

Handling Shape Variations in Geographical Distance Rankings For Local News Recommendation

Jon Espen Ingvaldsen
NTNU – Norwegian University of Science and
Technology
Trondheim, Norway
jon.espen.ingvaldsen@ntnu.no

Dirk Ahlers
NTNU – Norwegian University of Science and
Technology
Trondheim, Norway
dirk.ahlers@idi.ntnu.no

ABSTRACT

Geographical proximity is an important ranking feature in many context aware recommendation systems. News recommender systems are an example of such systems where the proximity between users and the news' geographical context is a particularly important relevance factor. In this paper, we will discuss ways to improve distance ranking by taking geographical entity sizes, shapes, and footprints into account when handling news items that are ranked for specific users.

CCS Concepts

•Information systems → Spatial-temporal systems; Recommender systems;

Keywords

Recommender System, News Retrieval, Distance Ranking, Footprints, Mobile, Geographic Information Retrieval, Knowledge Base

1. INTRODUCTION

Locations and place names are found everywhere. News is important for many people, and news articles usually report about occurrences connected to one or multiple locations. While place names in news are often ambiguous, they can mostly be disambiguated and placed on a map.

This paper addresses a complementary issue, how to handle news articles with location references and how to select and *rank* the best results to a user in a recommender system. News articles can contain multiple places and at different scales. The challenge is how to consider an article that for example talks about Oslo, Scandinavia, and the North Sea and how to compare it to an article referencing for example Trondheim, Norway, Stockholm. Such places have different spatial granularity. For example, for a person living in Trondheim, local news about the direct place would be interesting, general news from the capital Oslo a little less, but information about the whole of Norway may again

be ranked a bit higher, while also strongly depending on topical preferences. Our approach for geospatial relevance considers not only distance, but characteristics that are variously named as size, extent, footprint, coverage, or shape of places. Our main contribution thus is the improvement of distance ranking metrics with geographic shapes for news recommendation.

By indexing geographical footprints together with point-based geometries, we can extract media content relevant for the neighborhood of geographical reference locations and also rank them according to geographical properties [5]. Ranking items according to their distance to a reference point is an important feature to recommend news relevant to the geographical context and neighborhood of a given user.

A major difference between IR and RecSys is that in IR systems, we have a dedicated query. We may still have to calculate complex shape relationships (for example, oil platforms near the UK). In the recommender system, there is no explicit query. Instead, we have to select and rank items based on a general user profile, context, and location. The additional challenge for the news recommendation is that people may be interested in multiple results with overlapping scopes. Although recommender systems have no explicit query, they can be implemented using a query driver IR system. In such recommender system designs, the query is typically consisting of multiple factors, like users context, intents and interest profile.

Factors such as time, user interest similarities, and entity salience can describe news relevance. However, in this paper, we solely focus of the geospatial factor, and discuss challenges on how to measure the relationship between users' and news articles' geospatial context.

2. INCORPORATING AREA IN DISTANCE RANKINGS

Figure 1 shows an example of geographical entities with various shapes and areas and a user reference point. The map shows entities such as Melhus (municipality), Trondheim (city), Sør-Trøndelag (county), Norway(country) and Scandinavia (including Norway, Sweden and Denmark). Melhus and Trondheim are represented as overlapping circles, while the other location entities are represented as polygons. All of the location entities are commonly described in news articles, and their geospatial properties can be utilized in news recommender systems to derive geospatial relevance. The user reference point is within the intersection of Melhus and Trondheim, which would mean that the user is located in the outskirts of Trondheim, which merges into Melhus.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

GIR '15, November 26-27, 2015, Paris, France

© 2015 ACM. ISBN 978-1-4503-3937-7/15/11...\$15.00

DOI: <http://dx.doi.org/10.1145/2837689.2837702>



Figure 1: Illustration of geographical entities with different shapes and areas.

One approach for retrieving and recommending geo-relevant items is to apply “within” or “overlap” filters. The query is expressed as a shape (for a query usually a circle) and all items within or overlapping it are returned. This approach utilizes the shape property information, but it suffers from being very formal and strict on shape border definitions and it is not able to provide a result set ranking. By looking at the location centers of each geographical entity, we could easily calculate the distance and rank the entities, represented as points. Although the approach allows us to rank items, it would also mean that any user who is located close to the center of a larger geographical entity (like a country) would receive a lot of country-level-relevant news, and hardly any news from the local surroundings.

Neither of these approaches hold for news. People are interested in news at different scales and a news recommender system should ideally recommend news at different geospatial scale levels simultaneously. We propose to handle shape and scale variations in geographical distance rankings for local news recommendation by combining two decay functions. One is targeting the distance between a user’s reference point and the center of the location entities, and one is targeting the area of the location entities. Decay functions are commonly described as linear, exponential, or Gaussian functions. The distance function can be handled in the usual way. But we can handle the second one with an adapted function defined not on distance, but on area. For this, area values of entities are calculated with an exponential decay function taking origin, offset, scale and decay factor as arguments. An exponential decay function with origin = 0, offset = 5 km², scale = 10 km² and decay = 0.5 as arguments would give all entities with an area of less than the offset range a weight of 1. All entities with an area larger than 5 km² will get a lower score affected by an exponentially falling curve. This curve gives a decay weight of 0.5 at an entity area of 10 km². In short, size of entities matters; the area value allows to handle overlapping location footprints by specificity. As news often refer to multiple location entities, we need to specify a strategy for how to handle multiple values. Examples of such strategies are AVG, MIN, MAX

and SUM. The advantage of using decay functions is that several can be combined, each representing one relevance factor. Examples of other news recommender factors that can be implemented as decay functions include user interest similarity, popularity and time.

In the example, the proposed approach would mean that the location factor will have most impact for news articles describing Melhus and Trondheim. Even though the location center of Sør-Trøndelag is very close the user reference point, this entity is given less relevance for the local scale.

In related work, [1] uses several parallel place hierarchies to disambiguate list of places. We invert this approach to recommendations for multiple granularity levels [2]. Addressing similar issues of ranking for search, [4] explore combinations of similarity metrics. They use geographical scope of query and documents, a hierarchical distance over the taxonomy/gazetteer, area and degree of overlap of scopes, distance (and normalised distance) between centroid points of query and documents, and some more. This is most similar to our work. Our approach differs in that it ranks news for recommendation, that news may contain multiple distanced places, and that we need to work at multiple granularities.

3. CONCLUSION AND FUTURE WORK

We have discussed recommendation of news articles based on the similarity between the geographical context of a given user and the geographical entities mentioned in the articles. We have proposed a new approach that incorporates both distance and area properties as well as area size relations into an overall geo distance ranking. The proposed approach is implemented in a mobile news recommender system developed as a part of the Smartmedia project [3]. Experiences from this project show that the approach of decaying news according to distance and area has both benefits and disadvantages. Because news articles frequently refer to multiple geographical entities, we have to finetune how to combine decay results on the news article level. Area and distance to centroid work well on sets of entities with compact shape. However, for geographical entities such as roads or rivers, or vastly distanced sets (such as international capitals), there are limitations, which we need to address together with detailed user evaluations and possibly learning to rank approaches [4]. Handling a wider variety of geographical shapes and their influence to news recommendation as well as the dependencies between spatial ranking and news rankings will be addressed in future work.

4. REFERENCES

- [1] M. D. Adelfio and H. Samet. Structured toponym resolution using combined hierarchical place categories. In *GIR '13*, 2013.
- [2] D. Ahlers. Granularity as a Qualitative Concept for GIR. In *GIR'15*. ACM, 2015.
- [3] J. E. Ingvaldsen and J. A. Gulla. Taming news streams with linked data. In *RCIS 2015*. IEEE, 2015.
- [4] B. Martins and P. Calado. Learning to Rank for Geographic Information Retrieval. In *GIR '10*, 2010.
- [5] H. Samet, J. Sankaranarayanan, M. D. Lieberman, M. D. Adelfio, B. C. Fruin, J. M. Lotkowski, D. Panozzo, J. Sperling, and B. E. Teitler. Reading news with maps by exploiting spatial synonyms. *Communications of the ACM*, 57(10):64–77, 2014.