference system it belongs, we can calculate the number of different cells covered by the top results from each algorithm. This number gives an implication of how diverse in visual the results are on the map. Intuitively, a set of more spatially diversified results will cover more cells. We have chosen the 100km by

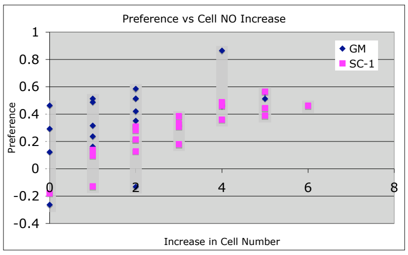

Figure 2: Scatter Plot of User Preference vs. Increase in Cell Number

100km level of grid, in which the entire UK was divided into 100km by 100km cells. A scatter plot has been produced, to examine the correlations between users' preferences and the increase in cell numbers. As shown in Figure 2, the horizontal axis represents the increase in cell number by a diversity algorithm over the standard method. The vertical axis is the average user preference. Each point represents a topic. Overall, it seems that the more increase in cell number, the more likely it is preferred by users.

5. CONCLUSIONS

According to our novel user experiment on MTurk, users do have strong preference on spatially diversified results. More diversified results seem to receive stronger preference.

6. ACKNOWLEDGMENTS

Work undertaken in this paper is supported by the EU-funded Tripod project (Contract No. 045335).

7. REFERENCES

[1] M. L. Paramita, J. Tang, and M. Sanderson. Generic and spatial approaches to image search results diversification. In *ECIR*, pages 603–610, 2009.

[2] J. Tang and M. Sanderson. The TRIPOD project technical report - modeling diversity in spatial information retrieval using spatial coverage. Technical report, University of Sheffield, 2009.

[3] J. Tang and M. Sanderson. Evaluation and user preference study on spatial diversity. In *ECIR*, 2010.

[4] M. van Kreveld, I. Reinbacher, A. Arampatzis, and R. van Zwol. Distributed ranking methods for geographic information retrieval. In *Proceedings of the 20th European Workshop on Computational Geometry*, pages 231–243, 2004.