MSS Population Estimation Privacy Protection

Technology Reports

"Mobile Spatial Statistics" Supporting Development of Society and Industry
— Population Estimation Technology Using Mobile Network Statistical Data and Applications —

Supporting Growth in Society and Industry Using Statistical Data from Mobile Terminal Networks — Overview of Mobile Spatial Statistics —

In this article, we give an overview of MSS, which are population statistics generated from operational data from mobile terminal networks. While most conventional population statistics are based on the residential population, (i.e., the population statistics based on where people live) MSS enables us to know actual population, which is constantly changing and reflects the movement of the people in their daily lives. Thus, they can be expected to contribute to developing and enhancing society. To preserve subscribers' privacy, MSS are generated through a three-step process of de-identification, estimation and disclosure limitation. Use of MSS can contribute to development of society and industry by improving standards and quality of life for individuals, and can contribute to realization of "Smart Life."

Research Laboratories
Ichiro Okajima
Satoshi Tanaka
Masayuki Terada
Daizo Ikeda
Tomohiro Nagata

1. Introduction

The national census and basic residency register conventionally used when studying population distribution in Japan contribute greatly to developing society. They are used for deciding national and regional government policy, and are also used in various industries. These existing statistics are compiled based on people's usual residence, so they are called residential population.

Since people are most likely at their residences during the night time, the residential population is expected to capture the actual population during the nighttime. However, many people travel away from their residences during the day time, and this results in changes to the population distribution. For example, on weekdays people tend to gather more in business districts, while on holidays, they gather in shopping and sight-seeing areas. These changes are not reflected in the residential population data, but they are reflected in actual population statistics which is a term for populations based on how many people are actually in a given area. These statistics change from one minute to the next as people move, so they have conventionally been difficult

to ascertain.

Accordingly, we focus on mobile terminal network systems as a means for estimating the actual population continuously on a national scale. Approximately 40% of the total population uses NTT DOCOMO's mobile phone services every day, so utilizing this high adoption rate, accurate estimations of current-location population can be expected.

Specifically, we estimate the actual population from operational data from the mobile terminal network, which is used for providing the mobile terminal service. This enables us to estimate the actual population (1) covering a national scope, (2) over arbitrary grid units or administrative boundaries (with granularity depending on the base station density), (3) with fluctuation over time, 24-hours a day, 365-days a year, and (4) with classifications such as gender, age-group or residential area. We call the actual populations estimated in this manner Mobile Spatial Statistics (MSS).

When generating statistics about the activities of people, it is very important to preserve the privacy of the individuals upon which statistics are based, as well as their utility. To achieve this, we have established a process for maintaining privacy when creating mobile spatial statistics, incorporating the results of a study done by experts[1] on the social, legal and technical aspects of MSS.

In this article, we give an overview of MSS, which apply mobile terminal infrastructure to estimate the actual population as it changes over time, and explain the approach to privacy protection with MSS.

2. MSS

2.1 Overview

Figure 1 shows the structure for MSS and the three types of population values estimated using MSS.

MSS can be used to estimate the actual population over wide areas and fluctuating over time. In other words, they can reveal geographical distributions of population over given time, such as what regions had high populations at a certain time. These values are called population distributions.

Also, by analyzing the changes in population continuously in a given area, they can reveal trends in population fluctuation over time, such as when the population in a given area is its highest. These values are called population transitions.

With MSS, analyzing distributions according to population composition, such as gender, age-group and residential area, can reveal information such as where young women tend to gather, or what business areas people from a given residential area tend to commute to. These values are called population composition values.

In other words, MSS provide methods for continuously estimating the constantly changing actual population

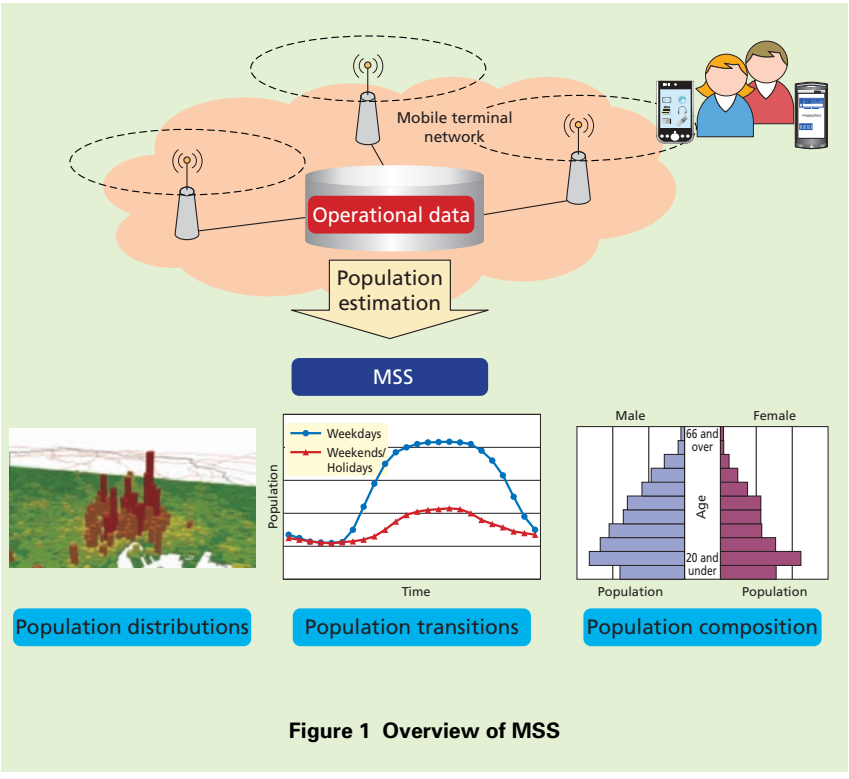

Figure 1 Overview of MSS

over months, days of the week, hours, gender, age-group and area of residence, and for studying changes and transitions in these over time and different regions.

In this way, they can provide support in fields such as urban development[2], disaster prevention planning[3] or local revitalization[4], where an understanding of population distribution, transitions or composition is important to help in making rational decisions on planning, drafting and verifying the effectiveness of policies at national or local government levels. This is expected to contribute to development in society and industry.

2.2 Features of MSS

MSS use operational data from a mobile terminal network to make population estimations, so it has the following features, depending on the properties of the mobile terminal network.

Coverage: Since MSS makes estimations based on the operational data from a mobile terminal network, the coverage of the estimates is essentially the same as the mobile terminal service area. As of the time of writing this article, the service area for NTT DOCOMO mobile terminal included 100% of local government offices for cities, towns and villages in Japan, so MSS has equivalent coverage.

Spatial resolution: The size of unit areas over which population estimations can be made using MSS depends on the distances between base stations in the mobile terminal network. This is not uniform since base stations are densely placed in urban areas where there is a high concentration of people, but they are placed more sparsely in rural areas. As a rough criterion, estimations can be done on a 500 m mesh within the Tokyo metropolitan area, and on a mesh of several kilometers outside of urban areas.

Temporal resolution: The temporal resolution of MSS depends on the frequency with which the base stations detect the mobile terminals within their regions. Typically, the temporal resolution of MSS is one hour.

2.3 Benefits and Limitations

The characteristics of MSS, which estimate of the actual population from the operational data of a mobile terminal network, differ from conventional residential population statistics, most of which are generated from questionnaires, so there are several benefits and limitations when comparing with existing statistics.

The first benefit is that it is easier to survey over wide areas. With conventional statistical studies using questionnaires, the cost increases as the area being surveyed expands or the detail increases, and so does the time required for planning and executing the survey, and for producing the results. In contrast, MSS are created automatically from operational data generated by a mobile terminal network, so the additional cost and time to expand the scope of the statistics are much lower.

The second benefit is that results can be estimated much more quickly than conventional statistics. For example, the census, the most widely-used residential population statistics, is done every five years. Considering the cost and labor involved in distributing and collecting questionnaires and analyzing and estimating the results, it would be difficult to perform the census more frequently. Conversely, MSS are created using the operational data from a mobile terminal network that continuously operates 24-hours-a-day and 365-days-a-year, so statistics at one-hour intervals can be implemented at practical costs.

On the other hand, there are limitations on the age-groups over which MSS can estimate data. Since MSS are estimated from the operational data of a mobile network, estimates cannot be made on age groups where mobile terminal penetration is particularly low. Specifically it cannot be applied to the age group of 80 and over or 14 and under. Thus, population estimations currently possible using MSS are limited to age ranges between 15 and 79.

MSS also have limitations in precision compared to data such as the national census. In principle, the national census involves distributing questionnaires to all residents of Japan, so population estimates from it cover the entire

population, while MSS are estimated from the operational data of the NTT DOCOMO mobile terminal network, so values are subject to estimation error. Thus, before using MSS in any particular field, ample attention must be paid to whether the accuracy needed for that field of application has been attained.

2.4 Safety of MSS

MSS are promising for contribution to development of society and industry by revealing changes in population over time and in composition, which have conventionally been difficult to comprehend, but on the other hand, they require careful attention to protecting privacy.

The mobile terminal network operational data used to create MSS can be divided into two main categories: location data and attribute data.

1) Location Data

Location data is data related to the area where each mobile terminal is located, determined through each of the base stations in the mobile terminal network in order to maintain the mechanism that allows mobile terminals to be paged at any time and any place. Mobile terminals connect to the mobile terminal network through the base stations, and the base stations provide wireless communications between the mobile terminal network and the mobile terminals that are in their coverage area (the base station area, or the "cell"). When a mobile terminal is paged to receive a call or data, the network must know the cell in which the mobile terminal currently exists. It would not be effective to search for the mobile terminal in all of the cells in Japan successively, so the base stations periodically interact with the phones to notify the mobile terminal network of the area they are currently in. This data, related to the state of mobile terminals within each base station area, is called location data.

2) Attribute Data

Attribute data is data related to attributes of the mobile terminal user. In order to appropriately provide services to users having mobile terminals, a mobile terminal network maintains attribute data regarding users such as name and address, gender and birth date.

Both location data and attribute data contain sensitive information related to user privacy, so ample care to protect privacy must be taken when handling it. Thus, when creating MSS, a three-step process is used to preserve user privacy, as shown in **Figure 2**.

(1) De-identification

First, data that can be used to identify individuals and is not needed for estimating populations is removed from the operational data. MSS provides statistics related to population distributions and composition, so there is no need to specify from whom each data item from the operational data is; it is sufficient to estimate values for attributes not specific to individuals, such as gender and age-group, for each time period and area. Thus, before performing the estimation process, which produces population estimates, the operational data are de-identified; i.e., information that can be used to identify individuals such as telephone numbers and names, is removed, and values such as birthdays and addresses are converted to age groups and administrative boundary codes.

(2) Estimation

Based on the de-identified data, the number of mobile terminals in each base station area is aggregated for each user attribute, and then estimates of geographic distributions of the populations, including people not using NTT DOCOMO, are created by taking into consideration NTT DOCOMO mobile terminal penetration rates and information regarding base station coverage areas[5].

(3) Disclosure limitation

The results of estimation must then be revised to protect the privacy of users under extreme conditions, such as in areas where the population is very small. This process is called disclosure limitation or disclosure control, and is also applied to official statistics published by the government. The

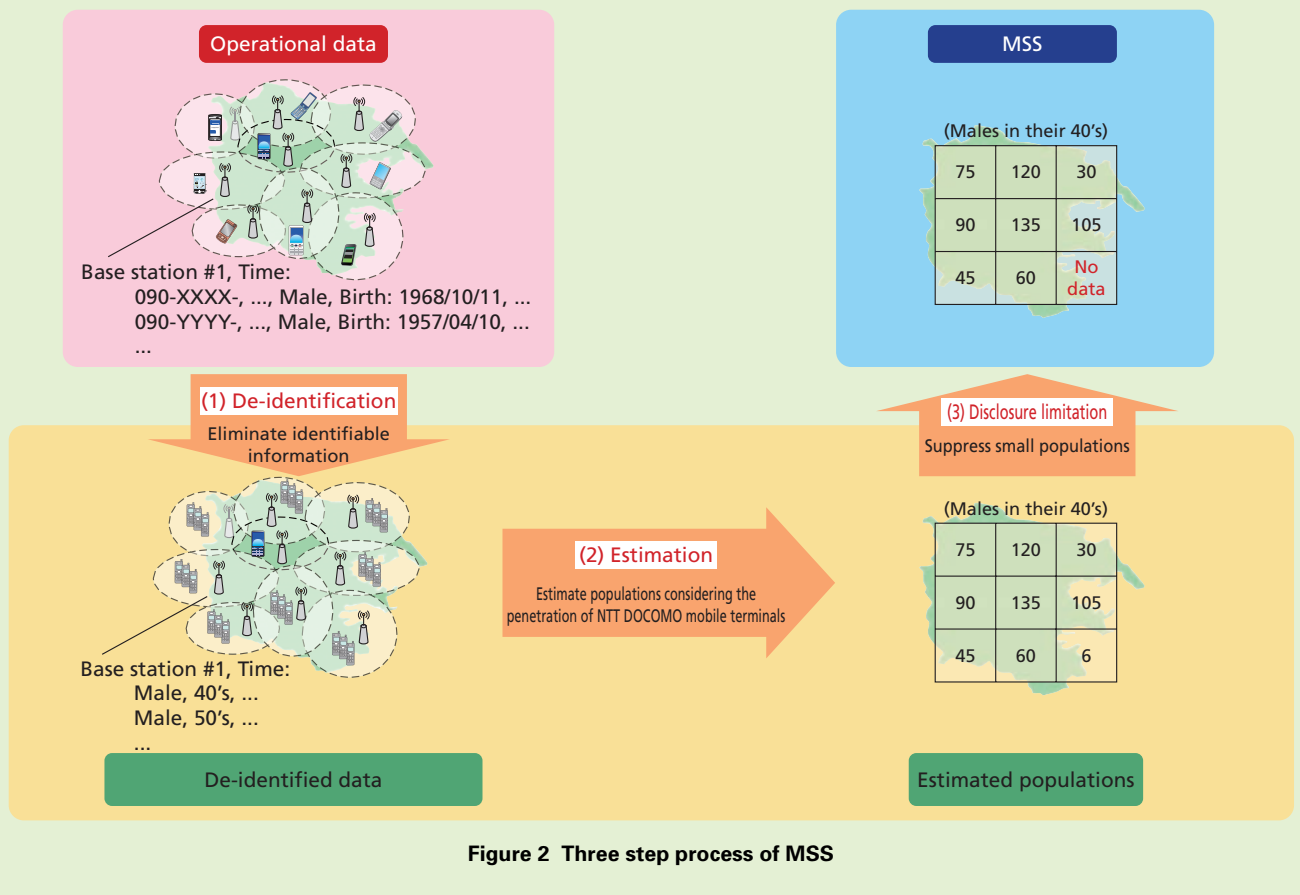

Figure 2 Three step process of MSS

disclosure limitation process used with MSS is based on standards established for official statistics, incorporating domestic and international technical development trends. Data revised in this process is used as MSS.

These procedures used to create MSS are published in a guideline[6], which summarizes the rules that must be observed when creating and providing MSS.

3. Conclusion

In this article, we have given an overview of MSS, which are able to estimate the actual populations over time and over wide areas. By using mobile terminal operational data, MSS are able to estimate populations as they change over one-hour time intervals, and by gender, age-group and area of residence. MSS are created through the three-step process to preserve the privacy of users, so information regarding individuals cannot be disclosed through MSS. We have introduced research on MSS done in collaboration with universities, and it is now in the phase of verifying its utility for applications in various fields. In the future, we plan to verify its utility in a broader range of fields, to continue testing the reliability of MSS, and to begin testing for more-practical applications such as supporting preparation of policy measures in public agencies. We hope to use this experience to contribute to efficient development of society and industry, and to increasing the standards and quality of life for individuals.

REFERENCES

[1] Mobile Society Research Institute: "Report on Requirements for 'Mobile Spatial Statistics' Contributing to Development of Society and Industry," Jun. 2010 (In Japanese).

http://www.moba-ken.jp/pdf/research10_01.pdf

[2] T. Odawara et al.: "Using Mobile Spatial Statistics in Field of Urban Planning," NTT DOCOMO Technical Journal, Vol. 14, No. 3, pp. 31-36, Jan. 2012.

[3] T. Suzuki et al.: "Using Mobile Spatial Statistics in Field of Disaster Prevention Planning," NTT DOCOMO Technical Journal, Vol. 14, No. 3, pp. 37-45, Jan. 2012.

[4] T. Nagata et al.: "Using Mobile Spatial Statistics for Local Revitalization," NTT DOCOMO Technical Journal, Vol. 14, No. 3, pp. 46-50, Jan. 2012.

[5] M. Terada et al.: "Population Estimation Technology for Mobile Spatial Statistics," NTT DOCOMO Technical Journal, Vol. 14, No. 3, pp. 10-15, Jan. 2012.

[6] NTT DOCOMO, "Mobile Spatial Statistics Guidelines." http://www.nttdocomo.co.jp/corporate/disclosure/mobile_spatial_statistics/guideline/