The 1st International Workshop on Diversity in Document Retrieval

Craig Macdonald
School of Computing Science
University of Glasgow
UK
craig.macdonald@glasgow.ac.uk

Charles Clarke
School of Computer Science
University of Waterloo
Canada
claclark@plg.uwaterloo.ca

Jun Wang
Department of Computer Science
University College London
UK
jun.wang@cs.ucl.ac.uk

Abstract

When an ambiguous query is received, a sensible approach is for the information retrieval (IR) system to diversify the results retrieved for this query, in the hope that at least one of the interpretations of the query intent will satisfy the user. Diversity is an increasingly important topic, of interest to both academic researchers (such as participants in the TREC Web and Blog track diversity tasks), as well as to search engines professionals. In this workshop, we solicited submissions of both technical papers and position papers on three themes: approaches and models for diversity; the evaluation of diverse search results; and on applications of diverse search results. In the workshop, two invited talks, along with ten refereed papers were presented before breakout groups considered questions and issues with the three workshop themes.

1 Introduction

The query is often the only evidence of the user's underlying information need available to an IR system. However, if the query is ambiguous, a sensible approach is to diversify the results retrieved for this query, in the hope that at least one of the interpretations of the query intent will satisfy the user. Diversification is now becoming an important topic in IR (driven by the TREC Web & Blog tracks, and through deployment by search engines).

The 1st International Workshop on Diversity in Document Retrieval (DDR 2011) was held in conjunction with the 33rd European Conference in Information Retrieval in Dublin, Ireland on 18th April 2011. The goal of the workshop was to create a forum for researchers interested in various aspects of diversity. In particular, we identified three themes for the workshop, and called for both technical and position papers in each of the three themes:

Modelling: Diversifying search results usually involves a departure from the independent document relevance assumption underlying the well-known probability ranking principle. An IR system should consider not only the relevance of each individual document, but also how relevant the document is in light of the other retrieved documents. Various models exist, but often form around two schools, namely implicit [4, 19] and explicit [1, 15], depending on how the query aspects are accounted for in the retrieved documents.

Evaluation: A diverse ranking of results should have documents matching as many of the possible aspects underlying a query (known as diversity or coverage), while still promoting novelty, whereby one aspect does not become over-represented. Several evaluation measures for diversity have been proposed, such as α-NDCG [8], Intent Aware-Precision [1], and Intent Aware-Expected Reciprocal Rank and Subtopic-Precision measures. The advent of the diversity tasks in the TREC 2009 Web and Blog tracks have created means for the evaluation of diversification approaches [7, 9]. Other problems include how the importance of different query aspects should be taken into account during evaluation.

Applications: Diverse rankings have applications in various areas. One primary application is of course in Web search (as investigated by the TREC Web track [7]). However, applications are also possible in social search (e.g. find me diverse opinions of a topic on the blogosphere - c.f. TREC Blog track [9]), in product search (e.g. I'm not sure what kind of digital camera I want to buy, show me a selection of different types), and in summarization [4].

Diversity is an increasingly important topic, of interest to both academic researchers, as well as to search engines professionals. In particular, there are now an increasingly large number of papers submitted to SIGIR, CIKM & WSDM on diversity, so there now is a sizable community working on the issue.

2 Program Committee

Ben Carterette, University of Delaware, USA
Olivier Chapelle, Yahoo! Labs, USA
Nick Craswell, Microsoft, UK/USA
Ben He, Chinese Academy of Science, China
Jaap Kamps, University of Amsterdam, the Netherlands
Jian-Yun Nie, Université de Montréal, Canada
Iadh Ounis, University of Glasgow, UK
Filip Radlinski, Microsoft, Canada
Tetsuya Sakai, Microsoft Research Asia, China
Rodrygo Santos, University of Glasgow, UK
Ryen White, Microsoft Research, USA
Jianhan Zhu, University College London, UK

3 Workshop Program

The workshop program [10] included two invited talks, five position papers and five technical papers, which culminated in three breakout groups:

Invited Talks

- *Challenges in Diversity Evaluation* [14] Tetsuya Sakai

- *Analysis of Document Diversity through Sentence-Level Opinion and Relation Extraction* [11] Alessandro Moschitti

Accepted Position Papers

- *Towards the Foundations of Diversity-Aware Node summarization on Knowledge Graphs* [18] Marcin Sydow
- *Diversifying for Multiple Information Needs* [16] Rodrygo Santos and Iadh Ounis
- *A Search Architecture Enabling Efficient Diversification of Search Results* [3] Gabriele Capannini, Franco Maria Nardini, Raffaele Perego and Fabrizio Silvestri
- *Diversity in Expert Search* [12] Vassilis Plachouras
- *SOPHIA: Bridging the Gap between Thematic Modelling to Interactive Diverse Search* [13] Niall Rooney, David Patterson and Vladimir Dobrynin

Accepted Technical Papers

- *Analysis of Various Evaluation Measures for Diversity* [6] Praveen Chandar and Ben Carterette
- *Novelty and Diversity Metrics for Recommender Systems: Choice, Discovery and Relevance* [5] Pablo Castells, Saúl Vargas and Jun Wang.
- *Diversification of Search Results as a Fuzzy Satisfiability Problem* [17] Steven Schockaert and Martine De Cock
- *A Comparative Study of Search Result Diversification Methods* [20] Wei Zheng and Hui Fang
- *Explicit Query Diversification for Geographical Information Retrieval* [2] Davide Buscaldi and Paolo Rosso

3.1 Session 1: Evaluation

The evaluation session kicked off with the invited talk of Tetsuya Sakai [14] who discussed his views on the evaluation of diversity in document retrieval. These views are particularly informed by his experience as an organizer of diversity-related tasks at NTCIR[1], and he provided a status report on NTCIR as part of his talk. In the remainder of the talk, he surveyed current approaches and measures used for evaluating diversity at TREC and NTCIR, including various intent-aware measures, with an eye to possible improvements. He reviewed a number of challenges related to evaluating diversity, illustrating shortcomings in current approaches. He highlighted the potential for problems when the number of distinct intents is large, the lack of proven methods for estimating intent likelihoods, and the need to carefully consider the balance between relevance and novelty (i.e., between overall precision and subtopic recall). Finally, looking beyond the ranked list, he considered approaches to evaluating whole page relevance and evaluation across search sessions, which might include user navigation and query reformulation as part of the session.

Following the invited talk, the evaluation session featured three contributed talks. The first, a position paper by Marcin Sydow [18], considered the problem of defining and evaluating the concept of diversity for graphical node summarization in knowledge multi-graphs.

[1] `http://research.nii.ac.jp/ntcir`

Given a knowledge multi-graph, with nodes representing entities and arcs representing relationships, node summarization extracts a subgraph that best summarizes knowledge related to a given node q under size and other constraints. The talk explored several ways that this problem could be formalized to capture a notion of diversity in the subgraph.

Next, Praveen Chandar, presented a technical paper analyzing evaluation measures for novelty and diversity, focusing on those used at TREC [6]. The talked echoed and extended some of the concerns raised in the invited talk. Working with the diversity runs from TREC 2009, and generating additional synthetic runs to complement them, he reported an analysis of variance that explored the sources of variation in current measures, including components related to relevance and novelty. This analysis indicates that different measures are sensitive to different factors, some of which are undesirable.

The final contributed talk explored novelty and diversity measures for recommender systems [5], presented by Pablo Castells. Again, participants heard about the importance of striking a balance between relevance and novelty (or in the case of recommender systems the corresponding concerns of popularity and similarity). Models of item novelty based on item choice and discovery were developed, along with estimation methods based on usage data and item ratings. Metrics for novelty and diversity were proposed, along with experimental results over MovieLens data.

3.2 Session 2: Modelling

In the modelling session, Rodrygo Santos [16] presented his XQuAD approach for user-driven diversification. xQuAD modifies the results retrieved for a query by using previous user interactions to identify possible interpretations or *aspects* of the query, and then diversifying the results to ensure that results for each of these aspects are well (but not over) covered.

Next up, France Maria Nardini presented the OptSelect architecture for diversification [3]. In particular OptSelect also uses query logs to identify specializations of a query. Then the initial set of documents identified for the query can be diversified by examining – in light of the specializations – the shingles that are stored for each document. These shingles are usually used to identify duplicate documents in the result lists, but have not previously been deployed for diversification. Moreover, the OptSelect architecture presents certain efficiency advantages in that it does not require access to either the direct index nor accessing the inverted index for multiple queries.

Steven Schockaert presented the notion that diversification is a combinatorial optimization problem [17]. He then draws the link to fuzzy logics, by defining a language for encoding constraints on fuzzy logic formulae. With this language, various intuitions for diversification can be encoded, as well as existing diversification approaches from the literature.

Finally, the work of Wei Zheng and Hui Fang (presented by Ben Carterette) contains an empirical investigation comparing three diversity approaches, namely MMR, WUME and xQuAD [20]. In particular, they performed several simulated experiments to compare and contrast each of the techniques, for instance changing the effectiveness of the underlying document ranking, or the number of aspects of the query that are modelled by explicit techniques. Among their findings, they found that the effectiveness of a diversification method is negatively correlated to the underlying document ranking.

3.3 Session 3: Applications

The applications session started with the invited talk of Alessandro Moschitti from University of Trento [11]. Moschitti presented an analysis on diversification through Sentence-Level Opinion and Relation Extraction. The argument was that using the current state-of-the-art in Relation and Opinion Extraction at the sentence level would improve document retrieval.

We then had three papers introducing the applications of document diversity ranging from expert search and interactive search to geographical information retrieval. Specially, Vassilis Plachouras demonstrated that diversity can be beneficial in expert search, where users are looking for people or organizations with relevant expertise [12]. In the paper, a new measure was defined to quantify the topical ambiguity, and it has been tested over an operating expert search engine. A more detailed discussion about the specific diversify problem in expert search was given at the end of the talk. Next, Rooney et al. from University of Ulster and sophiasearch.com presented an interactive retrieval system called Sophia [13]. They described various facilities to provide diverse search for knowledge workers in various domains of usage. At last, Buscaldi and Rosso provided their attempt to evaluate the potential benefits of diversity in the geographical information retrieval task [2]. In geographical information retrieval, documents are relevant not only from a thematic point of view, but also need to be spatially related. The diversity was achieved by reformulating queries with the help of a geographical ontology. The reported results show that a theoretical improvement is possible in the case that the relevant documents do not contain enough geographical data.

In summary, this was a very interesting session and three important applications areas were covered and well discussed from the speakers and audience.

3.4 Session 4: Breakout Groups

After the plenary sessions, the workshop split into breakout groups tackling the three main themes of the workshop. In the following, we report the discussion arising from each group

Evaluation The evaluation breakout session primarily discussed ways to make the current evaluation efforts at TREC and NTCIR better, both by improving the current tasks and introducing new tasks. Current TREC tasks require groups to generate retrieval runs over a very large (billion document) test collection, which some groups may find difficult. Participants suggested that the creation and distribution of a good baseline run would allow groups to submit runs that simply diversified this baseline, lowering the bar. Both evaluation efforts should provide more ways to explore individual components of the problem, particularly intent analysis and run diversification. The re-usability of the test collection should be considered, perhaps through crowdsourcing or other ways of cheaply growing the set of judgments.

Some ideas for new tasks were developed, as follows: 1) a snippet generation task, where subtopics would be highlighted in the snippets and judging would be done on the basis of the snippets; 2) a query suggestion task, where subtopics could be proposed and user browsing could be simulated and incorporated into the evaluation measures; 3) one or more tasks to examine diversity in other verticals, including news and microblogs. The moderator of the session (Clarke, who is one of the organizers of the TREC tasks) expressed his intention to carry these ideas forward as proposals for TREC 2012.

Modelling The modelling group firstly considered the dimensions within modelling of the diversification process, and thereby identified users, categories, documents (including semi-structure), entities, opinions and emotions, as well as graphs. The participants iden-

tified that there is not yet a killer model that can encapsulate all of these dimensions for diversification. Indeed, the nature and components of such a model may vary with the application, such as recommendation, Web search (cross- or within-vertical), ads, opinions and summarization. Future directions that were identified that are not yet well handled is the role of user context in diversity - for instance, a query which appears ambiguous at first sight may be better specified with knowledge of the user, allowing search results to be personalised instead of diversified. Moreover, there is thought to be further work needed on developing frameworks that permit diversification to be learned.

Applications The applications breakout session was intended to brainstorm the possible use of diversity in a boarder area such as data mining, information retrieval, natural language processing and multimedia. It started with a list of possible applications and discussed the fundamental challenges among those applications to see whether the diversify would be any help. The conclusion was that many applications suffer the following difficulties 1) the ambiguity of the information needs 2) relevant information can be found from multiple sources and media types. Diversifying the search results would certainly reduce the risk of not finding anything interesting and relevant. This calls for research on a more general diversity model that is not limited to text retrieval only.

4 Conclusions

After the breakout groups had finished, a representative of each breakout group presented the findings of their group. Specifically, each of the breakout groups contributed to our understanding of the way forward. In particular, the evaluation breakout group identified ways in which the evaluation of diversity approaches could be tackled in more fine-grained manners. The modelling group identified different dimensions that are not well integrated by diversification approaches, and noted the task specifics of the diversity problem. Finally, the applications group identified the role of multiple data sources when addressing ambiguous queries.

More generally, there was support for the interactive format and the topic of the workshop. With this in mind, we are pleased to announce that the Diversity in Document Retrieval workshop series will continue, with the next edition being held in conjunction with the 5th ACM Web Search and Data Mining conference (WSDM 2012), on February 12th 2012 in Seattle WA, USA. Further information about DDR 2012 can be found at `http://www.dcs.gla.ac.uk/workshops/ddr2012/`. Finally, directly relevant to the topic of this workshop is a special issue of the Information Retrieval Journal on Search Intents and Diversification, which is calling for submissions before April 2nd 2012.

Acknowledgements

We would like to thank ECIR 2011 for hosting this workshop, and in particular Leif Azzopardi and Cathal Gurrin. We would also like to thank the program committee named above in Section 2. Finally thanks are due to the paper authors, the invited speakers Tetsuya Sakai and Alessandro Moschitti, and the participants for a great and lively workshop. Details about the workshop including the contributed papers are available online at `http://www.dcs.gla.ac.uk/workshops/ddr2011`.

References

[1] R. Agrawal, S. Gollapudi, A. Halverson, and S. Ieong. Diversifying Search Results. In *Proceedings of WSDM 2009*.

[2] D. Buscaldi and P. Rosso. Explicit Query Diversification for Geographical Information Retrieval. In [10].

[3] G. Capannini, F.M. Nardini, R. Perego and F. Silvestri. A Search Architecture Enabling Efficient Diversification of Search Results. In [10].

[4] J. Carbonell and J. Goldstein. The use of MMR, Diversity-based Reranking for Reordering Documents and Producing Summaries. In *Proceedings of SIGIR 1998*.

[5] P. Castells, S. Vargas and J. Wang. Novelty and Diversity Metrics for Recommender Systems: Choice, Discovery and Relevance. In [10].

[6] P. Chandar and B. Carterette. Analysis of Various Evaluation Measures for Diversity. In [10].

[7] C. L. A. Clarke, N. Craswell, and I. Soboroff. Overview of the TREC-2009 Web track. In *Proceedings of TREC 2009*.

[8] C. L. A. Clarke, M. Kolla, G. V. Cormack, O. Vechtomova, A. Ashkan, S. Büttcher, and I. MacKinnon. Novelty and Diversity in Information Retrieval Evaluation. In *Proceedings of SIGIR 2008*.

[9] C. Macdonald, I. Ounis, and I. Soboroff. Overview of the TREC-2009 Blog track. In *Proceedings of TREC 2009*.

[10] C. Macdonald, J. Wang and C. Clarke (Editors). Proceedings of the 1st International Workshop on Diversity in Document Retrieval (DDR 2011). Dublin, Ireland, 2011. `http://www.dcs.gla.ac.uk/workshops/ddr2011/ddr2011.proceedings.pdf`

[11] A. Moschitti. Analysis of Document Diversity through Sentence-Level Opinion and Relation Extraction. Invited talk in [10].

[12] V. Plachouras. Diversity in Expert Search. In [10].

[13] N. Rooney, D. Patterson and V. Dobrynin. SOPHIA: bridging the gap between thematic modelling to interactive diverse search. In [10].

[14] T. Sakai. Challenges in Diversity Evaluation. Invited talk in [10].

[15] R. L. T. Santos, C. Macdonald, and I. Ounis. Exploiting query reformulations for Web search result diversification. In *Proceedings of WWW 2010*.

[16] R. Santos and I. Ounis. Diversifying for Multiple Information Needs. In [10].

[17] S. Schockaert and M. De Cock. Diversification of Search Results as a Fuzzy Satisfiability Problem. In [10].

[18] M. Sydow. Towards the Foundations of Diversity-Aware Node summarization on Knowledge Graphs. In [10].

[19] J. Wang and J. Zhu. Portfolio Theory of Information Retrieval. In *Proceedings of SIGIR 2009*.

[20] W. Zheng and H. Fang. A Comparative Study of Search Result Diversification Methods. In [10].