A study on monitoring method of urban commercial area

Ryomei Onishi
Graduate School of Frontier Sciences, the University of Tokyo
IIS Cw-503 4-6-1 Komaba, Meguro-ku Tokyo Japan 153-8505
ryoumei@iis.u-tokyo.ac.jp

Yusuke Akiyama
Graduate School of Frontier Sciences, the University of Tokyo
IIS Cw-503 4-6-1 Komaba, Meguro-ku Tokyo Japan 153-8505
aki@iis.u-tokyo.ac.jp

Ryosuke Shibasaki
Center for Spatial Information Science, the University of Tokyo
IIS Cw-503 4-6-1 Komaba, Meguro-ku Tokyo Japan 153-8505

Hiroya Tanaka
Tanaka Laboratory, Keio University
Keio University Shonan Fujisawa Campus 5322 Endoh, Fujisawa-shi, Kanagawa-ken
htanaka@sfc.keio.ac.jp

Abstract: In terms of urban area management, keeping urban area in modest scale is one of the most important issues. We focused on urban commerce area and make analysis based on various data source. Our analysis was carried as below. First, for the sake of preparing the base dataset, we integrate NTT data (Japanese internet yellow page, including commercial shops and firms data) and Zenrin map (a popular urban map) into one map data on GIS, through address matching and natural language processing. Secondly, we focused on the distribution of shops/firms in commercial area, and plotted them on the GIS map data, which covers the area around Tokyo. Then we make clusters about every shops based on minimal distance among plotted data, with Nearest Neighbor. Method. Through this clustering procedure, we generated a monitoring map for commerce area, with which we can monitor the rate of change of every shop/firm at individual building level, and the degree of concentration among them. Then we calculate the modest scale of each commercial area with this monitoring method.

In this study, we could examine monitoring method about urban commerce area. And our future prospect is dealing with more diversified aspects of urban area with this method, because it is necessary more indices to calculate the modest scale of urban area. Calculating those scales is also important issue in urban management, as typified by Compact City Theory. And our monitoring method should be the support of validating such theories.

Keywords: GIS, Data mining, commercial area, location theory

1. Introduction

1)background

There is a lot of data sources for analyzing region, or one area, but most of all is not united. In this study, we examined to unite such various data source on one platform. Off course, every dataset has different purposes, different format. And there is not only spatial information/maps), but also information only with language. If through modest algorithm, we can arrange them into one spatial information, and also can make close up analysis for one region.

Fig1. Background and purpose
2) Purpose
Our purpose is to make it possible to make various datasets through such data reforming processes, what we call “monitoring method for region”. Monitoring method for region is the data arranging method for detailed analysis for region. Firstly, we consider about how we develop this method, and put this method to practical use for analyzing regional problems.

2. Data arranging

As we mentioned, it is very important to make appropriate data arranging algorithms in this study. Data arranging process is divided into two processes. One is making literal information into spatial information. And after that, the other is making such spatial information into one dataset.

1) Literal information into spatial information

So we should, in order to make various data sources into one spatial data. This as below, and we name it as “monitoring method of region”

Fig 2. model of “monitoring method for region”

For example, we formed NTT Town page (phonebook) into GIS dataset. This digital phonebook is available on the web, and it includes basic information, such as We went through the following steps to set up this dataset.

NTT town page, Japanese phonebook is available and of course include address information. If information has address part, we can add longitude, and latitude information. But phone book is not normalized, with a lot of rack of information, items, and terms. So we have to add it and normalize the original file.

Fig 3. normalizing and geocoding

2) Setting spatial information into one dataset

Then we have gotten the set of spatial information, but there is no security for the data is of one accord. So we should measure the accuracy and tie them. Firstly, we consider the matching of positioning, because every spatial information is from different data sources, and sometimes out of synchronization between same point (buildings, etc). In this case, we are going to match phonebook (point data) into Zenrin Map (house map), polygon data. The latter has also a point data
(center of gravity of buildings), and if we tried to match them by the accuracy between point data, there is no matching, although there is 900,000 data per year. So we choose matching between polygon data and point data.

And among the point data which is not included in polygon data, there should be a correct data. We can tell through checking the name of both data, whether the point and polygon indicate same data or not. If both data has same name or similar name, we can distinguish it as same data. So we developed how to calculate the “similarity” between two name. It is called generically as k-gram, the idea of natural language processing. The method shown below is the concept of k-gram, and we tried to match the data which has similar name.

Shown as above, this is calculating method of similarity between language. Now we are trying to make it work, but it also has difficulty in appanage of language, such as coding of each character. And this problem is also depend on special characteristics of Japanese as language, which include more than one kind of character.
2) case study – commercial area and banking branch

The figure shown below is the result of data matching, between textual information and spatial information. Through monitoring method of region, we can renew this map as updated version.

Our future prospect is adding this map more textual information, which are not intended to be used as spatial information, such as quarterly corporate report or almanac of corporation. With adding such corporative information, we can analyze and validate appropriate strategy for one area.

3. Conclusions
Monitoring for urban commercial area is not possible with only spatial data. It is necessary more information just like corporative information, and regional statistics. With such various data, and make them into one data set, we can analyze urban commercial area, which include intertwining and complicated, subject. Our aim was to gather various data, such as spatial data and textual data, and take advantage of them for regional analyzing. Principally, we found most difficult part is normalizing textual data which include a lot of rack in terms and items. And for the first step, we could get phonebook data into analyzing data set. And about another textual data, such as quarterly corporate report, there are problems and difficulties in each data. Each data has indigenous and specific form and rack of information. Then our next step is developing method for normalizing disordered textual data and visualizing such a various textual information on one map.

References