My research interests lie in *information retrieval, sentence (and document) retrieval, text mining and computational linguistics*. In particular, my current research is focused on opinion mining in social media. I am a 2nd year Ph.D. student in the University of Santiago de Compostela (Spain) and I am working on a sentence-level approach to estimate the sequence of sentiments in textual documents. Given this estimation, I expect to detect useful patterns related to the way in which people express opinions in texts. Recently, I have made important contributions in this area that have been published in the ACM CIKM 2011 as a full paper (15% acceptance rate) [2].

Since 2009, I have been working as a hired researcher in three R&D projects on Information Retrieval and related technologies. I have gained indispensable knowledge and experience in building software prototypes, experimentally validating IR systems, and applying different programming languages (java, C++, perl, bash, etc.) to test my research ideas. Some of the prototypes in which I have been working can be found at `http://gsi.dec.usc.es/ir/findapps`. During this time I also participated in two technology transfer projects where I could improve my skills about web design & commercial IR technologies, such as Tapestry or Lucene.

Research Overview

In recent years, several studies have been conducted to determine opinions in social media (e.g. from blog posts). Most of these address this challenge as a two-stage process that involves a topic retrieval stage (i.e. retrieve on-topic documents given a user query), and a re-ranking stage that takes into account opinion-based features. This second Opinion Mining (OM) stage usually involves two further subtasks: a) an opinion-finding task, where the main aim is to find opinionated documents related to the query, and b) a subsequent polarity task to identify the orientation of an opinionated document with respect to the topic (e.g. positive, negative or neutral). In the last two years I have been working in both tasks.

- Topic Retrieval. I have proposed some adjustments to effective blog retrieval methods based on the distribution of sentence scores in the documents. I was able to successfully identify truly relevant documents by combining score features from document and sentences. This helps to detect right contexts related to the query. Some of my proposed variants, published in the 1st Spanish Conference on Information Retrieval as a full paper (46% acceptance rate) [1], can outperform state-of-the-art blog topic retrieval models.

- Opinion Mining. I have been working on polarity estimation, a challenging area that is much more demanding than topicality estimation. Usually, polarity estimation is severely affected by parts of the text that are off-topic or that simply do not express any opinion. Furthermore, there might be conflicting opinions in a document. For example, a blog writer may summarise pros and cons of a particular argument before settling on an overall recommendation. This mixed set of opinions severely affects the quality of automatic methods designed to estimate the overall orientation of the blog post. To overcome these issues I have explored how to combine topic-based, location-based and polarity-based information. More specifically, I have demonstrated that it is possible to successfully determine the polarity of documents guided by a sentence-level analysis that takes into account the topicality

and the location in the blog post of the subjective sentences. The large-scale experimental validation that I conducted showed that some of the methods proposed are both highly effective and computationally-lightweight [2].

Research Goals

Building upon what I have already done so far, I intend to work on challenges that lie ahead. In particular, I have the following immediate research plan:

- Sentiment Flow. As argued above, I have been trying to identify the key evaluative sentences in blog posts. A potential extension of this work is to define models of sentiment flow that effectively capture the sequence of opinions that are expressed in a typical social media document. Given these models, my aim is to mine patterns related to the way in which people express opinions in texts. For instance, a single positive sentiment may completely change the orientation of a document that is populated by negative sentiments. These patterns could help to understand the overall orientation of the text and, therefore, help to accurately estimate the polarity of the documents.

References

[1] Jose M. Chenlo and David E. Losada. Combining Document and Sentence Scores for Blog Topic Retrieval. In *1st Spanish Conference on Information Retrieval , CERI 2010, Madrid (Spain)*, 2010 (46% acceptance rate).

[2] Jose M. Chenlo and David E. Losada. Effective and Efficient Polarity Estimation in Blogs based on Sentence-Level Evidence. In *20th ACM Conference on Information and Knowledge Management,CIKM 2011, Glasgow (Scotland)*, pages 365–374, 2011 (15% acceptance rate).

josemanuel.gonzalez@usc.es