“Where Far Can Be Close”: Finding Distant Neighbors In Recommender Systems

Vikas Kumar Daniel Jarratt Rahul Anand Joseph A. Konstan Brent Hecht
GroupLens Research
Dept. of Computer Science
University of Minnesota, Twin Cities
Minneapolis,USA
{vikas, jarratt, anand037, konstan, bhecht}@cs.umn.edu

ABSTRACT

Location and its corollary, distance, are critical concepts in social computing. Recommender systems that incorporate location have generally assumed that the utility of location-awareness monotonically decreases as entities get farther apart. However, it is well known in geography that places that are distant “as the crow flies” can be more similar and connected than nearby places (e.g., by demographics, experiences, or socioeconomic). We adopt theory and statistical methods from geography to demonstrate that a more nuanced consideration of distance in which “far can be close” – that is, grouping users with their “*distant neighbors*” – moderately improves both traditional and location-aware recommender systems. We show that the distant neighbors approach leads to small improvements in predictive accuracy and recommender utility of an item-item recommender compared to a “nearby neighbors” approach as well as other baselines. We also highlight an increase in recommender utility for new users with the use of distant neighbors compared to other traditional approaches.

Keywords

location-aware recommendations, user clustering, distant neighbors

1. INTRODUCTION

Collaborative filtering starts with the assumption that past agreement is predictive of future agreement. Separating recommenders into user and item domains, within which users’ agreement is more predictive, increases the extent to which that core assumption is true. For this reason, we do not build recommender systems that mix movies, baseball cards, and sweaters into a single model. Konstan et al. [11] give empirical evidence that partitioning user-user recommenders by item domain improves predictive accuracy, showing that recommending within Usenix newsgroups rather than across them produces the lowest error. Similarly, partitioning users has resulted in better predictive accuracy than the full model [18, 25].

Location-aware recommenders, for example, have used geographic information as a filter to narrow the user-item rating space for improved performance. LARS [14] partitions users into geographic grid squares: an assumption that users within a contiguous, compact grid square are more alike than users not in that square. Other location-aware recommenders [28, 1] assume similarity is proportional to straight-line (geodesic) distance.

This paper extends that approach by recognizing that location-aware recommenders can incorporate non-proximate – distant, non-contiguous, and non-compact – geographies into their recommendation models. Current location-aware recommenders are grounded in Tobler’s First Law of Geography [22] (TFL): “*everything is related to everything else, but near things are more related than distant things*”. However, Tobler emphasizes that it is a rule-of-thumb rather than a law, and urges researchers to consider more sophisticated notions of distance (e.g., population density-controlled distance, border-aware distance, socioeconomic distance). Recent work in peer production communities finds evidence for different ways of calculating distance in regards to the First Law [7]. We do this by considering “ratings preference distance” between locations, combined with geodesic distance.

We use this non-trivial notion of distance to determine similar locations (specifically postal codes: e.g., 19104 in Philadelphia, Pennsylvania, is most similar to 61801 in Urbana, Illinois) and partition users based on their associated location, then build collaborative filtering models for each partition. We demonstrate that general collaborative filtering systems – even those that consider items that do not have a unique spatial footprint, such as movies – can benefit from an understanding of user geography, and specifically an understanding that goes beyond a “closer = more relevant” assumption. We provide the first evidence that similarity of locations (based on ratings) does not monotonically decrease with geodesic distance, meaning that locations often have “**distant neighbors**”. We demonstrate that creating a user space based on location similarity moderately improves predictive accuracy and recommender utility, relative to baseline approaches and location-aware recommender systems that rely just on proximate users.

At a high level, this paper demonstrates that we can harness geographic techniques to improve rec-

LocalRec’15, September 19, 2015, Vienna, Austria.

ommender systems by limiting the recommender model to groups of users in similar, though not necessarily adjacent, locations. Intuitively, the argument we make here for partitioning recommender systems based on users rests on whether one can find sets of users who have significantly different "views of the world" that make past agreement less predictive of future agreement. We can think of this in the context of item-item collaborative filtering, where we build a model of item-similarity (rating correlations) of item association (item co-rating or co-purchase). If items are considered related by one set of users, but not by others (e.g., if Belgians see french fries associated with mayonnaise while Americans see them associated with ketchup but not mayonnaise), then an item model built across these diverse groups of users may be less effective at recommendation than one based on partitions of users. Based on this intuition, we look at partitions based on location similarity by forming groups of distant neighbors, and make the following contributions (compared to an item-item recommender that uses no partitioning, a geographic partitioning based only on local neighbors, and a random-partitioning baseline):

1. We show improvement in prediction accuracy using partitioning on similar locations ("distant neighbors").
2. In new user cold start, we show improvement in recommender utility for Top-N lists.

2. RELATED WORK

This paper is motivated by research in two disciplines: geography and recommender systems. Below we outline related work in each of these areas.

2.1 Tobler's First Law

Current use of geography in collaborative filtering is almost always under the assumption that the most valuable ratings come from nearby people – an assumption grounded in the First Law of Geography (or "Tobler's First Law of Geography", 1970) [22]. In other words, the recommender systems community has used distance decay as the relevant property of location. But as noted by Tobler in 2004 [21], deviations from pure distance decay, such as population density, are common. Tobler's 2004 assertion has received empirical support in the work of Li et al. [16] and Hecht and Moxley [7], which found that the relatedness between geographic entities in Wikipedia generally decreases with geodesic distance, but that (as advocated by Tobler himself) we need to consider more sophisticated notions of distance to more completely model the relationship between relatedness and distance. Our distant neighbors approach uses "ratings preference distance" in combination with geodesic distance: we show evidence that people do have similarity with people around them, but that at a certain geographic scale (for which we examine postal codes), they are just as similar to a subset of distant people. This paper is the first, to our knowledge, to show that an understanding of TFL that combines ratings preference similarity and geodesic distance can lead to increased accuracy in recommendations.

2.2 Location-aware recommendation

Location-aware recommender systems consider location in two ways: (1) "nearby neighbors" filtering systems that explicitly or implicitly encode a strict form of TFL and (2) "just another feature" recommenders that treat location as profile features but do not apply geographic methods.

2.2.1 "Nearby neighbors" filtering systems

When recommenders are aware of location (i.e., from mining GPS tracks [15] or asking a person for her postal code [20]), the systems use that context to personalize suggestions to those relevant to nearby users or items [1]. Some location-aware recommenders [9, 27] operate within domains of items which have specific spatial footprints [6] (such as restaurants, people, or homes). For instance, recommenders meant for mobile queries can calculate an area around a person where her query remains valid [14][28] if, for instance, she requires updated coffee shop recommendations as she moves from neighborhood to neighborhood. Some point-of-interest recommenders [3] use co-visitation rates (implicitly geographic).

Other recommenders operate in domains of items which do not have specific spatial footprints, such as movies or consumer products. Levandoski et al. [14] tessellate users into geographic grid cells, which is discussed in detail below. Similarly, Das et al. [5] apply Voronoi tessellation, a strictly "nearby neighbors" method.

The primary motivation for this work is that of Levandoski et al. [14], who introduce the "Location-Aware Recommender System" (LARS). Like our work, LARS is a partitioning process that occurs before collaborative filtering model-building. LARS lays a hierarchy of grids over the surface of the Earth at various levels of granularity, so that each user belongs to one geographically contiguous and compact square. LARS then creates an item-item recommender model for each grid cell, suggesting movies – items without specific spatial footprints – based on the ratings of other people in the same grid cell. Its authors find an inflection point where quality of recommendations is maximized: a 4x4 grid where users are partitioned into 16 geographic squares. Smaller squares have too little data; larger squares include too many less-similar users. LARS thereby encodes physical proximity as a proxy for similarity, and is the first paper to contribute geographic user partitioning for collaborative filtering. Our contribution contra LARS is that geographic user partitions need not be contiguous nor a compact space.

2.2.2 Location as "just another feature"

Still other work in general recommender research treats location information as "just another feature". Seroussi et al. [20] use standard matrix factorization techniques to find latent relevant features in a person's preferences, with U.S. state names drawn from postal codes as possible features. Similarly, Kucuktunc et al. [12] match postal codes with census data, inferring demographics to match users in a question-and-answer site. We see promise in Kucuktunc's method, which uses postal codes as a way to guess demographic tags for users. Their method is based on Weber and Castillo's demographic analysis which is meant to "depersonalize and de-localize" location information [24], yet demographics are themselves geographically clustered [4], resulting in a "distant neighbors" type of approach through a demographic step. This approach has the benefit of accounting for locations a recommender has never before encountered – "location cold start" – since demographics are often widely available.[1]

[1] www.ipums.org

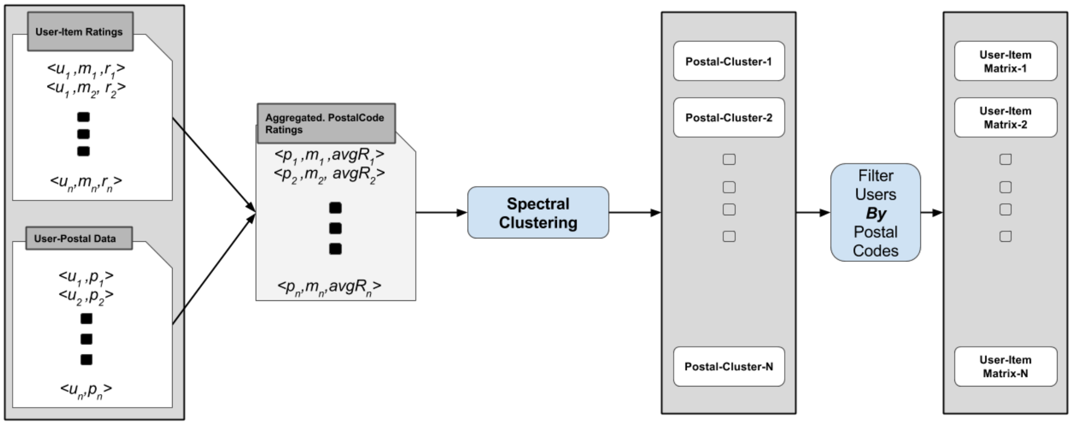

Figure 1: Illustration of Distant Neighbors Partitioning Approach

3. DATASET

We use movie ratings from the MovieLens rating community[2], whose users have the option of entering their postal code while signing up for the site. In MovieLens, users can rate movies on a rating scale from 0.5 to 5.0 in increments of 0.5. For our experimental dataset, we select only those users who entered a valid US postal code (which is about 22% of all users). To ensure data density, we further exclude postal codes that do not have at least 20 MovieLens users who rated at least 20 movies. We finally randomly sample 5000 users to create a test dataset, resulting in 1779 postal codes and 1.01 million ratings.[3] The rating count is equivalent to the extensively-used publicly available MovieLens dataset of 1 million ratings.[4]

4. PARTITIONING APPROACHES

We consider three partitioning approaches – distant neighbors, local neighbors and random – to partition the user item rating matrix.

4.1 Distant neighbors partitioning

Distant neighbors refers to users in a cluster of postal codes based on *rating similarity.* In this section, we explain how we identify distant neighbors from a user-item rating dataset. We then use a traditional collaborative filtering recommender on each of those clusters. Partitioning of the user-item matrix into distant neighbors clusters occurs in three stages:

1. Convert the user-item rating matrix to a location-item rating matrix
2. Determine location-to-location similarity
3. Cluster locations to form groups of users associated with their respective locations

We create a location-item rating matrix from the user-item rating matrix by matching postal codes with their respective user identifiers. We then determine the rating of a postal code for an item by averaging over the location's users' ratings for the same item. Based on previously published work on MovieLens data [10], we use the Bayesian average [26] with a damping value of 5, to dampen the effect of ratings of movies rated by only few users in a postal code.

In the above process, for example, we determine the *Toy Story*-15213 postal code (Pittsburgh, Pennsylvania) rating from the ratings of all users in that postal code who rated *Toy Story.*

We then cluster locations in the location-item rating matrix using spectral clustering.[5] After each process that generates C clusters of postal codes, we partition the original user-item rating matrix into C partitions. Recall that each user belongs to a single postal code and hence only to one cluster. For a cluster count $C = 1$, the user-item rating matrix remains the same as the traditional full model used for recommendation. The process is illustrated in Figure 1.

Why Distant neighbors?

To understand how postal codes reflect rating preference distance compared to geodesic distance, we use geostatistics. The correlogram, a geostatistical tool, helps us understand the correlation between geodesic distance and similarity among places [17]. We show this correlation in Figure 2, where on the x-axis we plot pairwise geodesic distance between postal codes (lag = 100 km) and on the y-axis, the average rating similarity of postal codes (separated by the corresponding distance on the x-axis). To determine similarity we compare the rating vectors for each postal code using (a) cosine similarity, (b) Spearman's rank correlation, and (c) Jaccard index. We also evaluate (d) cosine similarity on postal code's item-popularity vectors instead. As an example in Figure 2, we see that for postal codes separated by 1000-1100 km, their average cosine similarity is approximately 0.28.

Now, going back to the definition of Tobler's First Law based on purely geodesic distance leads to the assumption of the similarity of postal codes monotonically decreasing as the straight-line distance between them increases. However, as shown in Figure 2, it is immediately obvious that Tobler's First Law in its geodesic distance interpretation does not hold in the movie domain. While nearby postal codes

[2] www.movielens.org

[3] This dataset is public at http://cs.umn.edu/∼vikas

[4] www.grouplens.org/datasets/movielens

[5] We find that 20 features and 50 iterations for a given cluster results in better inter-cluster and intra-cluster density.

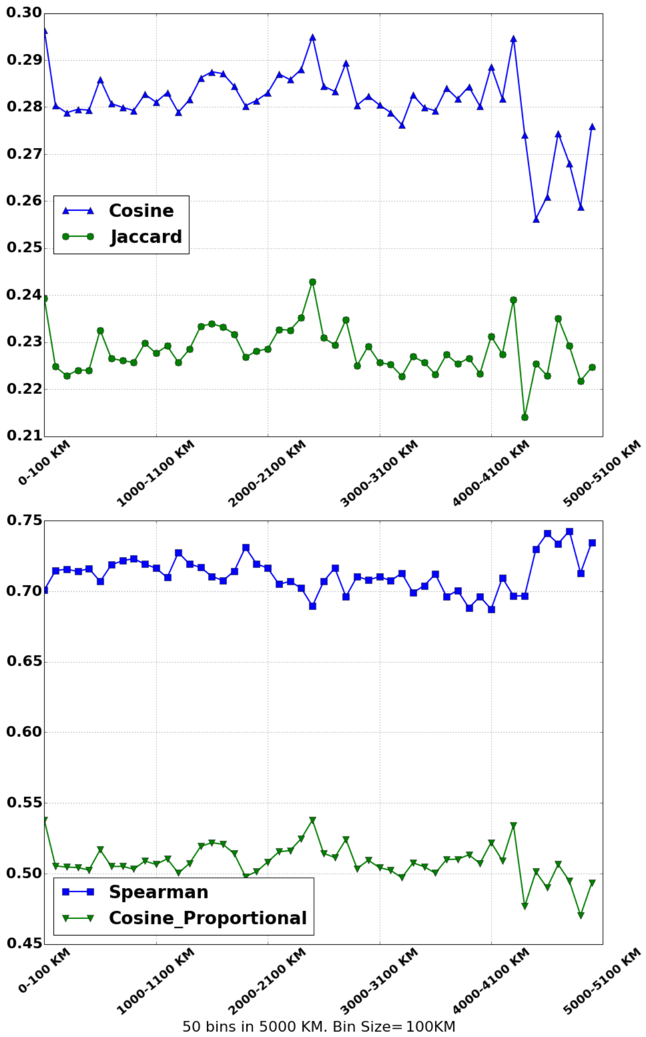

Figure 2: Correlograms: Distance Vs Similarity

are more similar than slightly more distant postal codes, average similarity does not decrease as distance increases. In fact, there are several points where the average similarity of locations separated by thousands of kilometers exceeds the average similarity of nearby locations.

Thus, recommender systems that only consider geodesic distance using (often) the correlogram's first bucket(s) ignore the higher average similarity displayed by distant postal code pairs. If these systems increase their geodesic distance threshold in an attempt to include more users, the newly-added nearby locations may actually have *lower* average similarity than distant locations.

We reasoned that perhaps the nationwide popularity of many movies causes the average similarity of distant postal codes to increase. To test our hypothesis, we re-evaluate the correlograms ignoring the most popular 20% of movies and found the same result.

4.2 Local neighbors

For each postal code, we build a "local neighbors" user-item matrix, which is a subsample of the entire dataset. This consists of the location's own users, plus users from the successively *nearest* locations (geodesic) until a given number of users is reached (matching with the number of users in the corresponding distant neighbors clusters). We use this method to compare our approach to location-aware recommenders that are based on location proximity instead of similarity.

4.3 Random partitioning

As a baseline, we use random partitions where each partition consists of a given number of users (matched with the number of users in the corresponding distant neighbors cluster), randomly selected from the original user-item rating matrix.

5. BUILDING RECOMMENDATIONS

We use the item-item collaborative filtering approach with log-likelihood similarity [19] for the prediction and evaluation of the different partitioning approaches on recommendation. We build models for each partitions (or clusters) independently i.e. if a partitioning approach creates N user-item matrices then we build N recommendation models, one for each matrix. For item recommendations to a user we select her respective partition (determined by her postal code) and use the respective model.

Note, it is possible that with many clusters we may have small user and item proportions per cluster. In such cases, a recommender may fail to predict or recommend resulting in lower number of successful predictions thus affecting the *coverage* of the recommender. For example, item-item collaborative filtering fails to predict rating of an item for a given user, if that item has no correlation with any other item that user has rated. Hence, we compare the accuracy of different partitioning approaches with respect to their coverage. Using fallback predictions, we also compare the accuracy of each partitioning approach at 100% coverage i.e. whenever recommender fails to predict we either use (a) the *item mean rating*, if item exists in the partition's training set, or (b) the *global mean rating* if item does not exist in the partition's training set.

6. EVALUATION

At a high level, our methodology involves three stages. First, we generate user-item rating matrix partitions in several sizes for two approaches, distant neighbors and random. For local neighbors, we build user-item rating matrices in several sizes for each postal code, which we call "partitions" for convenience. Next, for each partition, we build an independent item-item recommender. Finally, we evaluate prediction accuracy using RMSE (root mean squared error) and recommendation accuracy with MAP (mean average precision). We compare the performance of each approach to that of an item-item recommender built using the full rating matrix. We evaluate the techniques on four different numbers of disjoint clusters – 1, 5, 20 and 50 – with user proportion sizes of 100%, 20%, 5% and 2% respectively.

For RMSE evaluation, we use 90% of ratings for each user in the training data and keep 10% of their ratings in the test set. In Top-N evaluation, for each user we build an evaluation set of relevant items [23]: those rated 4 (of 5) or higher, following [10]. We then keep only 80% of that user's rating in the training data with all other users' ratings. We repeat this process for a random sample of 10% of users in the dataset and take the average of MAP for Top-50 recommendations (MAP@50) over this sample of users.

6.1 Results

The distant neighbors approach with 50 clusters shows the best RMSE among all other partitioning approaches and number of clusters, including the full model ("1 partition") recommender, shown in Table 1. We find the improvement against the full model to be statistically significant ($p < 0.001$, per user using the Wilcoxon Rank-Sum test). We also note that for any given cluster count, distant neighbors have more accurate predictions than other partitioning approaches i.e. local neighbors and random partitions; however, we find the difference between them to be statistically significant only for 20 clusters.

The distant neighbors approach's low coverage of 88.75% compared to the full model's 99.87% means that we cannot directly imply better performance. We note that at 20 and 50 partitions, distant neighbors has higher coverage than local neighbors and random partitioning. Controlling for 100% coverage using fallback predictions, we observe that distant neighbors at 20 clusters remains statistically significantly more accurate than full-model item-item ($p < 0.05$), but the other two partitioning methods are no longer significantly different than the full model.

For recommendation metrics, distant neighbors shows significantly better MAP@50 for cluster sizes of 20 and 50, especially over the full model. We hypothesize here that due to very small user proportions (and therefore items) per cluster, the recommendations are limited to a specific set of items liked by most of the users in that cluster. To understand this, we calculate the intra-list similarity (or diversity) between the items, using the full training data, and find a much lower diversity using any partitioning method compared to the full model.

7. NEW USER COLD START EVALUATION

New users are a crucial part of a recommender system's success, and this problem forms an important part of system design decisions. Previous work has used location information to solve the cold start problem [2], because user context data can substitute for user rating data. We analyze the effect of our location-aware approach by simulating the new user experience. Here we randomly sample 10% of users from the training data (that contains 90% of ratings for each user) and retain only a few ratings for those users in the training data – 5 ratings per user – and discard other ratings for that user [10]. We evaluate the metrics on the test data containing only these users.

7.1 Results

In case of cold start with no fallback predictions, we observe distant neighbors with 20 clusters to have the statistically significantly best predictive accuracy against the full model ($p < 0.05$, significance per user). However, we find this difference to be not significant against local neighbors based partitions for the same clusters.

The coverage, like the previous evaluation, remains low for larger number of clusters. With fallback predictions, although distant neighbors has better RMSE compared to local neighbors and random (not significant), no partition method is able to beat the full model RMSE. Moreover, the RMSE gets better with fallback predictions than the RMSE without fallback predictions. This suggests that the item baseline predictor is better than the full model in cold start; we find this result consistent with the results from Kluver et al. [10] on evaluation of item-item for new users.

For recommendation utility, we observe local neighbors to have better MAP@50 for 20 and 50 clusters compared to any other partition or the full model. Distant neighbors is able to do better than other partitions only for 5 clusters. In contrast, with fallback predictions for all number of clusters, we observe that distant neighbors is able to produce better MAP@50 results (0.1972 with 50 clusters) with improvement over local neighbors (0.14755), random (0.09753) and the full model (0.0533).

8. DISCUSSION

In this section, we summarize the results in light of the assumptions and the limitations of our approach. We further highlight the key takeaways and some anecdotes from our dataset.

We show that by controlling for coverage, grouping users based on their location similarity – distant neighbors – shows significant improvement in prediction accuracy over other partitioning approaches for 20 clusters. In cold start, though, we notice that partitioning fails to improve prediction accuracy (RMSE). However, as Kluver et al. [10] states that recommendation utility is more important for new users, we perhaps regard distant neighbors better mean average precision a positive result.

Also, we note that the Top-N metrics, like MAP, have popularity bias [10]. Mean average precision will be higher for any algorithm that favors popular items. To understand if such bias exists in our results, we look at popularity of the recommended items. We find that the local neighbors recommender favors relatively more popular items than distant neighbors, which recommends more popular items than random, and than the full model. We calculate popularity based on the number of users who rated the item in the full matrix. We also determine the average user rating for the items for the cold start situation to understand the relevance of recommendation, by taking the mean of ratings by the user for the relevant items in the recommendation list. We find that distant neighbors consistently provides recommendations that have higher user average ratings: 4.57, compared to local neighbors (4.25) and random (4.23). We therefore find the distant neighbors performance on Top-N metrics to be a positive result.

We consider these results as a proof of concept. We note that the magnitude of the improvements is small, even if statistically significant. We recognize that such small decreases in RMSE are unlikely to, in and of themselves, result in a significant change in user experience. Rather, we hope to see further development and optimization of geographic similarity into recommender systems with the promise of further improvements in performance. We also note that the Top-N results show even more promise, and deserve future user-centered evaluation.

To further interpret our results we revisit our intuition of user worldviews that can manifest in rating space. By "worldviews" we mean the agreement among a subset of users who have significantly different "views of the world" compared to other subsets. We see empirical evidence of how

	No fallback predictions			With fallback predictions (100% coverage)		
Partitions	Distant neighbors	Local neighbors	Random	Distant neighbors	Local neighbors	Random
1	0.98253 RMSE (99.87% coverage) 0 MAP			0.9875 RMSE 0 MAP		
5	**0.98076** (98.81%) 0	0.98206 (98.68%) 0	0.98124 (**99.12%**) 0	0.9822 0	0.98131 **0.02632**	**0.98073** 0
20	**0.96914** **(96.34%)** **0.03846**	0.97346 (95.82%) 0.03015	0.97063 (96.26%) 0.03014	**0.98131** **0.03849**	0.98367 0.02381	0.98234 0.02042
50	**0.96774** **(88.75%)** **0.13854**	0.96946 (87.57%) 0.06105	0.96812 (87.27%) 0.03107	**0.98996** **0.06639**	0.99461 0.0407	0.99373 0.04532

Table 1: Evaluation of partitioning approaches. The best results for each partition count per coverage condition are in bold. The best results in each coverage condition are underlined.

	No fallback predictions			With fallback predictions (100% coverage)		
Partitions	Distant neighbors	Local neighbors	Random	Distant neighbors	Local neighbors	Random
1	1.09396 RMSE (99.87% coverage) 0.0533 MAP			1.07119 RMSE 0.0533 MAP		
5	1.09025 (**99.58%**) **0.05309**	1.08774 (98.85%) 0.03452	**1.08122** (98.61%) 0.0303	1.08078 **0.125**	**1.07376** 0.10346	1.07596 0.09515
20	**1.09** (95.36%) 0.11268	1.09278 (**95.74%**) **0.1241**	1.10356 (94.74%) 0.05956	**1.08069** **0.13668**	1.08502 0.09609	1.08672 0.097
50	**1.10404** (**92.46%**) 0.08049	1.10414 (91.36%) **0.15347**	1.11138 (91.04%) 0.11289	**1.07392** **0.1972**	1.0867 0.14755	1.07834 0.09753

Table 2: Evaluation of partitioning approaches in simulated cold start. The best results for each partition count per coverage condition are in bold. The best results in each coverage condition are underlined.

Place	Most Similar	2nd Most Similar	3rd Most Similar
Jersey City, NJ 07087	San Francisco, CA 94104	Redwood City, CA 94063	Bala Cynwyd, PA 19004
Philadelphia, PA 19104	Urbana, IL 61801	New York, NY 10003	Cambridge, MA 02139
Ann Arbor, MI 48104	Minneapolis, MN 55455	Brooklyn, NY 11215	Minneapolis, MN 55414
Palo Alto, CA 94305	Chicago, IL 60614	Urbana, IL 61801	Ann Arbor, MI 48103

Table 3: Most similar locations for selected US postal codes

college towns are more similar to other college towns than to other postal codes in their own city, and hypothesize that people on a particular college campus may share an understanding of concepts (for example: registration, fraternities, and late-night pizza) with people in places far apart. For instance, as seen in Table 3, the postal code 48104 (representing a university campus in Ann Arbor, Michigan) is most similar to 55414, another university campus in Minneapolis, Minnesota, about 900 km away. We further hypothesize that similarity among users is not primarily based on straight-line distance but on a more nuanced distance – cultural or socioeconomic – that manifests in a community consensus of agreement on item likeness, in the same way that the semantic relatedness of spatially-referenced Wikipedia articles is not fully consistent with geodesic promiximity [7].

We also believe that clusters based on location provide advantages over clusters based directly on users' ratings [18]: because (a) the number of postal codes is few compared to the number of users, an important factor when clustering within a very large rating matrix; (b) the exogenous information from location can help explain the clusters.

9. LIMITATIONS

An important limitation of our work is that our experiment considered only United States postal codes. The United States has a large variety of cultures and demographics spread throughout the country. Different countries may display different properties.

The other limitation is the low user density within postal codes used for evaluation. We find 1779 postal codes associated with 5000 users which means we have on average 3 users per postal code. Recognizing this limitation, we performed a predictive accuracy experiment on an equivalent sized dataset by picking postal codes (which has its own bias of oversampling users from dense locations) and found RMSE results were consistent.

In addition, our experiment considered only a Bayesian average to calculate postal codes' item rating midpoints. Different similarity methods, such as Kullback–Leibler divergence [13] of rating distributions may provide better understanding of similarity between two places.

While we calculate local neighbors by proximity, LARS tessellated users into geographic grid squares. One of LARS' major contributions was spatial data structure optimization for best storage and recommendation performance. The use of a spatial database is out of the scope of our paper.

Finally, MovieLens users enter postal codes at registration and rarely update that information, even if they move.

10. FUTURE WORK

There is significant complexity around what kind of location-aware partitioning may improve recommendations. Other methods might include **ESRI's Tapestry dataset**, which groups postal codes not on item rating similarity but with "lifestyle" – that is, demographic and socioeconomic – weighted tags. For example, the top lifestyle segments for 94043 (Mountain View, California) are Enterprising Professionals (38%), Trendsetters (14%), and Laptops and Lattes (10%).

We are interested in how this process applies to item domains other than movies, especially for **items with specific spatial footprints**. We intuit that a distant neighbors approach may work less well considering individual places (say, a specific restaurant) but may work just as well considering place features (say, a cuisine that applies to many restaurants).

Location-aware recommenders should be able to use the context of location to **explain** their suggestions: e.g., that people in a similar location offered relevant opinions. Future work involves exploring the value of location-based explanations over those focused on other dimensions of similarity. Herlocker et al. find that 86% of recommender system users value additional explanations [8], and location-based partitions are explainable. Placenames are immediately available, and demographic inference (e.g., "another college town") is possible.

Finally, although we chose postal codes as the unit of **spatial granularity** in our system, location is defined as part of a hierarchy [3] and we could have been more or less precise. In an area of the world with smaller nations and stricter cultural borders – say, Europe or Central America – we may find the applicable scale is at the level of nations and languages. Spatial granularity may also be inconsistently scaled within the same dataset: if speaking German is predictive, it may be at the national level in Europe, regional level in the United States, and city level in South America.

11. CONCLUSION

We demonstrate that we can harness geographic techniques to improve recommender systems by limiting the recommender model to groups of users in similar locations. We show similar locations are determined by a combination of ratings preference and geodesic distance, an understanding that goes beyond a "closer = more relevant" assumption. We provide the first evidence that rating similarity of locations does not monotonically decrease with geodesic distance, and describe a geostatistical method to discover "distant neighbors". We show that creating a user space based on location similarity improves predictive accuracy and recommender utility, including the new user cold start scenario and controlling for full coverage.

Over-relying on the opinions of nearby places can lead to the inclusion of proximate but less similar people, whereas the power of distant neighbors in recommender systems is smart partitioning of the underlying dataset to include the right users among all those far away.

We urge a reconsideration of geographic relationships in recommender systems, where the value of geography goes beyond geodesic distance.

12. NOTES

Our code and datasets, as well as results for additional metrics, are available at http://cs.umn.edu/∼vikas.

13. ACKNOWLEDGMENTS

This work was supported in part by National Science Foundation (IIS-0808692 and IIS-0964695) and the University of Minnesota's Social Media and Business Analytics Collaborative (SOBACO). We also would like to thank Daniel Kluver (University of Minnesota) and Michael Ekstrand (Texas State University) for their continuous feedback and input in the paper.

14. REFERENCES

[1] G. Adomavicius and A. Tuzhilin. Context-aware recommender systems. In *Recommender systems handbook*, pages 217–253. Springer, 2011.

[2] J. Bao, Y. Zheng, and M. F. Mokbel. Location-based and preference-aware recommendation using sparse geo-social networking data. In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*, pages 199–208. ACM, 2012.

[3] J. Bao, Y. Zheng, D. Wilkie, and M. Mokbel. Recommendations in location-based social networks: a survey. *GeoInformatica*, pages 1–41, 2015.

[4] B. Bill. *The Big Sort: Why the Clustering of Like-Minded America Is Tearing Us Apart*. New York: Houghton Mifflin Company, 2008.

[5] J. Das, S. Majumder, and P. Gupta. Voronoi based location aware collaborative filtering. In *3rd National Conference on Emerging Trends and Applications in Computer Science (NCETACS)*, pages 179–183. IEEE, 2012.

[6] K. E. Grossner, M. F. Goodchild, and K. C. Clarke. Defining a digital earth system. *Transactions in GIS*, 12(1):145–160, 2008.

[7] B. Hecht and E. Moxley. Terabytes of Tobler: evaluating the first law in a massive, domain-neutral representation of world knowledge. In *Spatial information theory*, pages 88–105. Springer, 2009.

[8] J. L. Herlocker, J. A. Konstan, and J. Riedl. Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference on computer supported cooperative work*, pages 241–250. ACM, 2000.

[9] T. Horozov, N. Narasimhan, and V. Vasudevan. Using location for personalized poi recommendations in mobile environments. In *International Symposium on Applications and the Internet (SAINT)*. IEEE, 2006.

[10] D. Kluver and J. A. Konstan. Evaluating recommender behavior for new users. In *Proceedings of the 8th ACM conference on recommender systems*, pages 121–128. ACM, 2014.

[11] J. A. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, and J. Riedl. Grouplens: applying collaborative filtering to usenet news. *Communications of the ACM*, 40(3):77–87, 1997.

[12] O. Kucuktunc, B. B. Cambazoglu, I. Weber, and H. Ferhatosmanoglu. A large-scale sentiment analysis for Yahoo! answers. In *Proceedings of the fifth ACM international conference on web search and data mining*, pages 633–642. ACM, 2012.

[13] S. Kullback and R. A. Leibler. On information and sufficiency. *The annals of mathematical statistics*, pages 79–86, 1951.

[14] J. J. Levandoski, M. Sarwat, A. Eldawy, and M. F. Mokbel. Lars: A location-aware recommender system. In *28th International Conference on Data Engineering (ICDE)*, pages 450–461. IEEE, 2012.

[15] Q. Li, Y. Zheng, X. Xie, Y. Chen, W. Liu, and W.-Y. Ma. Mining user similarity based on location history. In *Proceedings of the 16th ACM SIGSPATIAL international conference on advances in geographic information systems*, page 34. ACM, 2008.

[16] T. J.-J. Li, S. Sen, and B. Hecht. Leveraging advances in natural language processing to better understand Tobler's first law of geography. In *Proceedings of the 2014 SIGSPATIAL Conference*. ACM, 2014.

[17] N. L. Oden. Assessing the significance of a spatial correlationogram. *Geographical Analysis*, 16(1):1–16, 1984.

[18] A. M. Rashid, S. K. Lam, G. Karypis, and J. Riedl. ClustKNN: a highly scalable hybrid model- & memory-based CF algorithm. *Proceeding of WebKDD*, 2006.

[19] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, pages 285–295. ACM, 2001.

[20] Y. Seroussi, F. Bohnert, and I. Zukerman. Personalised rating prediction for new users using latent factor models. In *Proceedings of the 22nd ACM conference on hypertext and hypermedia*, pages 47–56. ACM, 2011.

[21] W. Tobler. On the first law of geography: A reply. *Annals of the Association of American Geographers*, 94(2):304–310, 2004.

[22] W. R. Tobler. A computer movie simulating urban growth in the detroit region. *Economic geography*, pages 234–240, 1970.

[23] S. Vargas and P. Castells. Rank and relevance in novelty and diversity metrics for recommender systems. In *Proceedings of the 5th ACM conference on recommender systems*, RecSys '11, pages 109–116. ACM, 2011.

[24] I. Weber and C. Castillo. The demographics of web search. In *Proceedings of the 33rd international ACM SIGIR conference on research and development in information retrieval*, pages 523–530. ACM, 2010.

[25] B. Xu, J. Bu, C. Chen, and D. Cai. An exploration of improving collaborative recommender systems via user-item subgroups. In *Proceedings of the 21st international conference on World Wide Web*, pages 21–30. ACM, 2012.

[26] X. Yang and Z. Zhang. Combining prestige and relevance ranking for personalized recommendation. In *Proceedings of the 22nd ACM international conference on information & knowledge management*, pages 1877–1880. ACM, 2013.

[27] Y. Yu and X. Chen. A survey of point-of-interest recommendation in location-based social networks. 2015.

[28] J. Zhang, M. Zhu, D. Papadias, Y. Tao, and D. L. Lee. Location-based spatial queries. In *Proceedings of the ACM SIGMOD international conference on management of data*, pages 443–454. ACM, 2003.