Inferring the Character of Urban Commercial Areas from Age-biased Online Search Results

How place recommendation data can reveal dynamic Seoul neighborhoods

David Lee
Department of Civil and
Environmental Engineering
KAIST
Daejeon, South Korea
david733@kaist.ac.kr

Seolha Lee
Department of Civil and
Environmental Engineering
Seoul National University
Seoul, South Korea
joeunseolha@gmail.com

ABSTRACT

We analyze the consumer-age-specific patterns of restaurant preferences in commercial areas of Seoul, through the mining of place recommendation results from the Naver Place online service. We calculate indices for 188 distinct areas of Seoul measuring the heterogeneity of taste across age groups, and the dominance of any one age group over the general options presented to the public. Our results suggest that both high-traffic and rapidly changing commercial areas present diverse options appealing to all age groups, and that this diversity is primarily driven by the tastes of younger age groups. Recognizing these patterns may help stakeholders predict gentrification and proactively shape neighborhood transformation from business turnover. This study contributes to the broader literature on applying online behavioral data to study urban economic activity.

CCS CONCEPTS

• Human-centered computing~Geographic visualization • Social and professional topics~Age • Applied computing~Economics

KEYWORDS

gentrification, urban economy, online search behavior, point-of-interest recommendation

1 Introduction

Many online portals provide geographic search tools to help users locate or discover points-of-interest in unfamiliar or dense urban environments. These results may be presented alongside website links and other resources returned from traditional text search (e.g. Google uses the context of the user's IP address or device geolocation as contextual cues to filter such returned locations), or directly from a search option in a map/directions application like Google Maps. These place recommendations are themselves indexed and their presentation influenced by the activities of many prior users of the service, in order to provide more relevant results to subsequent users with similar interests.

The Korean web portal and search engine Naver, one of the two dominant online platforms in South Korea, launched its Naver Place recommendation engine in 2018. This geographic search tool allows users to query for points of interest like restaurants or cafes within a geographic area, but adds a novel feature that arranges the results based on the user's age (as selected from a list of age ranges). This algorithm, which remains opaque to the end user, decides on which results to show and in what order to display them, based on past search and online review activity from other registered users within that age range.
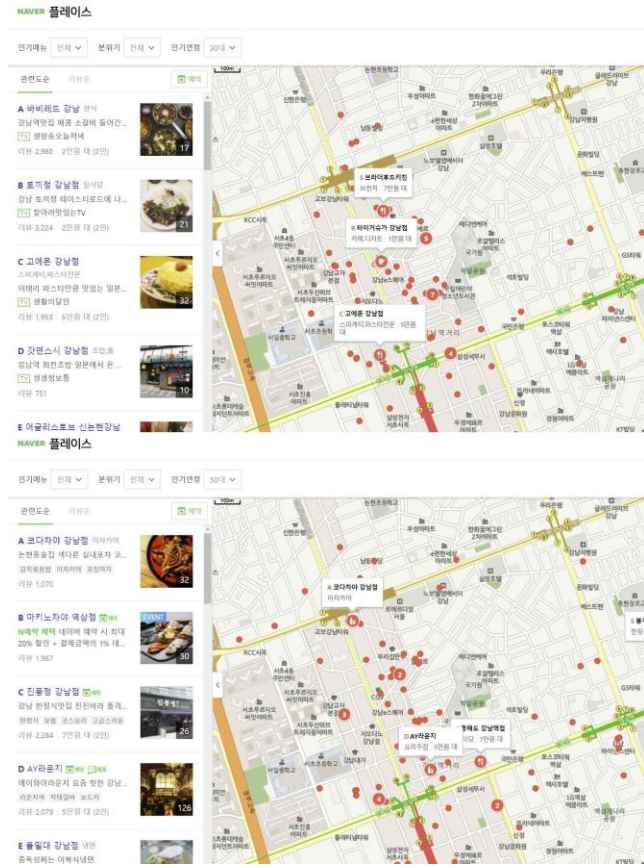
Figure 1: Results from searching for restaurants, when applying the age 30-40 filter (top) vs age 50+ filter (bottom) in the Gangnam district of Seoul. Each of top five results differs between the two queries. Screen capture from Naver Place web application (https://store.naver.com).

In addition to providing more relevant information for discerning users, these results also reveal how the experience and appeal of these urban areas may differ for people of different age groups. For instance, a dense neighborhood like the famous Gangnam district of Seoul may offer unique sets of restaurant and entertainment choices for different generations of users. Younger visitors may prefer Western-style restaurants and bars, while older visitors look for more traditional Korean cuisine. The heterogeneity of these suggested results reflects on the density, variety, and vibrancy of these areas.

Alternatively, we can compare the tailored search results from each age group with the default, non-age-filtered results from the same area, to identify to what degree these results are skewed towards one particular age group over the others. This could provide evidence that some urban areas are becoming dominated by businesses targeted at particular age groups, such as recently gentrifying areas that mainly attract visitors in their 20s with trendy restaurants and coffeeshops.

In this study, we used the Naver Place engine to generate place recommendation lists for 188 distinct areas of Seoul, across five age groups. By cross-comparing these lists, we generated indices that measured the heterogeneity of taste across age groups, and the dominance of any one age group on the general options presented to the public. Our results suggest that both high-traffic and rapidly changing commercial areas present diverse options appealing to all age groups, and that this diversity is primarily driven by the tastes of younger age groups.

2 Related Work

2.1 Point-of-interest recommendation

With the rise of online mapping and location-based social networking, many researchers and firms have sought to develop effective ways to search for and recommend points of interest (POIs) to users. Understanding a user's preferences and intent are critical to providing relevant results, and any approach must rely on various contextual clues, rich geographic datasets, and algorithms to infer the best results. Influential approaches include:

- Using user's real-time-location, distance filters, and preferred search type (global ratings, personal ratings, similar user-ratings, etc.) to filter and present POI recommendations [1].
- Combining social influence through location-based social network activity and geographical influence through spatial clustering of user activities [2].
- Accounting for time-of-day to predict user activity and intent [3].

Many contemporary approaches rely on "big data", compiling many online user interactions, social media content, and recorded mobility traces, and developing predictive algorithms to provide personalized POI search results. However, there are few explicit mentions of user age as a criterion to segment user populations in the learning or application phase.

2.2 Mining POI data to describe urban place

Many studies of urban planning, economics, and spatial patterns rely on mined POI data to provide context to the geography of cities. Some sources and uses of POI data mined from user-facing applications include:

- Volunteered geographic information (VGI), in the form of place listings in the Yahoo! Local service, used to identify and classify POIs, which are then used to disaggregate land use data to finer detail [4].
- VGI from OpenStreetMap used to collect POIs and estimate urban population at the building scale [5].
- Combining POI databases and human mobility data (as historic taxi GPS traces) to classify the functional regions of a city [6].

2.3 Segmenting urban locations by user preferences

A common application of "location intelligence" tools, for businesses deciding on where to locate, is to understand the

potential market and types of clients that inhabit a region. One example is the Tapestry segmentation dataset sold by ESRI, which classifies neighborhoods based on demographic and socioeconomic characteristics of their residents, such as age, occupation, family size, etc., going so far as to infer some of the behavior preferences of these defined groups [7]. A variety of other digital traces can be used to classify urban places based on users' connected behavior, such as data traffic [8] and credit card transactions [9].

3 Methods

3.1 Extracting information from Naver API

Though Naver provides API endpoints for many of its search products, there was no documented service for the Naver Place engine. Instead, we reverse-engineered how the Naver Place web application formed its queries as GET requests in the body of the URL. Each query we used included the following fields:

- Bounding box filter: Restricts search to rectangular area defined by southwest and northeast coordinates.
- Age: Lower-bound of the 10-year age range. Age 50 returned results meant for users 50 and older. Omitting this field removed this age filter altogether.
- Query text: Text string for the search. In this experiment we only used the string "맛집", a common term used in Korean slang to identify popular or well-regarded restaurants.

We divided the Seoul metropolitan city area into a grid of 1.5 km x 1.5 km squares, removing areas sparsely populated due to mountains and water bodies, to form 188 separate bounding boxes to search within.
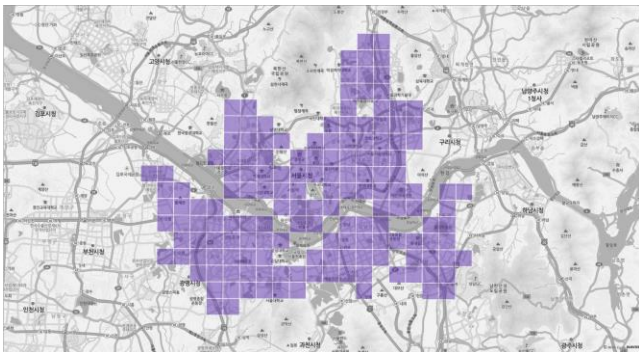
Figure 2: Map of 1.5 km x 1.5 km squares covering extent of Seoul Metropolitan City. Unpopulated areas removed from the grid. Basemap from OpenStreetMap.

From a Python script, we ran the same query six times over each bounding box, once each for the 10-20, 20-30, 30-40, 40-50, and 50+ age ranges, along with the default "all ages" query. The script would scrape the returned webpage, which included the structured search results in JSON format in its source code, and save these to separate tables. We ran these queries on June 22, 2018.

We kept the first 10 results from each query. Each result included the restaurant name, address, ranking within the list of results, location in coordinates (these had to be re-projected from Naver's custom coordinate reference system to WGS84 coordinates using a separate script), and unique identifier. These identifiers persisted across queries, allowing us to cross-compare results and rankings from different age groups.

3.2 Forming the indices

3.2.1 Taste heterogeneity index. To capture how much the preferences of different age groups diverge within a given area, we calculated each area's THI, or taste heterogeneity index. The THI is a weighted sum of the number of unique places that appear in the combined top 10 results of all age groups; the more often a place appears, the lower it is weighted. Thus, THI will be lowest in areas where the result lists from each age group are identical (ignoring order), and highest where the lists are completely unique.

Where k represents the number of age groups queried, and c_i represents the count of each of the n unique respective elements from the combined result list from all age groups:

$$THI = \sum_{i=1}^{n}(k - c_i + 1)c_i \qquad (1)$$

3.2.2 Taste dominancy index. To capture how an area might be dominated by the preferences of one age group at the expense of others, we calculated each area's TDI, or taste dominancy index. The TDI is simply the number of items from each age group's top 10 result list that overlap with items in the default, non-age-filtered result list. TDI will be lowest when these lists do not share any elements (0), and highest when they contain identical sets of elements (10).

$$TDI_{age\ group} = \left|S_{all} \cap S_{age\ group}\right| \qquad (2)$$

Note that while each grid square area has one THI calculated using all five age groups, each area has five separate TDI respective to each age group. When an area has divergent TDI values, with one higher than the others, it can be interpreted as being "dominated" by that age group.

3.3 Analysis

After calculating the indices, we visualized them in choropleth maps overlaid on a base map of Seoul, using QGIS. We also measured the correlation between the TDI values across pairs of age groups, and between THI and TDI.

4 Results
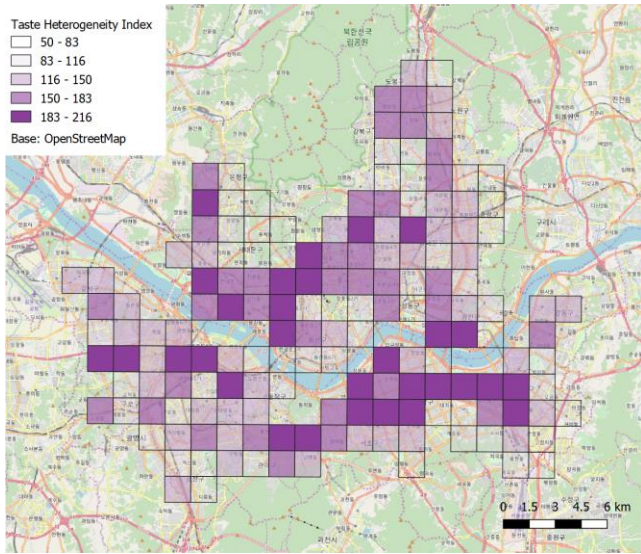
Figure 3: Choropleth map of calculated THI in Seoul. Darker zones represent areas with more heterogeneous preferences across age groups. Basemap from OpenStreetMap.
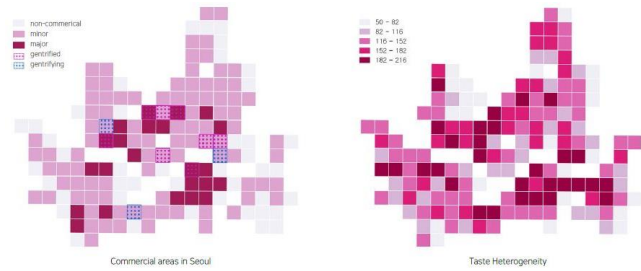

Figure 4: Side-by-side comparison of commercial areas in Seoul (left) and calculated THI (right). Commercial areas and gentrified/gentrifying areas were identified and officially designated by the city government.

When overlaid on a map of Seoul, the high-THI areas of the map roughly overlap with three types of commercial area in the city:

- Highly trafficked commercial corridors, such as the Gwanghwamun – Seoul Station axis, or Gangnam – Yeoksam – Jamsil.
- University-adjacent entertainment districts, such as near Hongdae, Korea University, or Konkuk University.
- Gentrifying areas, such as Mangwon-dong or Yeongdeungpo-dong.

This aligned with our intuitive understanding of these areas. The major commercial corridors offer a wide variety of restaurant choices for all tastes, through sheer density, and host both legacy establishments and newly trending options. On the other hand, university-adjacent districts and gentrifying areas rapidly adapt to the tastes of younger visitors, offering choices distinct from the older businesses that they threaten to eventually displace.
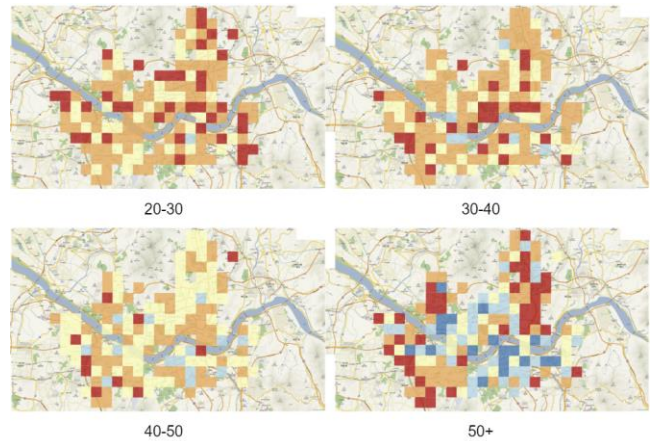
Figure 5: Choropleth maps of calculated TDI in Seoul for age groups 20+. Redder zones represent areas where that age group's tastes align closest with the default, non-age-adjusted results.

Visualizing TDI across the different age groups, we see clear differences in how closely the groups align with the default results. In particular, TDI for the 20-30 group is high throughout the city, but especially so in the commercial districts highlighted earlier; by contrast, these are the same places where the 50+ group are weakly aligned with the default results and show distinct preferences. Correlation within the age groups reveals that the 20s and 30s groups are the most correlated (0.64) while the 20s and 50+ groups are the least (0.32).
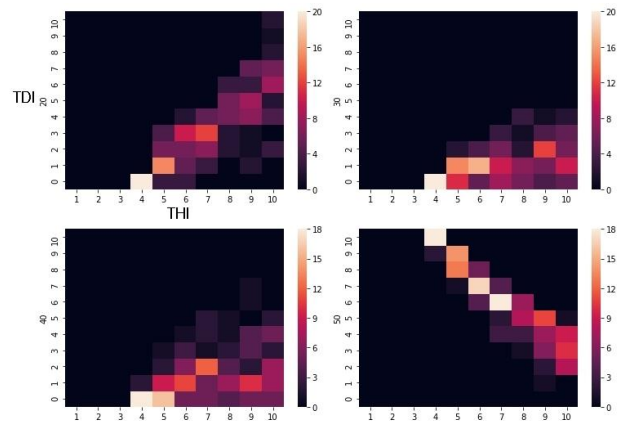
Figure 6: Scatter density plot between TDI and THI for age groups 20-30 (upper left), 30-40 (upper right), 40-50 (lower left) and 50+ (lower right).

Finally, when correlating THI with TDI, we again see clear differences among the age groups. The 20s group sees the strongest correlation (0.87) between the indices; in the places that present more diverse options to the various age groups, they tend to dominate default search results, more so than the other groups.

By contrast, the 50+ group shows a negative correlation (-0.7) between the indices; as presented options become more diverse, their restaurant options tend to diverge the most from the default results.

5 Conclusion

South Korea is an aging country, with a majority of the population expected to shift into the 50+ age range in the coming decades. Tailored-search tools like Naver Place provide ways for older urban residents to surface content more relevant to their needs, but in doing so also reveal the distinctness of those needs in relation to the overall user base.

These age-based preferences offer us a new way to observe the character of commercial areas in the city. By highlighting areas in the city where tastes appear to diverge among the age groups, we can identify areas that offer a dense or dynamic number of options to a diverse set of patrons. These places are the core of the urban experience that appeals to many different people in different ways.

However, the results also imply that such places are significantly biased toward younger users, whose preference dominate the overall search results at the expense of older generations. The preferences of older users are marginalized, either online through the skew of search results away from them, or in the actual distribution of restaurants in those areas catering to their tastes. In a sense these places have been gentrified in virtual and/or real space.

The indirectness of these indices is a key limitation of their usefulness. Because the place recommendation engine algorithm is opaque to the end user, as are the datasets it operates on, we can only speculate on how closely these results reflect user preference vs user behavior. Furthermore, it is unclear how much the results can tell us about the actual commercial activity in these neighborhoods; not every business is listed in Naver's database, and the likelihood of a business appearing there may correlate with the age of its clientele.

For future work, we will explore the use of these datasets to predict neighborhood gentrification. Many parts of Seoul have experienced such change for the last twenty years. However, the context is different from that of U.S. and European cities. In Seoul, the significant change signaling gentrification is not a change of residents, but a change of business target. Starting with a few trendy shops entering an old neighborhood, local merchants are rapidly replaced with new shops, especially food establishments, which mainly target outside visitors as the area gains popularity. This abrupt change eventually leads to increased rent prices and exodus of the original retailers. Even though these changes have been anecdotally reported, there hasn't been a reliable indicator or predictor of this phenomenon.

To characterize these economic changes more precisely, we can combine these indirect indicators of diverging local tastes with other conventional datasets, and identify the dominating trend of popular food establishments in terms of age, type, and concurrent building usage. These attributes can directly reveal the trend of certain new food establishments preceding rapid business turnover, as well as the increase of rent price traditionally associated with gentrification.

REFERENCES

[1] M. A. Sheha, A. Sheha & S. Petilli. (2006). U.S. Patent No. 7,082,365. Washington, DC: U.S. Patent and Trademark Office.

[2] M. Ye, et al. (2011, July). Exploiting geographical influence for collaborative point-of-interest recommendation. In Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval (pp. 325-334). ACM.

[3] Q. Yuan, et al. (2013, July). Time-aware point-of-interest recommendation. In Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval (pp. 363-372). ACM.

[4] S. Jiang, et al. (2015). Mining point-of-interest data from social networks for urban land use classification and disaggregation. Computers, Environment and Urban Systems, 53, 36-46.

[5] M. Bakillah, et al. (2014). Fine-resolution population mapping using OpenStreetMap points-of-interest, International Journal of Geographical Information Science, 28:9, 1940-1963

[6] J. Yuan, Y. Zheng & X. Xie. (2012, August). Discovering regions of different functions in a city using human mobility and POIs. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 186-194). ACM.

[7] Tapestry. Retrieved June 28, 2019 from https://www.esri.com/en-us/arcgis/products/tapestry-segmentation/overview.

[8] S Grauwin, et al. (2014), Towards a Comparative Science of Cities: Mobile Traffic Records in New York, London and Hong Kong, in *Computational Approaches for Urban Environments* (Ed. Springer)

[9] S. Sobolevsky, et al. (2015). Cities through the Prism of People's Spending Behavior. PLOS ONE.