



COMMIT/

The Role of Geographic Information in News Consumption

Gebrekestos G. Gebremeskel and Arjen P. de Vries
gebre@cwi.nl

LocWeb2015, Florence, Italy

- Does geographic proximity play a role in news consumption?
- At what level?
 - At portal (publisher) level?
 - Local category level?

Dataset

- Data collected from Plista during our participation in CLEF NEWSREEL: Benchmark News Recommendations in a Living Lab
 - Contains one month's impressions
 - 53 million impressions (item viewings by users)

Information Portals

URL	Type	Short Name
cfoworld.de	Business	Cfo
cio.de	IT news	Cio
computerwoche.de	IT news	woche
gulli.com	IT & Games	Gulli
ksta.de	News	ksta
motot-talk.de	Automotive	M-talk
tecchannel.de	IT	Channel
sport1.de	Sports	Sport1
tagesspiegel.de	News	Tage
wohnen-und-garten.de	Garden	WH

Two Types of Information Portals

- 10 information portals
 - 8 Special purpose portals (sports, IT and games, Automotive, business, gardening)
 - 2 Traditional news portals (providing politics, opinion, and current events)

Machen Sie tagesspiegel.de zu Ihrer Startseite [X]




Deutsche Bahn und GdL


Kann der Bahnstreik doch noch verhindert werden?

Noch hat der Ausstand bei der Bahn nicht begonnen, da überrascht GdL-Chef Wesel mit Gedanken zu einer möglichen Schlichtung. Bisher lehnte er diese Möglichkeit aber strikt ab. *Von Simon Frost, Heike Jahberg mehr... [7 Kommentare]*

- Anzeige -



Visit Greece
Have you ever been to the majestic is...

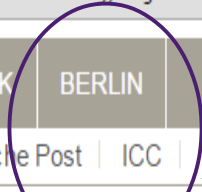


+938 ↻ 133 45 [Hover to Expand](#)

EL CHECKPOINT

Der **Berlin-Newsletter** von
Tagesspiegel-Chefredakteur **Lorenz**

 **NACH OBEN**



Local news category



Deutsche Bahn und GdL

Kann der Bahnstreik doch noch verhindert werden?

Noch hat der Ausstand bei der Bahn nicht begonnen, da überrascht GdL-Chef Wesel mit Gedanken zu einer möglichen Schlichtung. Bisher lehnte er diese Möglichkeit aber strikt ab. *Von Simon Frost, Heike Jahberg mehr... [7 Kommentare]*

- Anzeige -

Visit Greece
Have you ever been to the majestic is...

+938 133 45

Hover to Expand

Kann der Bahnstreik doch noch verhindert werden? EL CHECKPOINT

Der Berlin-Newsletter von
Tagesspiegel-Chefredakteur Lorenz

NACH OBEN

Item and User Geographic Information

Item's geographic Information

- Publisher
 - Are some portals more related to some regions?
- Local category
 - Within traditional news portals, are local news categories more appealing to users from some geographic regions?

User's geographic Information



- User's state-level postcode
- 52 states of
 - Germany
 - Austria, and
 - Switzerland

User's Geographic Information



Correlations?



- Portals
 - › Tagesspiegel
 - › Ksta
 - › Sport1
 - › ..
 - › ..
- Categories
 - › Local news
 - › Non-local news

Method

- Compute geographic likelihood distribution
 - $P(\text{Portal}|\text{user's state})$, and $P(\text{category}|\text{user's state})$
- Compute Jensen-Shannon distance (JSD) score based on the geographic likelihood distribution
 - Jensen-Shannon is a symmetric version of KL-Divergence
 - Its square root is true distance metric, called JSD
 - A higher JSD score, a more different geographic user distributions

Results

Distance Scores between portals

	WH	M-Talk	Tage	Woche	Cio	Cfo	Chanel	Ksta	Sport1
Gulli	0.067	0.057	0.187	0.066	0.101	0.129	0.043	0.322	0.102
Sport1	0.099	0.080	0.192	0.091	0.105	0.131	0.119	0.305	
Ksta	0.330	0.314	0.368	0.323	0.321	0.332	0.331		
Chanel	0.067	0.062	0.209	0.055	0.087	0.11			
Cfo	0.140	0.127	0.229	0.082	0.053				
Cio	0.110	0.093	0.215	0.044					
Woche	0.076	0.060	0.198						
Tage	0.221	0.210							
M-talk	0.033								

The highest distance is between Tagespiegel and Ksta, the two traditional news portals

Distance Scores between portals

	WH	M-Talk	Tage	Woche	Cio	Cfo	ChaneKsta	Sport1	
Gulli	0.067	0.057	0.187	0.066	0.101	0.129	0.043	0.322	0.102
Sport1	0.099	0.080	0.192	0.091	0.105	0.131	0.119	0.305	
Ksta	0.330	0.314	0.368	0.323	0.321	0.332	0.331		
Chanel	0.067	0.062	0.209	0.055	0.087	0.11			
Cfo	0.140	0.127	0.229	0.082	0.053				
Cio	0.110	0.093	0.215	0.044					
Woche	0.076	0.060	0.198						
Tage	0.221	0.210							
M-talk	0.033								

Each portal's highest distance score is from Ksta

Distance Scores between portals

	WH	M-Talk	Tage	Woch	Cio	Cfo	Chane	Ksta	Sport1
Gulli	0.067	0.057	0.187	0.066	0.101	0.129	0.043	0.322	0.102
Sport1	0.099	0.080	0.192	0.091	0.105	0.131	0.119	0.305	
Ksta	0.330	0.314	0.368	0.323	0.321	0.332	0.331		
Chan	0.067	0.062	0.209	0.055	0.087	0.11			
Cfo	0.140	0.127	0.229	0.082	0.053				
Cio	0.110	0.093	0.215	0.044					
Woch	0.076	0.060	0.198						
Tage	0.221	0.210							
M-talk	0.033								

Each portal's second highest distance score is from Tagesspiegel

- The highest score between the traditional news portals indicates that the two portals differ the most in their geographic readerships
- Their big distance scores from the special portals indicates that the two traditional news portals have different geographic readerships from the special portals.
 - Geography plays a role in their readership
- Thus we focus on the traditional news portals and examine if the geographic information also manifests at local categories level

Local vs. Non-local Categories

- We extracted two categories for each traditional portal
 - Tagesspiegel: Berlin (Tage+Ber) and Non-Berlin (Tage-Ber)
 - Ksta: Cologne (Ksta+Col) and Non-Cologne (Ksta-Col)
- For comparison, we also included a sport category for Tagesspiegel (Tage+Sport)

Local vs. Non-local categories

	Tage	Ksta	Tage+Ber	Ksta+Col	Ksta-Col	Tage-Ber
Tage+Sport	0.038	0.360	0.207	0.465	0.358	0.046
Tage-Ber	0.031	0.354	0.230	0.465	0.351	
Ksta-Col	0.366	0.003	0.483	0.133		
Ksta+Col	0.474	0.130	0.561			
Tage+Ber	0.200	0.485				
Ksta	0.368					

The highest distance is between Berlin and Cologne, followed by between Berlin and Ksta

Local vs. Non-local categories

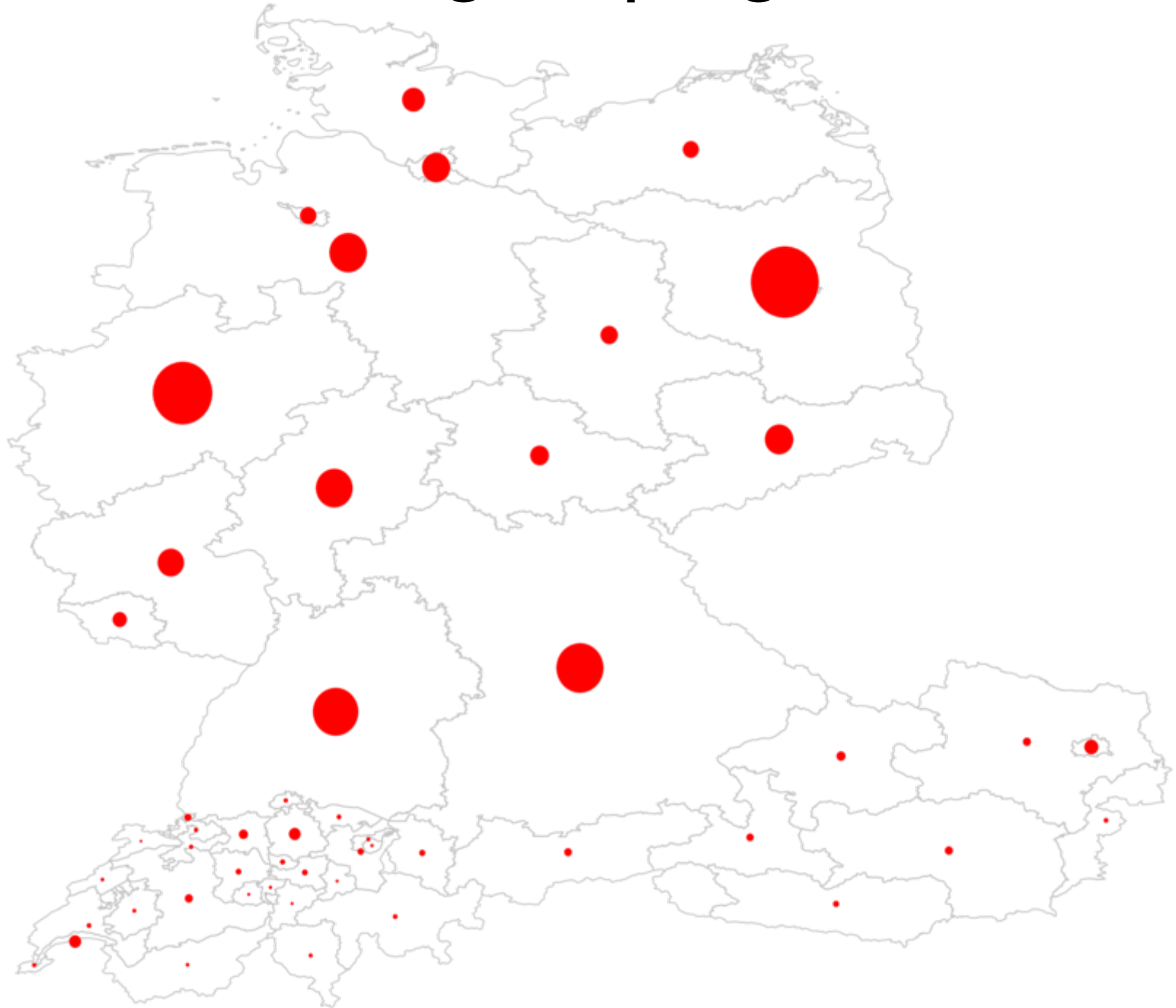
	Tage	Ksta	Tage+Ber	Ksta+Col	Ksta-Col	Tage-Ber
Tage+Sport	0.038	0.360	0.207	0.465	0.358	0.046
Tage-Ber	0.031	0.354	0.230	0.465	0.351	
Ksta-Col	0.366	0.003	0.483	0.133		
Ksta+Col	0.474	0.130	0.561			
Tage+Ber	0.200	0.485				
Ksta	0.368					

- More interesting is the distance scores between categories in the same portal.
 - Tagesspiegel's Berlin with Tagesspiegel's non-Berlin (compare with Tagesspiegel Sport)
 - Ksta's Cologne with Ksta's non-Cologne

Local Vs. Non-local Categories

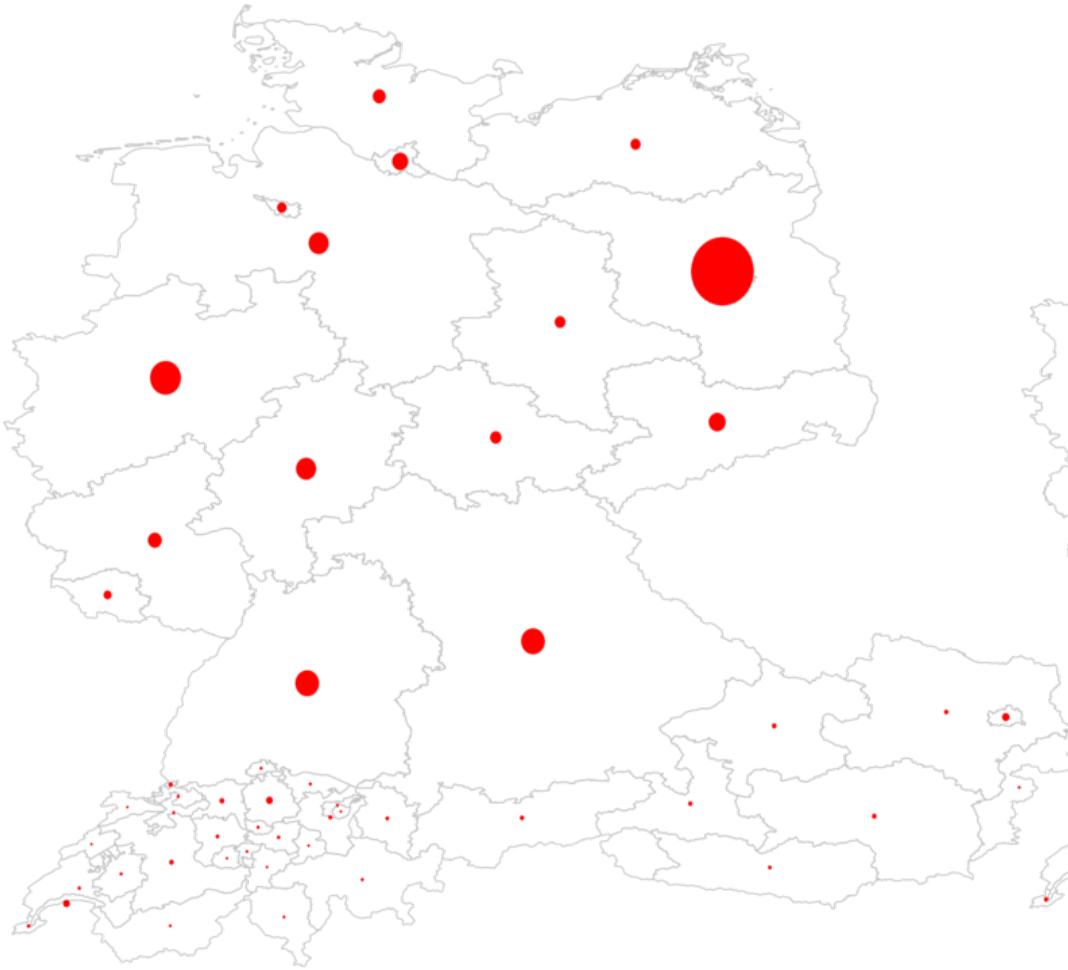
- The local categories have distinct geographical distributions of readership different from their non-local categories
- Tagesspiegel's local category has a more geographically distinct readership from Tagesspiegel's non-Berlin than Ksta's local category from Ksta's non-local
 - Tagesspiegel's national nature, and Ksta's regional character may explain this.

Tagesspiegel

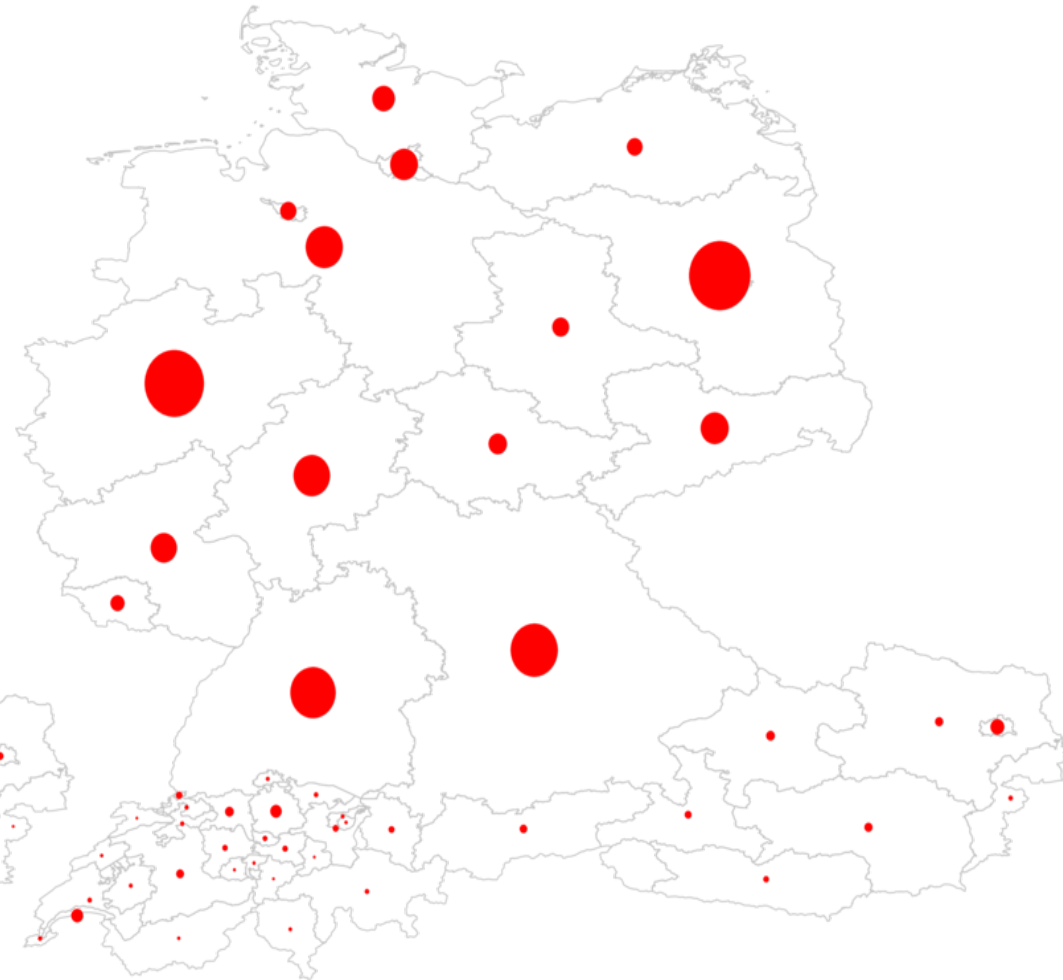


Tagesspiegel's Berlin vs Non-Berlin

Berlin



Non-Berlin



Conclusion

- Geographical information as represented by user's state-level postcodes for users, and portals (and local categories) for items plays a role in news consumption of traditional news portals at two levels
 - At the portal level: user's seem to ascribe geographical focus to traditional news portals
 - At local category level: local news categories attract a more geographically proximate users to themselves
- Might be useful to incorporate in news recommendation

Preview of Results of Geographic Information in Live Recommendation

- We Incorporated geographic information into recency in live recommendation systems in Plista
 - Recency is a recommendation system that recommends the most recently viewed items to the user
 - For Tagesspiegel and Ksta, a geographic recommender system generates geographical recommendations which are then intersected with recency recommendation

A preview of geographical Information in News Recommendation

- We incorporated the geographical factor in a news recommender system.
- Experimented with two instances of the same algorithm (recency, and recency2), a geographical recommender (GeoRec) and a random recommender (Random)

Results

	Requests	Clicks	CTR(%)
Recency	37,520	296	0.79
GeoRec	35,789	310	0.87
Random	23,232	149	0.64

- The GeoRec seems to do better.
- But, is it an improvement?

Results

	Requests	Clicks	CTR(%)
Recency	37,520	296	0.79
Recency2	35,668	255	0.71
GeoRec	35,789	310	0.87
Random	23,232	149	0.64

- Recency and Recency2 have different performances.
- What explains this?

Open Questions

- What would be better ways of incorporating the geographic information into live recommendation?
 - Specifically to recency recommender so that we have a spatio-temporal recommender system?
- What is the time needed to compare two algorithms online?
- What does the difference in performance of the same recommender system signify?