# **Classifying user states in next generation networks**

Y. He $^1$  and A. Bilgic $^2$ 

<sup>1</sup>Institute for Integrated Systems, Ruhr University Bochum, 44780 Bochum, Germany <sup>2</sup>Infineon Technologies AG, 81739 Munich, Germany

Abstract. In this paper we apply a classification method to learn geographic regions using Location Based Services (LBS) in IP Multimedia Subsystem (IMS). We assume that the information in Local Network (cellular network) can be freely exchanged with Global IP Network (IMS) and the information can be gathered in a data base. LBS in the IMS also provide location information for the data sets. Statistic classification methods are applied to the data sets in the data base. Depending on the information provided by the users, they are divided into different user groups (event classes) using Type Filters (TF). Then discriminant analysis is applied to the position information offered by LBS in IMS to determine the geographic regions of the different classes. The learned geographic regions can be used to inform the users in this region or other regions over IMS. This kind of service can be used for any location-based events.

## 1 Introduction

With the increased computing power as well as the evolving wireless networking technology (e.g. 3GPP Long Term Evolution (LTE) (Liu et al., 2006)) many new applications become available for NGN, which are supposed to be all IP based networks.

IP Multimedia Subsystem (IMS) (Poikselkae et al., 2006; Vingarzan and Weik, 2007; OpencourceIMS, 2007) is specified by the 3rd Generation Partnership Project (3GPP) as part of NGN. It is an architectural framework for delivering Internet Protocol (IP) to mobile users, across wireless and wireline terminals. IMS uses both packet switched and circuit switched domain, so it offers network access independent of the different standards (GPRS, UMTS, LTE...). It can be regarded as a platform for IP-based applications, which pro-



*Correspondence to:* Y. He (yuheng.he@ruhr-uni-bochum.de)

vides the possibility to develop an application fast and easy. IMS offers a lot of services, like presence, messaging, quality of service, charging and security (Monsmondo et al., 2006; Khlifi and Gregoire, 2007).

Almost all applications and services in IMS can be supported by Location Based Services (LBS) (Monsmondo et al., 2006; Perusco and Michael, 2007). LBS are wireless "mobile content" services which are used to provide location-specific information to mobile users moving from location to location. Currently, GPS technique (Djuknic and Richton, 2002), network positioning methods (Drane et al., 1998) as well as other positioning methods (He et al., 2008) can be utilized for location estimation for LBS.

In the following, we assume a NGN, which is all IP based, and where information of the users in the network can be exchanged freely. Therefore this information can be collected in data bases accessible by the application servers of IMS. Furthermore, location information is also provided by LBS, such that the position of the users is also available in the data base. We do not consider security issues here, i.e. how to make the gathered information/position anonymous or to encrypt it.

As an exemplary application scenario vehicle to network (V2N) communication is taken in our paper, since vehicles are on the edge of becoming one of the biggest IMS users (Varsheny, 2004). There are different geographic regions on the road when vehicles move from one position to another. The on-road information gained from the sensors of the vehicles or from the vehicle devices manually switched on/off by drivers is gathered in a data base in the IMS application layer. For example, whether there is fog or rain in the position of a vehicle can be obtained according to the states of fog lamps and rain sensors in the cars. Or traffic jam situations can be predicted according to the speed of the cars in a certain area. LBS in IMS (IMSLS: IMS Location Server) provides location information for the users' data sets. The obtained data base can be used to yield traffic scenario mirrors or forecasts.

| User     | Date     | Cell            | Location | Fog      | Rain            |
|----------|----------|-----------------|----------|----------|-----------------|
| ID       | time     | ID              |          | lamp     | sensor          |
| $s_{i1}$ | $s_{i2}$ | s <sub>i3</sub> | $s_{i4}$ | $s_{i5}$ | s <sub>i6</sub> |

Table 1. Data sets in IMS data base.

In this paper, we present how to apply classification (Seber, 2004; McLachla, 2004) for learning geographic regions using LBS in IMS. Depending on the information provided by the users, they are divided into different groups (classes). This process is defined as Type Filter (TF) in IMS. A classification application server uses discriminant analysis to perform the learning of the geographic regions. The learned geographic regions of the groups (classes) can be used to define Location Filters (LF). This case is denoted as "TF-CAS-LF". This kind of case can be used to establish services (e.g. warning services for the vehicles) in this region or other regions over IMS network. Therefore, the presented concept can be used for realizing up-to-date driver support and active safety services.

Section 2 presents filter functions and classification method which are essential in our concept. Implementation and experimental results are discussed in Sect. 3. After that conclusion is made in Sect. 4.

### 2 Classification and filter functions

In order to learn the geographic region, multivariate elements in this region should be analyzed. The statistical analysis of many variables effectively refers to the study of vectors of correlated random variables, e.g. *n* random vectors  $\mathbf{s}_1$ ,  $\mathbf{s}_2$ ,...,  $\mathbf{s}_n$ , each of dimension *d*. These vectors arise from taking measurements or characteristics on *d* variables for each of *n* objects. The vectors may or may not come from the same probability distribution and the variables  $s_{ij}$ within each  $\mathbf{s}_i = [s_{i1}, s_{i2}, ..., s_{id}]$  will generally be correlated. The variables may be quantitative (discrete or continuous) or qualitative (ordered or unordered).

For example, let **S** be the collection of *n* data sets  $\mathbf{s}_i = [s_{i1}, s_{i2}, ..., s_{ij}]$ , i = 1 : n, j = 1 : d, i.e. *d* variables are obtained for each of the *n* objects. It is convenient to express the data base in matrix **S** with random variables  $s_{ij}$ , namely,

$$\mathbf{S} = \begin{bmatrix} s_{11} \ s_{12} \ \cdots \ s_{1d} \\ s_{21} \ s_{22} \ \cdots \ s_{2d} \\ \vdots \ \vdots \ \cdots \ \vdots \\ s_{n1} \ s_{n2} \ \cdots \ s_{nd} \end{bmatrix}.$$
(1)

Figure 1 shows the overall procedure for learning geographic regions by applying filters and classification.

A data set  $\mathbf{s}_i$  of the data base **S**, which includes d = 6 variables, is given as in Table 1. We record the following d = 6



**Fig. 1.** Learning geographic regions by applying filters and classification to the data base in IMS.

variables: User-ID  $\mathbf{s}_{i1}$ ; Date-Time  $\mathbf{s}_{i2}$  (the time stamp of the data collected by the user); CellID  $\mathbf{s}_{i3}$  (the Cell IDs of the users); Location  $\mathbf{s}_{i4} = [\mathbf{x}_{i4}, \mathbf{y}_{i4}]$  (the exact position of the user); FogLamp  $\mathbf{s}_{i5} \in \{\text{``f'', ``n''}\}$ , where  $\mathbf{s}_{i5} = \text{``f''}$  if the fog lamp is turned on and  $\mathbf{s}_{i5} = \text{``n''}$  otherwise; RainSensor  $\mathbf{s}_{i6} \in \{\text{``r'', ``n''}\}$ , where  $\mathbf{s}_{i6} = \text{``r''}$  if the rain sensor has turned on the windshield wiper automatically and  $\mathbf{s}_{i6} = \text{``n''}$  otherwise. Here  $\mathbf{s}_{i1}, \mathbf{s}_{i2}, \mathbf{s}_{i3}$  and  $\mathbf{s}_{i4}$  are discrete quantitative variables, while  $\mathbf{s}_{i5}$  and  $\mathbf{s}_{i6}$  are unordered qualitative variables.

Because we are interested in vehicles, it is assumed that only the data sets from vehicles are analyzed. Furthermore, only the data sets for a certain period of time are taken. After that, a data base remains, to which we apply classification methods.

### 2.1 Type filter

Based on the different information in the variables, the data sets can be divided into different groups (classes). This procedure is defined as Type Filter (TF).

Suppose the data sets of car-sensor states are divided into groups, which are "fog", "rain" and "none", denoted as "f", "r" and "n" in the following. On the basis of n data sets, we wish to assign the position of cars to three states (fog/rain/none).

The well-defined clusters are obtained by type filter which is realized by accessing only a part of the whole data set:

$$g_{f} = \{ [\mathbf{s}_{i4}, \mathbf{s}_{i5}] | \mathbf{s}_{i5} = "f" \} 
 g_{r} = \{ [\mathbf{s}_{i4}, \mathbf{s}_{i6}] | \mathbf{s}_{i6} = "r" \} 
 g_{o} = \{ [\mathbf{s}_{i4}, \mathbf{s}_{i6} (\text{or } \mathbf{s}_{i5})] | \mathbf{s}_{i5} = "n" \land \mathbf{s}_{i6} = "n" \}$$
(2)

where  $\mathbf{g}_{f}$  is the group in which all fog lamps are on;  $\mathbf{g}_{r}$  is the group in which all rain sensors are on; and in group  $\mathbf{g}_{o}$ , neither the fog lamps nor the rain sensors are on. Note that both fog and rain hardly happen at the same time, so it will not be considered here.

In the following example, there are n = 7 users. The information of the data base required here is a subset of the data

base S (the 4th, 5th and 6th column of S) and is presented in matrix  $S_s$ :

$$\mathbf{S}_{\mathbf{s}} = \begin{bmatrix} \mathbf{s}_{i4} \ \mathbf{s}_{i5} \ \mathbf{s}_{i6} \end{bmatrix}$$

$$= \begin{bmatrix} \begin{bmatrix} x_{1}, y_{1} \end{bmatrix} \mathbf{n} \ \mathbf{n} \\ \begin{bmatrix} x_{2}, y_{2} \end{bmatrix} \mathbf{f} \ \mathbf{n} \\ \begin{bmatrix} x_{3}, y_{3} \end{bmatrix} \mathbf{n} \mathbf{r} \\ \begin{bmatrix} x_{4}, y_{4} \end{bmatrix} \mathbf{n} \mathbf{r} \\ \begin{bmatrix} x_{5}, y_{5} \end{bmatrix} \mathbf{f} \ \mathbf{n} \\ \begin{bmatrix} x_{6}, y_{6} \end{bmatrix} \mathbf{n} \ \mathbf{n} \\ \begin{bmatrix} x_{7}, y_{7} \end{bmatrix} \mathbf{f} \ \mathbf{n} \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} \begin{bmatrix} x_{1}, y_{1} \end{bmatrix} \mathbf{n} \\ \begin{bmatrix} x_{2}, y_{2} \end{bmatrix} \mathbf{f} \\ \begin{bmatrix} x_{3}, y_{3} \end{bmatrix} \mathbf{r} \\ \begin{bmatrix} x_{4}, y_{4} \end{bmatrix} \mathbf{r} \\ \begin{bmatrix} x_{5}, y_{5} \end{bmatrix} \mathbf{f} \\ \begin{bmatrix} x_{6}, y_{6} \end{bmatrix} \mathbf{n} \\ \begin{bmatrix} x_{7}, y_{7} \end{bmatrix} \mathbf{f} \ \mathbf{n} \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} x_{1}, y_{1} \end{bmatrix} \mathbf{n} \\ \begin{bmatrix} x_{1}, y_{1} \end{bmatrix} \mathbf{n} \\ \begin{bmatrix} x_{2}, y_{2} \end{bmatrix} \mathbf{f} \\ \begin{bmatrix} x_{3}, y_{3} \end{bmatrix} \mathbf{r} \\ \begin{bmatrix} x_{4}, y_{4} \end{bmatrix} \mathbf{r} \\ \begin{bmatrix} x_{5}, y_{5} \end{bmatrix} \mathbf{f} \\ \begin{bmatrix} x_{6}, y_{6} \end{bmatrix} \mathbf{n} \\ \begin{bmatrix} x_{7}, y_{7} \end{bmatrix} \mathbf{f} \end{bmatrix} = \mathbf{S}_{\mathbf{c}}$$
(3)

After applying Eq. (2) to matrix **S**, the second column  $\mathbf{s}_{i5}$  and third column  $\mathbf{s}_{i6}$  of matrix **S** are rearranged into one column resulting in matrix  $\mathbf{S}_{\mathbf{c}}$  in Eq. (3). The position of the users, which form the training data, is the first column of matrix  $\mathbf{S}_{\mathbf{c}}$ :

$$[\mathbf{x}_{\mathbf{c}}, \mathbf{y}_{\mathbf{c}}] = \mathbf{S}_{\mathbf{c}}(:, 1). \tag{4}$$

The group vector is the second column of matrix  $S_c$ :

$$\mathbf{g}_{\mathbf{c}} = \mathbf{S}_{\mathbf{c}}(:, 2). \tag{5}$$

#### 2.2 Classification

Classification is done by discriminant analysis using as input data: the  $n \times 2$  matrix  $[\mathbf{x}_c, \mathbf{y}_c]$  as training data, the vector  $\mathbf{g}_c$  as group vector, and the sample data which are all the possible positions in the region. In our example, training data  $[\mathbf{x}_c, \mathbf{y}_c]$  is a  $7 \times 2$  (n = 7) matrix and group vector  $\mathbf{g}_c$  is a  $7 \times 1$  (n = 7) vector. A quadratic discrimination function (Seber, 2004) is used to separate different swarms of points in the region.

The number of groups will be calculated and the matched indices of group members are also determined. Suppose there are *K* groups, where  $n_k$  is the number of elements in group *k*. Let  $[x_1 \ y_1]...[x_{n_k} \ y_{n_k}]$  be the subset of the training data belonging to group *k*. For every group we compute the mean value of the training data:

$$\left[\overline{x}_{k} \ \overline{y}_{k}\right] = \frac{1}{n_{k} - 1} \sum_{i=1}^{n_{k}} [x_{i} \ y_{i}]$$

$$(6)$$

and the covariance matrix of the training data:

$$\mathbf{S}_{k} = \frac{1}{n_{k} - 1} \sum_{i=1}^{n_{k}} ([x_{i} \ y_{i}] - [\overline{x}_{k} \ \overline{y}_{k}])^{T} ([x_{i} \ y_{i}] - [\overline{x}_{k} \ \overline{y}_{k}]).$$
(7)

With  $|\mathbf{S}_k|$  being the determinant of  $\mathbf{S}_k$  we can compute the following parameters for each group:

$$H_{k} = \frac{1}{2} \left( \left[ \overline{x}_{k} \ \overline{y}_{k} \right] \mathbf{S}_{k}^{-1} \left[ \overline{x}_{k} \ \overline{y}_{k} \right]^{T} + \log |\mathbf{S}_{k}| \right), \tag{8}$$

$$\mathbf{r}_{k} = \mathbf{S}_{k}^{-1} \left[ \overline{\mathbf{x}}_{k} \ \overline{\mathbf{y}}_{k} \right]^{T}, \tag{9}$$

$$\mathbf{Q}_k = -\frac{1}{2}\mathbf{S}_k^{-1}.\tag{10}$$



Fig. 2. Using filters for warning service.

Now with  $\mathbf{v} = [x_s \ y_s]^T$  be the sample data in the considered region (local map), we can compute the quadratic discriminant function between group *k* and  $\ell$  as follows:

$$f_{k,\ell}(x_s, y_s) = (H_k - H_\ell) + \mathbf{v}^T [\mathbf{r}_k - \mathbf{r}_\ell] + \mathbf{v}^T [\mathbf{Q}_k - \mathbf{Q}_\ell] \mathbf{v}.$$
(11)

The positions  $[x_s \ y_s]^T$  for which  $f_{k,\ell}(x_s, y_s) = 0$  determine the border between groups k and  $\ell$ . Once we have done the discriminant analysis for all possible pairs of groups  $(k, \ell)$ , the geographic regions are successfully learned (see Fig. 3 for an example) and a location filter can be defined.

#### 2.3 Location filter

The learned region of each group can be used to define a Location Filter (LF) (Mahy, 2007). Depending on different LFs in various regions, IMS sends different warnings to the users who are in these areas (see Fig. 2). Cars in the fog or rain area, which have not turned on their fog lamp/windshield wiper, will get a warning from IMS. The warning (fog/rain-warning) service from the IMS network will also be demanded by users who are going to enter that LF-region for safe-driving support. This service helps the IMS clients to notice the geographic surrounding immediately, so that they can have enough time to prepare to upcoming situations.

#### **3** Implementation and experimental results

A desktop-class machine equipped with an Intel<sup>®</sup> Core<sup>TM</sup> 2 Quad Q6600 CPU with 2.4 GHz, 4 GB of memory and a gigabit ethernet controller is used. Besides the software specific to the individual tests, this computer runs a nameserver for resolution of the IMS domain and a MySQL data base. Fraunhofer's OpenIMSCore (OpencourceIMS, 2007) is implemented as IMS server. The server machine runs all of the Call Session Control Functions (CSCFs) (Poikselkae et al., 2006) in parallel. Clients for all scenarios are laptops running



Fig. 3. Learned geographic regions using deterministic analysis.

pjsua, a reference implementation of PJSIP, an open source SIP library. The experiments are done by running several clients on the laptops, which send their data (location, sensor states, etc.) to the data base. The black circle on the road means that the car devices are not on. Red cross means that the fog lamp is on, and the green triangle means that the rain sensor is on. Because in the fog area, the cars drive more slowly than in the normal weather areas, the density of the cars in the fog area is higher than in good weather area. The same holds for the rain area. TFs divide the positions into different groups (fog, rain, none). Then the CAS can determine the respective geographic regions. The result of the discriminant analysis for a specific example is shown in Fig. 3. Inside the pink line ellipse is the fog region, while the area inside the green dotted line ellipse is the rain region. From now on, the geographic regions are learned and LFs can be generated.

#### 4 Conclusion

This paper presents how to deploy classification method to detect geographic regions using LBS in IMS. Exchanging the user information (data sets) between Local Network and Global IP Network through the IMS kernel sets up the data base in IMS network. TF divides all IMS clients into groups due to different variables in users' data sets. Classification method using discriminant analysis together with location server in IMS determines the position of the user group depending on the location of each member in every group. After that, the geographic regions are learned by IMS and LF can be defined. Depending on LF, different services (e.g. warning service) can be established for the users in different positions.

The assumption that the variables of the groups are normally distributed seems reasonable for fog/rain areas (i.e. in the middle of the region it is more likely that the devices are turned on). However for other events other distributions might be more reasonable, and more sophisticated methods yield better results (e.g. SVM (Abe, 2005)). One important issue of future research is the dynamics of the geographic regions (fog areas move over time, entrance/exit of users in geographic regions). Overall we believe that in future vehicle communication systems, where location based services gain more and more importance for many applications, classification of geographic regions of events and the deduced services will play an important role.

#### References

- Abe, S.: Support Vector Machines for Pattern Classification, Springer in series Advances in Pattern Recognition, Springer, 2005.
- Djuknic, G. M. and Richton, R. E.: Geolocation and Assisted GPS, IEEE Computer Society, 34, 123–125, 2002.
- Drane, C., Macnaughtan, M., and Scott, C.: Positioning GSM telephones, Communications Magazine, IEEE, 36, 46–54, 1998.
- He, Y., Hueske, K., Gotze, J., and Coersmeier, E.: Efficient Computation of Joint Direction-Of-Arrival and Frequency Estimation, IEEE Signal Processing and Information Technology, IEEE International Symposium, pp. 144–149, December 2008.
- Khlifi, H. and Gregoire, J.: IMS for Enterprises, IEEE Communications Magazine, 45, 68–75, 2007.
- Liu, G., Zhang, J., Zhang, P., Wang, Y., Liu, X., Li, S.: Evolution Map from TD-SCDMA to FuTURE B3G TDD, IEEE Communications Magazine, 44, 54–61, 2006.
- Mahy, R.: A Document Format for Filtering and Reporting Location Notifications in PID-LO, Current Internet Draft <draft-ietfgeopriv-loc-filters-01.txt> (work in progress), March 2007.
- McLachla, G. J.: Discriminant Analysis and Statistical Pattern Recognition, Wiley series in probability and statistics, John Wiley and Sons Ltd., 2004.
- Monsmondo, M., Kapov, L., and Kovacic, M.: Bringing Location Based Services to IP Multimedia Subsystem, IEEE Electrotechnical Conference Melecon 2006, pp. 746–749, May 2006.
- Fraunhofer FOKUS NGNI: OpensourceIMS; http://www. openimscore.org/, 2007.
- Perusco, L. and Michael, K.: Control, Trust, Privacy, and Security: Evaluating Location-Based Services, IEEE Technology and society magazine, 26, 4–16, 2007.
- Poikselkae, M., Khartabil, H., and Niemi, A.: IMS IP Multimedia Concepts and Services, Wiley Series in Mobile and Wireless Communications, John Wiley and Sons Ltd., 2006.
- Seber, G. A. F.: Multivariate Observations, Wiley series in Multivariate analysis, John Wiley and Sons Ltd., 2004.
- Varshney, U.: Vehicular Mobile Commerce, IEEE Computer Society, 37, 116–118, 2004.
- Vingarzan, D. and Weik, P.: Experiments Using the Open IMS Core, IEEE Vehicular Technology Magazine, 2, 28–34, 2007.