

Development of spatio-temporal urban data by the integration of digital maps and yellow page and comparison with statistical data

Yuki Akiyama[1]; Ryosuke Shibasaki[2]

[1] Frontier Science, The University of Tokyo; [2] CSIS,UT

<http://shiba.iis.u-tokyo.ac.jp/>

Distributional variations of offices and shops in urban areas have a great impact for vitalities and attractions of cities. It is indispensable to quantitatively monitor these variations for continuance and renewal of urban areas. However, statistical data used widely for urban analysis does not have fine spatial resolution for detailed urban monitoring.

In contrast, time-series variations of every shop and office can be monitored by field surveys or time-series manual decipherment of house maps between two times. Though, a large amount of labor is needed to use these methods, thus, these methods can cover in a very limited area.

In this research, we have developed the methodology to make detailed spatio-temporal urban dataset through the integration of digital house maps and yellow page data that can cover all over Japan and detection of time-series variations of all tenants automatically. There are two steps to make this dataset. First step is to integrate digital maps and yellow page data in same time. Second is to integrate old maps that have attributes of yellow page data and new maps, namely spatio-temporal integration.

Our methodology realizes the 3D spatial integration by automatic analysis of building information and the identify determination of tenant names that can consider the name fluctuation of description between two tenant names. The name fluctuation of description is that descriptions of tenant names between each data or between different two years are subtly different even same tenants. We have proved this problem to quantify the similarity of tenant name using one of the methods of natural language processing called the n-gram.

In addition, we have made libraries of noise words, such as frequently appearance words, geographic names and station names that affect results of identify determination of tenant names. Using these libraries, noise words are deleted form tenant names, therefore, our methodology can integrate each data and old and new data with a high degree of accuracy. Using this methodology, we can monitor time-series variations of all tenants between different any two times in any areas of Japan. Time-series variations mean continuation, change, emergence and demise of tenants.

We have already confirmed that our methodology can deal with all digital maps and yellow page data in south Kanto region between 1995 and 2000, and between 2000 and 2005 accurately. Our system has technically viable environment that can deal with source data in any part of Japan.

In addition, we have revealed reliability and correction method of our data through the cross-check existing statistical data with our data. We use the Establishment and Enterprise Census published by Statistics Bureau of Japan. The Census includes variations of tenants totaled by each municipality and by 17 kinds of business categories in all prefectures of Japan. We have compared total number of tenants in each 17 business categories of that Census and our data in all wards of 23 wards of Tokyo. Many business categories of our data discord business categories of the Census. Therefore, our data have been matched the Census using matching table of each business categories, and replacing business categories of our data with the Census. The figure shows matching results and comparisons of our data with the Census of each 17 business categories in 23 wards of Tokyo. Many business categories have similar results without agriculture, forestry and fishery that number of tenants is very few despite one-year time lag between our data with the Census.

The Census is updated every 2 or 3 year, in addition, time span of tenant variation is fixed as 5 years. On the other hand, these restrictions can be removed using our data.

In the past, few studies have tried to develop methodology such as our research. The detailed time-series urban dataset with high flexibility like our data is unexampled not only in Japan but also foreign countries in past times.

