(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 July 2013 (25.07.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/107669 A1

(51) International Patent Classification:
H04W 4/02 (2009.01)

(21) International Application Number:
PCT/EP2013/050225

(22) International Filing Date:
8 January 2013 (08.01.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
P201230075 20 January 2012 (20.01.2012) ES

(71) Applicant: **TELEFÓNICA, S.A.** [ES/ES]; Gran Vía, 28, E-28013 Madrid (ES).

(72) Inventors: **LARA, Ruben**; Telefónica, S.a., Gran Vía, 28, E-28013 Madrid (ES). **MARTINEZ, Rocio**; Telefónica, S.a., Gran Vía, 28, E-28013 Madrid (ES). **RODRIGUEZ, Miguel Ángel**; Telefónica, S.a., Gran Vía, 28, E-28013 Madrid (ES). **MARTÍN, Alberto**; Telefónica, S.a., Gran Vía, 28, E-28013 Madrid (ES). **ARMENTA, Ana**; Telefónica, S.a., Gran Vía, 28, E-28013 Madrid (ES).

(74) Agent: **GONZÁLEZ - ALBERTO, Natalia**; GARRIGUES IP, S.L.P, Hermosilla, 3, E-28001 Madrid (ES).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: A METHOD FOR THE AUTOMATIC DETECTION AND LABELLING OF USER POINT OF INTEREST

(57) Abstract: The method comprises acquiring information from signals exchanged between a user's mobile computing devices and a plurality of Base Transceiver Stations, or BTSs, analyzing said acquired information for determining, over a period of time, the locations of said user's mobile computing device and deduce through a statistical model the points of interest, identifying and labelling at least part of said user's mobile computing device determined locations as points of interest.

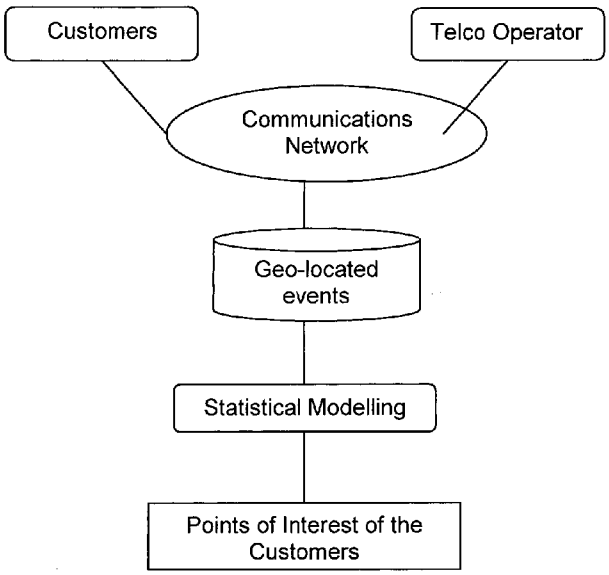


Figure 1

WO 2013/107669 A1

Published:

— *with international search report (Art. 21(3))*

A method for the automatic detection and labelling of user point of interest

Field of the art

The present invention generally relates to a method to automatically detect and label one or more points of interest (PoIs) of a user of mobile phone services, said method based exclusively on geo-located phone usage information and with no customer interaction. Based on the geo-located usage events generated in the telecommunications operator network and using statistical methods, the invention allows the identification of, from all the locations visited by the user, the locations that are most relevant to him: his PoIs. Moreover, the invention automatically assigns labels to the PoIs detected, thereby attaching meaning to such locations.

Background of the invention

The study of human mobility patterns has received growing attention over the past few years, especially due to the increasing availability of location data coming from both global positioning systems (GPS) and mobile telephone usage, which leaves geo-located traces on the operators' networks.

Understanding how and when human movements take place across towns, cities or countries is of interest in many areas, such as traffic management, transport network design or diseases spread control. However, not only a global view of population flows, but also individual mobility patterns of a user, are of great interest in a number of fields. The knowledge of what locations a user periodically visits, over what period, with what frequency, what days of the week and at what times of the day, etc. can be exploited for the provision of contextual services, relevant advertising, targeted offers to address the particular mobility needs of the user, itinerary planning... In general, knowing locations that are relevant for a user can enable the personalisation of commercial communications or service interactions and improve their relevance.

In order to estimate PoIs needs to be assumed that human movements follow some pattern and, thus, the location of a user is to some extent predictable. In this sense, several authors have recently worked on the predictability of human mobility patterns trying to find the limits of such predictability. Based on the study of the trajectory of 100,000 anonymised mobile phone users whose position was tracked for a six-month period, it was found in [1] that human trajectories show a high degree of

temporal and spatial regularity, each individual being characterized by a time-independent characteristic travel distance and a significant probability to return to some few highly frequented locations.

Reference [2] tried to answer the question of "to what degree is human behavior predictable?" by studying the mobility patterns of anonymised mobile phone users. The authors measured the entropy of each individual's trajectory, and found a 93% potential predictability in user mobility across the whole user base. They also found a remarkable lack of variability in predictability, largely independent of the distance users cover on a regular basis.

Location prediction models developed in the recent years take into account individual and collective behaviours. As for example [3] , where a model was based on the person's past trajectory and the geographical features of the area where the collectivity moved, both in terms of land use, points of interests and distance of trips.

The prediction method can be addressed in different ways. [4] for example presents a location and dwell time prediction using kernel density estimation based on communication, proximity, location, and activity information of the subjects.

Collective communication behavior has also been used to detect the occurrence of anomalous events as in [5] where they study how spatiotemporal anomalies can be described using standard percolation theory tools.

Human mobility patterns have also been inferred from GPS traces [6] where a clustering method is proposed to extract the main points of interest, called geo-locations, from GPS data. Starting from geo-locations they propose a definition of community, the geo-community, which captures the relation between a spatial description of human movements and the social context where users live. A statistical analysis of the principal characteristics of human walks provides the fitting distributions of distances covered by people inside a geo-location and among geo-locations and pause time. They also analyze factors influencing people when choosing successive locations in their movement.

Reference [7] discloses study mobility in an activity-aware map that describes the most likely activity associated with a specific area of space. This allows them to capture the individual daily activity pattern and analyze the correlations among different people's work area's profile. For this purpose they understand the work location of each user as the most frequent stop during day hours. Based on a large mobile phone data of nearly one million records of the users in the central Metro-Boston area, they find a strong correlation in daily activity patterns within the group of people who share a

common work area's profile. In addition, within the group itself, the similarity in activity patterns decreases as their work places become apart

Some authors [8] focus on Human Centered Mobility Models, that is, how social networks and mobility patterns match together, trying to extend the social network mobility to the geographical movements, which they call "opportunistic networks".

So, the scientific community is working and researching around human mobility as a topic, and several approaches yield to better knowledge of different aspects of human mobility, but from the perspective of its exploitation for personalisation, what matters most to us is to know when and for what purpose users frequently visit certain locations, that is, what are the points of interest of users.

Different inventions work on the topic of PoIs for different purposes:

The document US2010121803 "Predictive ephemeral Points of Interest" wireless application users are provided with the ability to record locations and retrieve maps of past locations and predicted future locations of specific interest. Within this invention, to predict a location, data about previous reported locations are gathered, and statistical analysis is used to present a visual guide to finding the PoIs at a particular time in the future.

The document WO2011076988 "Methods and apparatus for grouping points of interest according to area names" an approach is provided for crowd sourcing and grouping point-of-interest based on cell broadcast message information. Reception of a message from a mobile terminal is caused, at least in part. The message specifies point-of-interest information and an associated area name corresponding to one of a plurality of cells of a communication network. The message is parsed to determine the point-of-interest information and the associated area name.

The document WO2011072882 a method for evaluating an attribute of a point of interest comprises associating a region with the point of interest and evaluating the attribute according to a comparison of position data of a plurality of users with position data defining the associated region.

The document US2011166957 "Biasing of search result clustering to ensure more effective point of interest targeting": Directory service results responsive to a request for a desired good or service provider may be provided based on one or more user-selected locations. The user may seek a desired good or service provider that is close to a location from which the user may begin travelling to the point of interest,

referred to as a source location, and satisfies a beneficial objective by the user (a directional travel preference for instance).

The document US2010023259 "Discovering points of interest from users map annotations" A method that facilitates generating a point of interest related to a map. An interface component can collect a portion of annotation data from two or more users, wherein the portion of annotation data is associated with a digital map and includes at least one of a map location and a user specific description of the map location. An annotation aggregator can evaluate annotation data corresponding to the map location on the digital map. The annotation aggregator can create a point of interest for the map location based upon the evaluation and populates the digital map with at least one of an identified location extracted from two or more users.

The document US2003191578 "Method and System for providing reminders about points of interests while travelling" A navigation system includes a feature that allows a user to specify a type of point of interest and then receive a reminder when the user is in proximity to a location of the point of interest of the specified type while travelling in a geographic region.

The document US2009097710 "Methods and system for communication and displaying points-of-interest" A method for displaying point-of-interest coordinate locations in perspective images and for coordinate-based information transfer.

The document US2008076451 "Point of Interest Spatial Rating Search Method and System" A system and method for searching and retrieving location information associated with one or more points of interests, whereby the search criteria can be dependent upon the location of a point of interest with respect to the real-time position of the user, and any preferences or search restrictions selected by the user, such as rating information about the point of interest.

The document US7890254 "Point of Interest Display System" A point of interest display system includes an updateable database which interfaces with a microprocessor which receives data from a GPS receiver providing the system with current vehicle location and direction of travel information.

The document US2004236504 "Vehicle Navigation Point Of Interest" The present invention provides a navigation system for assisting locating points of interest during vehicle navigation. The system includes a processor enabled by software to receive and store a user selection of preferred visitation points and a user selected time and to determine and indicate a subset of the preferred visitation points that are

located within a predetermined location relative to the vehicle position at the user selected time or relative to a selected destination.

The document US2011125359 "Navigation Apparatus, Server Apparatus and Method of Providing Point of Interest Data" A navigation apparatus includes a communications interface for communicating data via a communications network and a processing resource coupled to the interface and arranged to receive a request for point

of interest information, and to communicate via the communications interface a message constituting a point of interest data request for receipt by a remote server.

The document WO2011072745 "Dynamic Point of Interest Suggestion" A system, method and device for recommending a POI via a navigation device that includes receiving a recommendation of a POI from a third party at a server and determining information related to the POI. The determined information is correlated to with data related to navigation devices associated with the server. Navigation devices are selected to receive the third party recommended POI based on results of the correlation of the determined information with the data related to the navigation devices and the recommended POI is forwarded from the server to a targeted navigation device based on results of the correlation of the determined information with the data related to the navigation devices.

The document EP1939797 "Method and apparatus for automatically determining a semantic classification of context data" A method for automatically determining a semantic classification for context data obtained by a mobile device, said method comprising sampling by said mobile device one or more context data streams over time; applying a clustering algorithm to identify one or more clusters in the sampled context data; running a logic engine to automatically determine a concept name from a set of predefined concept names as a semantic classification of said one or more clusters; assigning said concept name to said one or more cluster or suggesting said assignment to the user.

Problems with existing solutions

As seen, *Points of Interest* is a term that a number of existing works refer to. Most of them use GPS data, and information about Pols is gathered from particular users, proposed and displayed, or routes between given Pols are computed. But in all the works where Pols are said to be computed, relevant locations have to be given either by the user, or by a navigation system.

For instance, inventions US2009097710 and US7890254 work respectively on a method and a system for displaying point of interest. US2010121803 gathers data about previously reported locations and uses statistical analysis to present a visual guide to finding PoIs at a particular time in the future, US2004236504 and WO2011072745 work on a dynamic point of interest suggestion for navigations devices, and US2011125359 consists of a point of interest data request remote system.

Based on the evaluation of user annotated data US2010023259 creates points of interests for map locations, and according to the user location US2008076451 provides a point of interest searching method while US2003191578 generates a reminder when the user is in the proximity of the point of interest. WO2011072882 also works with the location information of a point of interest and by comparing it to an associated region evaluates an attribute of the point of interest itself.

Classification techniques are applied by US2011166957 to cluster and so ensure more effective point of interest targeting among user selected locations, by WO2011076988 to group points of interest according to area names, and also by EP1939797 which provides a method for automatically determining a semantic classification for context data obtained by a mobile device.

Several references as [2][3][4][5] also work with Call Detail Records (CDRs) or GPS data [6] and build predictive models, but these are usually isolated experiments carried for a specific town and do not comprise a fully automatic detection and labelling of the most relevant locations for a user.

It has also found many inventions that are in some way related to PoIs. Some of them gather information directly from the users, others record the previously visited locations (via GPS), and many are focused on POI display systems for GPS navigation.

Summary of the Invention

It is necessary to offer an alternative to the state of the art which covers the gaps found therein, particularly related to the lack of proposals which really present an efficient method to detect and label automatically and in a non-intrusive way the PoIs of user mobile services.

To that end, the present invention provides a method to detect and label PoIs, said method based exclusively on geo-located phone usage information and with no customer interaction.

The method of the invention comprises:

a) acquiring information from signals exchanged between a user's mobile computing devices and a plurality of Base Transceiver Stations, or BTSs;

b) analyzing said acquired information for determining, over a period of time, the locations of said user's mobile computing device based on the locations of the BTSs with which said signals exchange has occurred; and

c) detecting and labelling at least part of said user's mobile computing device determined locations as points of interest at least on the basis of the number of times said user's mobile computing device has been at said determined locations over said period of time.

wherein said steps b) and c) comprises applying said analysis and identification through a statistical model.

This statistical model according to one embodiment comprises a Partitioning Around Medoids, or PAM, clustering algorithm based on a Pearson distance.

The mentioned clustering algorithm returns twenty different representations of clusters.

The representations of clusters are represented by its medoid curve and are labelled considering social habits and cultural characteristics of the region under study.

Other features of the method of the invention are described according to appended claims 2 to 10, and in a subsequent section related to the detailed description of several embodiments.

Brief Description of the Drawings

The previous and other advantages and features will be more fully understood from the following detailed description of embodiments, with reference to the attached drawings (some of which have already been described in the Prior State of the Art section), which must be considered in an illustrative and non-limiting manner, in which:

Figure 1 shows current scheme used for the location of the points of interest.

Figure 2 shows current scheme used for the location of the points of interest worked on geo-located information.

Figure 3 shows the different processes to be performed to estimate the statistical model.

Figure 4 shows an example of use cases and applications that could be performed knowing users' points of interest, according to an embodiment of the present invention.

Figure 5 shows a possible inclusion of the invention to third parties for the correct customization of their marketing campaigns or commercial activities, according to an embodiment of the present invention.

Figure 6 shows an example of several BTS usage vector, according to an embodiment of the present invention.

Figure 7 shows an example of set of clusters, each one represented by its medoid curve, according to an embodiment of the present invention.

Detailed Description of Several Embodiments

The invention works on geo-located events obtained from the operator's network in a non-intrusive way for a given period of time, capturing any geo-located traces that mobile phone usage leave on the network, from signalling (data sessions attach/detach events, initiation and termination of voice calls, SMSs sent, etc). This acquired information, i.e. geo-located events must contain at least the following information:

-The number associated to the event
-The date and time for the event
-The BTS associated to the event

When talking about voice calls, the CDRs contain (for every customer of the operator):

-The number making the call
-The number receiving the call
-Date and time of the call
-Duration of the call
-The BTS where the call begun
-The BTS where the call ended

The invention consists of a series of processes that based on geo-located information lead to the points of interest of the customers that generated that geo-located information

1. Gathering of Location Information

By processing a certain amount of signalling traces (for instance CDRs) events are obtained (in the case of CDRs, calls) related to every customer over the period of

time covered, and as we have the event date, time and BTS, we also obtain the number of events for every BTS the customer has been using over the days of the period.

$$Events(custX) = \{BTS, date\}$$

$$date - date(t)$$

A BTS identifier is linked to a geographical point, more precisely to the geographical area covered by the Base Transceiver Station. So, having the count of events for every BTS means that you have the count of events for every location the customer visited (every location where the customer did or received a call in the case of CDRs).

$$BTS = BTS(x, y)$$

$$Events(custX) = Events[cust, BTS(x, y), date(t)] = Events(cust, x, y, t)$$

2. Customer Communication Filters

This method uses statistical methods to build the Mobility Models that among other questions extract the more characteristic communication per BTS patterns across the customer base. In order to allow reliable and extendible statistical models, we filter out the cases where communication behaviour seems too extreme to be modelled, this is, customers that talk too much or too few (in general, across every location they visit), or events that might not be modelled because do not respond to a common BTS usage pattern.

This is only considering the cases where the overall communication (number of events in general) is limited by a lower and an upper threshold:

$$T_L < \sum_{bts} \sum_{t} Nevents(custX) < T_U$$

3. BTS Usage Filter

For each customer it is also filtered out every BTS that does not reach a representative communication threshold T_R. This threshold can be expressed as an absolute quantity of geo-located events, as a percentage of the customer geo-located

events, or as a combination of both. For the case of a threshold given by a percentage:

$$\sum_{t} Nevents(custX, btsY) \quad > \quad \mathbf{TR} * \sum_{bts} \sum_{t} Nevents(custX)$$

4. Relevant BTS

The Customer-BTS pairs that remain after the two filtering phases are what we call the "relevant BTSs". For every customer we have a set of BTSs that represent the locations where the customer communicates (or registers any kind of events) at least the necessary to be modelled.

Cust1 →{BTS11,BTS21,....BTSn1}

Cust2 →{BTS12,BTS22,....BTSn2}

.....

Custm →{BTS1m,BTS2m,....BTSnm}

Mobility Analysis Models

In order to better explain the Mobility Analysis Models a zoom-in in those components is made. The numbered boxes are explained in the following paragraphs.

5. Customer- BTS Usage Vectors

For every "Customer-Relevant BTS" couple, a vector containing the number of registered positions is built (calls or in general, any kind of geo-located events N_{ge}) of each of the customers at each of those BTS for every hour of the days of the week. But as not every week day has the same meaning in terms of life activity patterns, we group Monday, Tuesday, Wednesday and Thursday while Friday, Saturday and Sunday remain separated.

{Cust1,BTS11}→{Nge_{mt00},...,Nge_{mt23}, Nge_{fr00},...,Nge_{fr23}, Nge_{st00},...,Nge_{st23}, Nge_{sn00},...,Nge_{sn23}}

{Cust1,BTS21}→{Nge_{mt00},...,Nge_{mt23}, Nge_{fr00},...,Nge_{fr23}, Nge_{st00},...,Nge_{st23}, Nge_{sn00},...,Nge_{sn23}}

.....

{Cust1,BTSn1}→{Nge_{mt00},...,Nge_{mt23}, Nge_{fr00},...,Nge_{fr23}, Nge_{st00},...,Nge_{st23}, Nge_{sn00},...,Nge_{sn23}}

So, for every customer several curves are obtained, as many as relevant BTS's that gather the communication pattern of that customer over its representative BTSs across the four different kinds of days. It is what we call the BTS usage vectors.

Figure 6 shows several BTS usage vector examples. They refer to customer X, and contain the aggregated count of calls that the customer X made or received through bts1, bts2 and bts3 respectively at the 24 hour intervals from Monday to Thursday (mt00-mt23), on Fridays (fr00-fr23), Saturdays (st00-st23) and Sundays (sn00-sn23).

6. Normalizations

A first normalization is done dividing every value by the number of days of the corresponding type present in the period of time under consideration. This normalization allows us to compare the 4 different parts of the curves among them:

{Cust1,BTS1}' = {Cust1,BTS1}/ {Nmt,Nfr,Nst,Nsn}

where

Nmt is the number of Mondays, Tuesdays, Wednesdays and Thursdays over the period of time considered

Nfr is the number of Fridays over the period of time considered

Nst is the number of Saturdays over the period of time considered

Nsn is the number of Sunday over the period of time considered

{Cust1,BTS1}' ={ Nge_{mt00}/Nmt ,...,Nge_{mt23}/Nmt,
Nge_{fr00}/Nfr,...,Nge_{fr23}/Nfr,
Nge_{st00}/Nst,...,Nge_{st23}/Nst,
Nge_{sn00}/Nsn,...,Nge_{sn23}/Nsn }

$$\{Cust1,BTS1\}' = \{ N'ge_{mt0},...,N'ge_{mt23},\ N'ge_{fr00},...,N'ge_{fr23},\ N'ge_{st00},...,N'ge_{st23},\ N'ge_{sn00},...,N'ge_{sn23}\}$$

After this first normalization, a second one needs to be made. In order to make possible the comparison between different BTS usage curves with very different mean communication levels, and to be able to focus on the curve shape itself (and not only on the amplitude levels), the curves are also normalized dividing them by the sum of values for every point and thus giving a resultant sum equal to 1:

$$\{Cust1,BTS1\}'' = \{Cust1,BTS1\}' / \sum_{i=0}^{95} N'ge_i = \{Cust1,BTS1\}' / N't$$

$$\{Cust1,BTS1\}'' = \{ N'ge_{mt00}/N't,...,N'ge_{mt23}/N't,\ N'ge_{fr00}/N't,...,N'ge_{fr23}/N't,\ N'ge_{st00}/N't,...,N'ge_{st23}/N't,\ N'ge_{sn00}/N't,...,N'ge_{sn23}/N't \}$$

Where

$$N't = \sum_{i=0}^{95} N'ge_i = \sum_{j=0}^{23} N'ge_{mtj} + \sum_{k=0}^{23} N'ge_{frk} + \sum_{l=0}^{23} N'ge_{stl} + \sum_{m=0}^{23} N'ge_{snm}$$

$N't = \sum_{i=0}^{95} N'ge_i$

7. Sample

A representative sample of the normalized BTS usage vectors is then extracted to feed a non-supervised classification method to identify the main classes of BTS Usage Patterns.

8. Clustering

There are several possible implementations for the classification method; one option is to use a clustering algorithm as for example, a partitioning around medoids (PAM) method based on a Pearson distance.

PAM clustering

The Partitioning Around Medoids (PAM) clustering can be considered a more robust version of the classical k-means approach. It is described in Chapter 2 of the book "Finding Groups in Data: An Introduction to Cluster Analysis", Kaufman & Rousseeuw, 1990.

It has some desirable features:

- It can work directly on a set of data, but it also accepts a dissimilarity matrix of those data as input
- It is more robust than k-means because it minimize a sum of dissimilarities instead of a sum of squared Euclidean distances

The PAM algorithm is based on the search for k representative objects (medoids) among the observations of the dataset. These observations should represent the structure of the data. After finding a set of k medoids, k clusters are constructed by assigning each observation to their closest representative object, based on a given distance. In the case of our invention a Pearson distance is used. A medoid can be defined as that object of a cluster whose average dissimilarity to all the objects in the cluster is minimal.

By default, the initial set of medoids is not specified. The algorithm first looks for a good initial set of medoids (build phase). Then it finds a local minimum for the objective function (swap phase).

Pearson distance

If we have samples of two variables X and Y, it is very common to calculate the sample Pearson correlation coefficient to reveal if there is a linear relationship between the two variables:

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

This coefficient is always between -1 and 1. It is 1 if there is a perfect positive linear relation between the two variables, and is -1 if there is a perfect negative linear relation.

An equivalent expression gives the correlation coefficient as the mean of the products of the standard scores:

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{s_x}\right)\left(\frac{Y_i - \bar{Y}}{s_y}\right)$$

where sx and sy are the sample standard deviation of X and Y.

Pearson correlation can be taken as a measure of similarity between the paired data (Xi, Yi). So we can also obtain a correlation based distance as an expression of the dissimilarity between those data vector:

$$d = 1 - r$$

The greatest value of this distance will be 2 when the vectors can be considered "opposite", and the lowest values will be 0 when the vectors can be considered as having the same shape or profile (if represented in sequence).

The invention uses a PAM clustering method based on a Pearson distance to group the different {Cust, BTSi} curves in several classes following an unsupervised strategy.

The output of the clustering method is a set of classes of BTS usage vectors, each of the classes represented by its medoid. The resulting classes must be as different as possible among them but the vectors belonging to the same class must be as similar as possible. Similarities are considered according to the given distance, in this case the Pearson's one.

9. Clusters, Medoids and Centroids

The clustering process aims to cover as much variability as possible in the sense of detecting many different groups for the Relevant BTS communication patterns. This is why we initially work with a relatively high number of clusters.

For instance, in one of the implementations, the clustering algorithm is forced to return 20 different classes.

Figure 7 shows a set of clusters, each one represented by its medoid curve.

10. Medoid labelling

Once the medoid and centroid are obtained for each class a different label is assigned to them taking into account the social habits and cultural characteristics of the

region under study. In this case the algorithm would give 20 different labels for "level 0" labels (the wider set of labels). Looking at the patterns of the class representatives we later group them into 5 "level 1" labels.

The following text-box shows an example set of level 0 labels created for the BTS-Usage clusters:

1	work (commercial afternoon)
2	nightlife leisure
3	Saturday evening leisure
4	Sunday evening leisure
5	Friday evening leisure
6	work (office, morning)
7	work (commercial)
8	work Mo-Th (lunch time)
9	Sunday late evening leisure
10	work days afternoon shopping
11	home
12	Saturday morning shopping
13	work days evening leisure
14	work (afternoon)
15	Sunday afternoon leisure
16	Saturday afternoon leisure (shopping)
17	Friday afternoon shopping
18	afternoon shopping
19	Friday afternoon shopping (Friday lunch, leaving home)
20	Friday evening leisure

And as previously mentioned, a "level 1" set of labels is created thinking on practical applications that would not need such detail:

1	office work
2	commercial work
3	home / evening leisure
4	night / evening leisure
5	afternoon leisure / shopping

11. Intra-Class Dispersion Measures

The Pearson distance to the cluster representative (medoid) is analyzed for every cluster, to obtain the average and the standard deviation of the distance to the

cluster centre for every classified item. Such values are later used to decide about the POI classification accuracy, to see whether to assign a Point of Interest label or not.

12. Model Results

- The distance
  - A function implementing the distance, needed to be applied every time we want to automatically assign a POI label to a new instance (normalized communication vector)
- Medoid Set:
  - Cluster representatives : 96 position vectors
- Threshold Set
  - Average distance inside each cluster
  - Distance standard deviation
- Label Set (POI labels)
  - Level 0 label sets
  - Level 1 label sets
  - Correspondence table

13. Using the Model

For any Customer-BTS usage vector (present in the sample or not) we assign the "level 0" label of the closest centroid in terms of the same distance that has been used in the clustering process (in this case the Pearson distance).

The equivalent "level 1" label is also assigned based on the knowledge of the "level 0" label and on the correspondence table. For some cases, the instance (customer-BTS usage vector that is being automatically labelled) and the centroid are very similar and the "level 0" labels are a good choice. But this does not always happen, and the level 1 labels represent a more general labelling, with a lower error margin.

One possibility of assigning the labels is to use the value of the average distance and the distance standard deviation to obtain a threshold to decide which of the labels should be used.

For instance, if

Distance({Cust,BTSi},Centroid) > Avg(distance) + stdDev(distance)

then the vector is far from the cluster centre, so it's not reliable enough and we might not want to provide that label. For that case we could for instance return a code that implies that the cluster and label obtained are not reliable enough.

For the cases where

Avg(distance) < Distance({Cust,BTSi},Centroid) < Avg(distance) + stdDev(distance)

It can be said that the vector is nor too far neither too close from the cluster centre. In this case we could consider the assigned cluster is not reliable enough to return the level 0 label and could instead consider only the level1 label.

And for the rest of the cases where

Distance({Cust,BTSi},Centroid) < Mean(distance)

The vector is close enough to the cluster centroid and the level 0 label seems to be an acceptable solution for the automatic point of interest labelling process.

This invention allows a customer to have several POIs with the same label. In some cases, we can modify one of those labels producing new specialized level 1 labels. For instance, a customer may have more than one location labelled as "home". Taking into account other information as the usual area of activity of the customer during week days and week ends, and calculated over the given period of time, one of those POIs can be labelled as "2nd residence". In these cases the initial level 1 labels set is expanded by the addition of new specialized labels.

14. Point of Interest

Finally, as an output of our invention we obtain for every customer:

- A set of BTS that are of special interest for the customer, each one automatically labelled. Those labels explain the particular meaning of the locations for that particular customer.

Example of several embodiments

The knowledge of our customers' points of interest enables a wide variety of use cases and applications.

Knowing the places our customers live, work and prefer for their leisure activities at the different hours of the week allows the operator to develop specific services and applications that exploit such location-segmented information.

That information can also be provided to institutions that might find it useful for any public service planning such us public transport network designing, disease spread control, or other public initiative based on the knowledge of citizens' points of interest.

And other third parties (companies) can also be interested in the use of the points of interest for the correct customization of their marketing campaigns or commercial activities.

Imagine a big textile company that sells different kinds of clothes for the different segments, where segments can be defined by the crossed age and socio-economic level (SEL). That company then is really interested in knowing which locations are of the interest of the teenagers to decide where to locate shops that sell clothes for that segment. Similarly, that company is also very interested in knowing which places are visited through the daily routes of people in their thirties with a probably higher purchasing power. The purchasing power could be correlated with the ARPU (average revenue per user).

So, the information automatically generated by the invention can be sold to 3rd parties or combined with other predictive models inside the operator to support (as in this use case) the shop network designing process (based on the mixed used of the locations of interest of our customers and the more traditional segments as age, SEL or ARPU).

Advantages of the Invention

This invention takes input information that is already available in the usual activity of a telecommunications operator, so no special processes have to be developed in order to obtain it.

This invention allows a fully automatic obtainment of the points of interest of the customers, both individually and at a global and aggregated fashion.

This invention is a non-intrusive method, so the users are not disturbed or bothered from their activity while communicating.

This invention can be easily extended to incorporate new features and can also be extended or applied to any regions.

This invention allows the developments of new and disruptive services that take into account the knowledge of the locations a user goes into, and what they mean for them.

ACRONYMS

POI	Point of Interest
BTS	Base Transceiver Station
CDR	Call Detail Record
GPS	Global Positioning System
SMS	Short Message Service

REFERENCES

[1] González, Hidalgo, Barabási (2008) Understanding individual mobility patterns. Nature, 453, 779-782

[2] Song, C., Qu, Z., Blumm, N. and Barabási, A-L. (2010). Limits of Predictability in Human Mobility. Science, 327, 1018–1021.

[3] Calabrese, Di Lorenzo, Ratti (2010) Human Mobility Prediction based on Individual and Collective Geographical Preferences. ITSC

[4] Firouzi, Liu. Sadrpour (2009) Mobility Pattern Prediction Using Cell-phone Data logs. EECS Final Project Report.

[5] Candia, González, Wang, Schoenharl, Madey, Brabási (2008) Uncovering Individual and Collective Human Dynamics from Mobile Phone Records. Journal of Physics A: Mathematical and Theoretical 41

[6] Zignani, Gaito (2010) Extracting Human Mobility Patterns From GPS-based Traces. 978-1-4244-9229-9/10/$26.00 ©2010 IEEE

[7] Phithakkitnukoon, Horanont, Di Lorenzo, Shibasaki, Ratti (2010) Activity Aware Map: Identifying Human Daily Activity Pattern Using Mobile Phone Data. LNCS 6219, pp. 14–25

[8] Boldrini, Conti, Passarella (2009) The Sociable Traveller: Human Travelling Patterns in Social-Based Mobility. MobiWac'09, October 26–27, 2009, Tenerife, Canary Islands, Spain.

Claims

1. - A method for automatic detection and labelling of user points of interest, comprising:

d) acquiring information from signals exchanged between a user's mobile computing devices and a plurality of Base Transceiver Stations, or BTSs;

e) analyzing said acquired information for determining, over a period of time, the locations of said user's mobile computing device based on the locations of the BTSs with which said signals exchange has occurred; and

f) identifying and labelling at least part of said user's mobile computing device determined locations as points of interest at least on the basis of the number of times said user's mobile computing device has been at said determined locations over said period of time.

wherein said steps b) and c) comprises applying said analysis and identification through a statistical model.

2. - The method of claim 1, further comprising limiting said acquiring information from user's mobile computing device by a lower and an upper threshold.

3. - The method of claim 1, further comprising filtering each of said BTS for each of said user mobile computing device when the communication between them is lower than a threshold.

4. - The method of claim 1 to 3, wherein said acquiring information of step a) further comprising for each couple user- relevant BTS, a vector containing said locations for every hour of the days of the week.

5. - The method of claim 1, wherein said statistical model further comprises a Partitioning Around Medoids, or PAM, clustering algorithm based on a Pearson distance.

6. - The method of claim 5, wherein said clustering algorithm returns twenty different representations of clusters.

7. - The method of claim 6, wherein said representations of clusters are represented by its centroid curve and are labelled considering social habits and cultural characteristics of the region under study.

8. - The method of claim 7, wherein a first set of 20 labels are used to identify points of interest taking into account said habits and said cultural characteristics of the region.

9. - The method of claim 8, wherein from said first set of 20 labels a second set of 5 labels are used to identify said points of interest based on practical applications.

10. - The method of claim 1, wherein said acquiring information of said step a) includes the number of said user mobile computing device, the date and time, and the BTS associated to said signals exchanged.

5

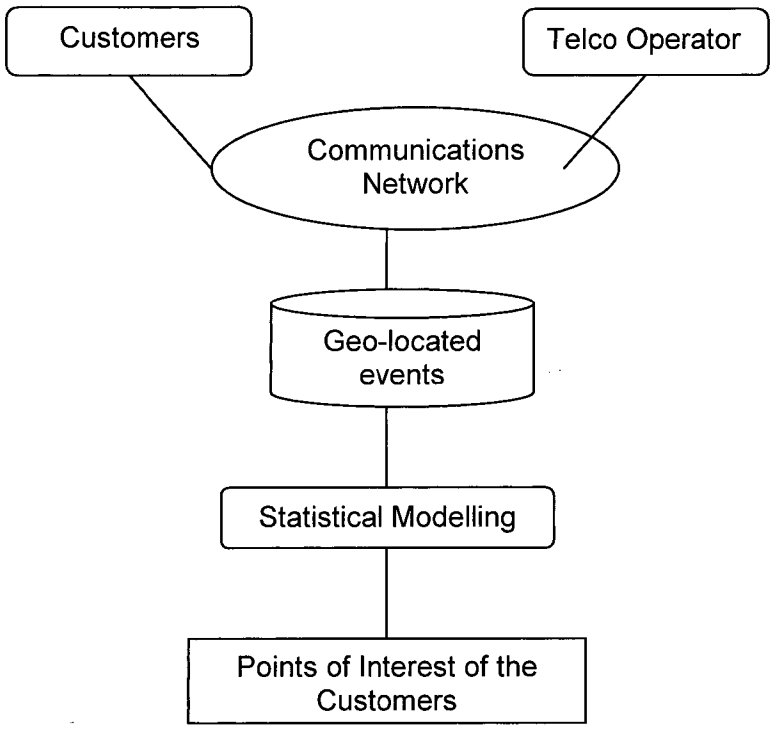

Customers
Telco Operator
Communications Network
Geo-located events
Statistical Modelling
Points of Interest of the Customers

Figure 1

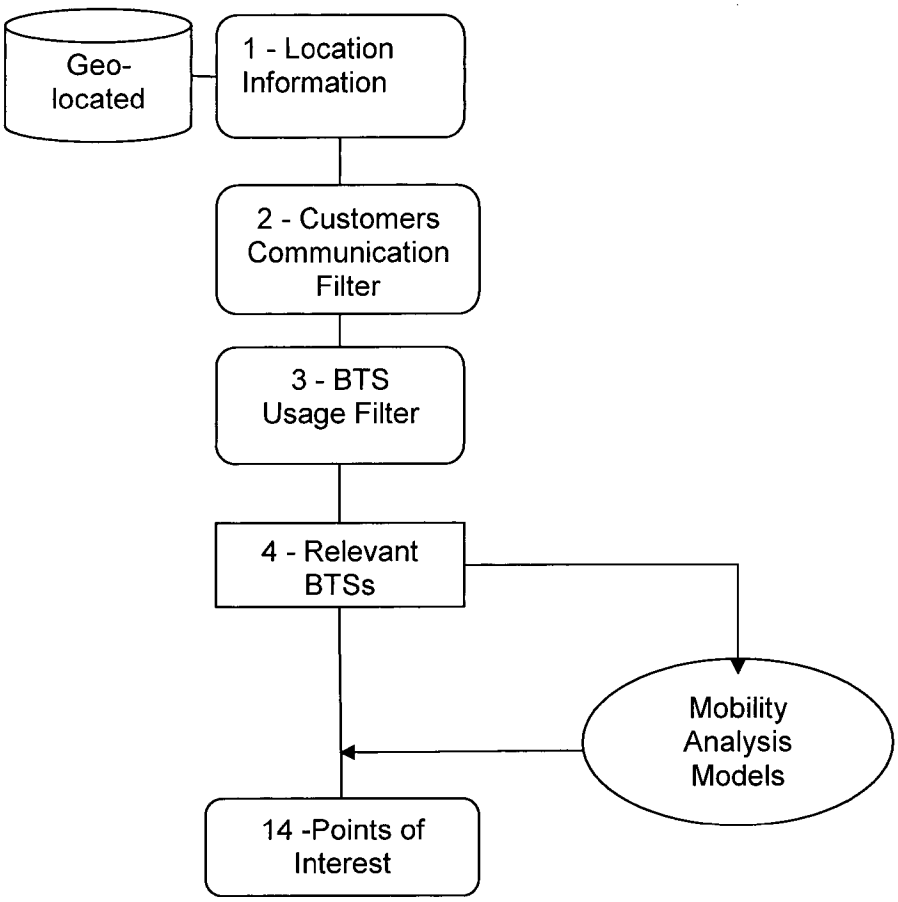

Geo-
located
1 - Location
Information
2 - Customers
Communication
Filter
3 - BTS
Usage Filter
4 - Relevant
BTSs
Mobility
Analysis
Models
14 -Points of
Interest

Figure 2

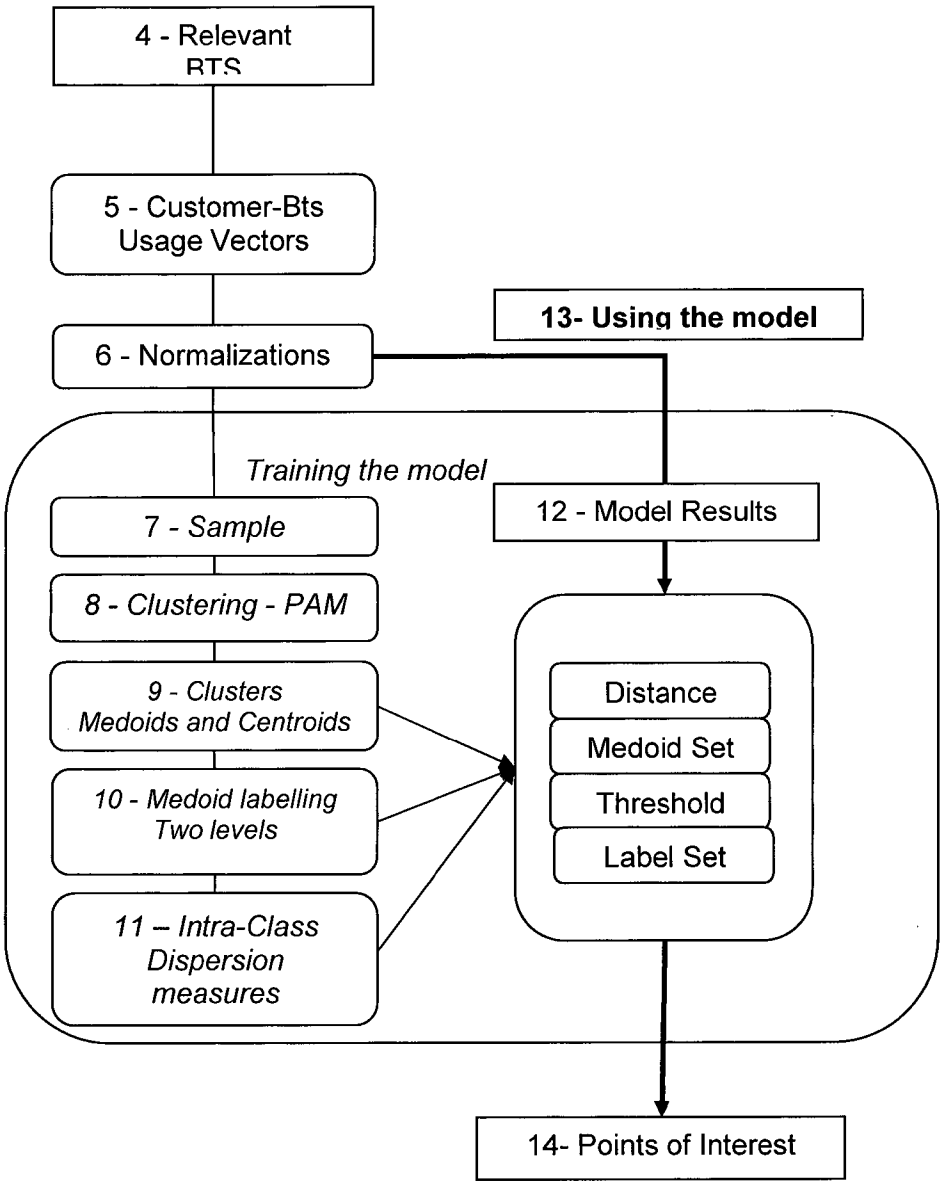

4 - Relevant BTS
5 - Customer-Bts Usage Vectors
6 - Normalizations
13- Using the model
Training the model
7 - Sample
8 - Clustering - PAM
9 - Clusters Medoids and Centroids
10 - Medoid labelling Two levels
11 – Intra-Class Dispersion measures
12 - Model Results
Distance
Medoid Set
Threshold
Label Set
14- Points of Interest

Figure 3

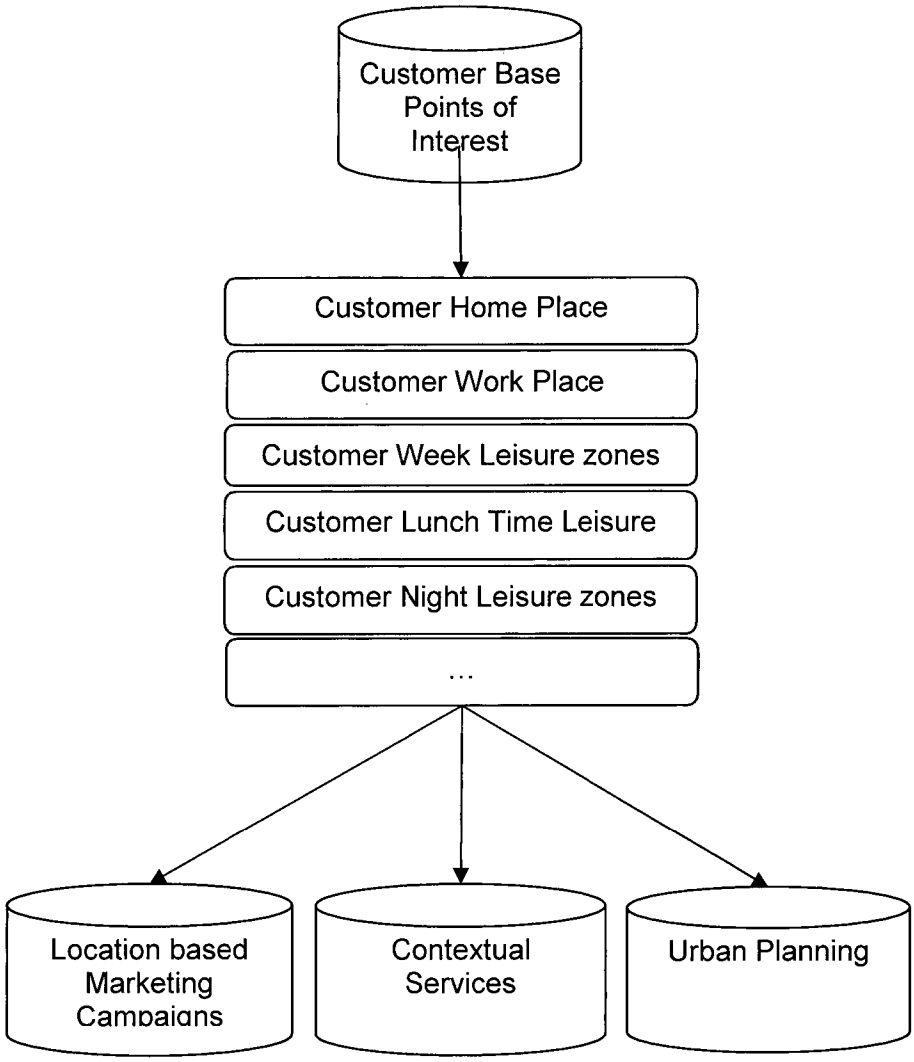

Customer Base Points of Interest
Customer Home Place
Customer Work Place
Customer Week Leisure zones
Customer Lunch Time Leisure
Customer Night Leisure zones
...
Location based Marketing Campaigns
Contextual Services
Urban Planning

Figure 4

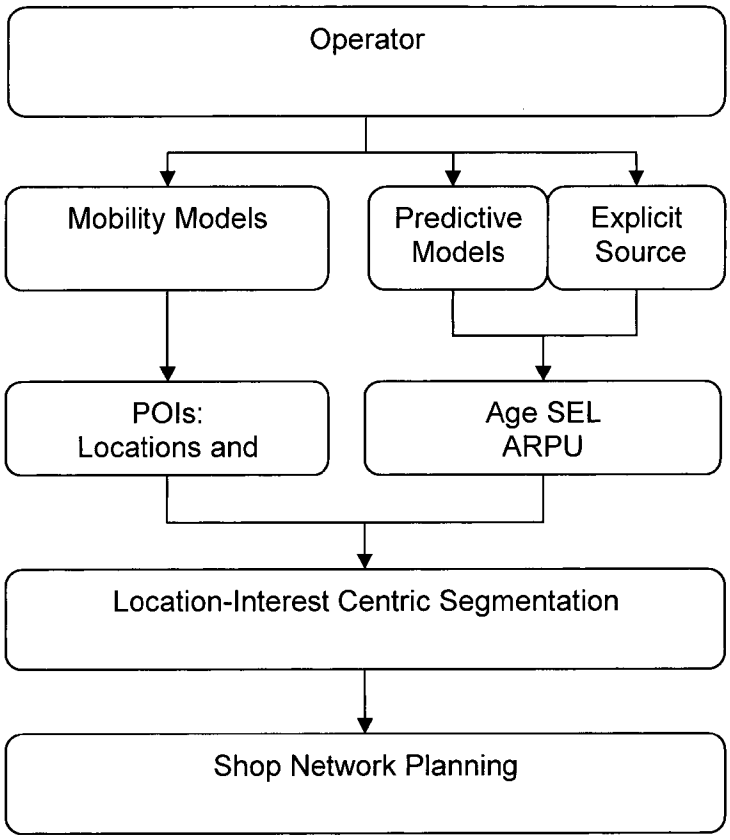

Operator
Mobility Models
Predictive Models
Explicit Source
POIs: Locations and
Age SEL ARPU
Location-Interest Centric Segmentation
Shop Network Planning

Figure 5

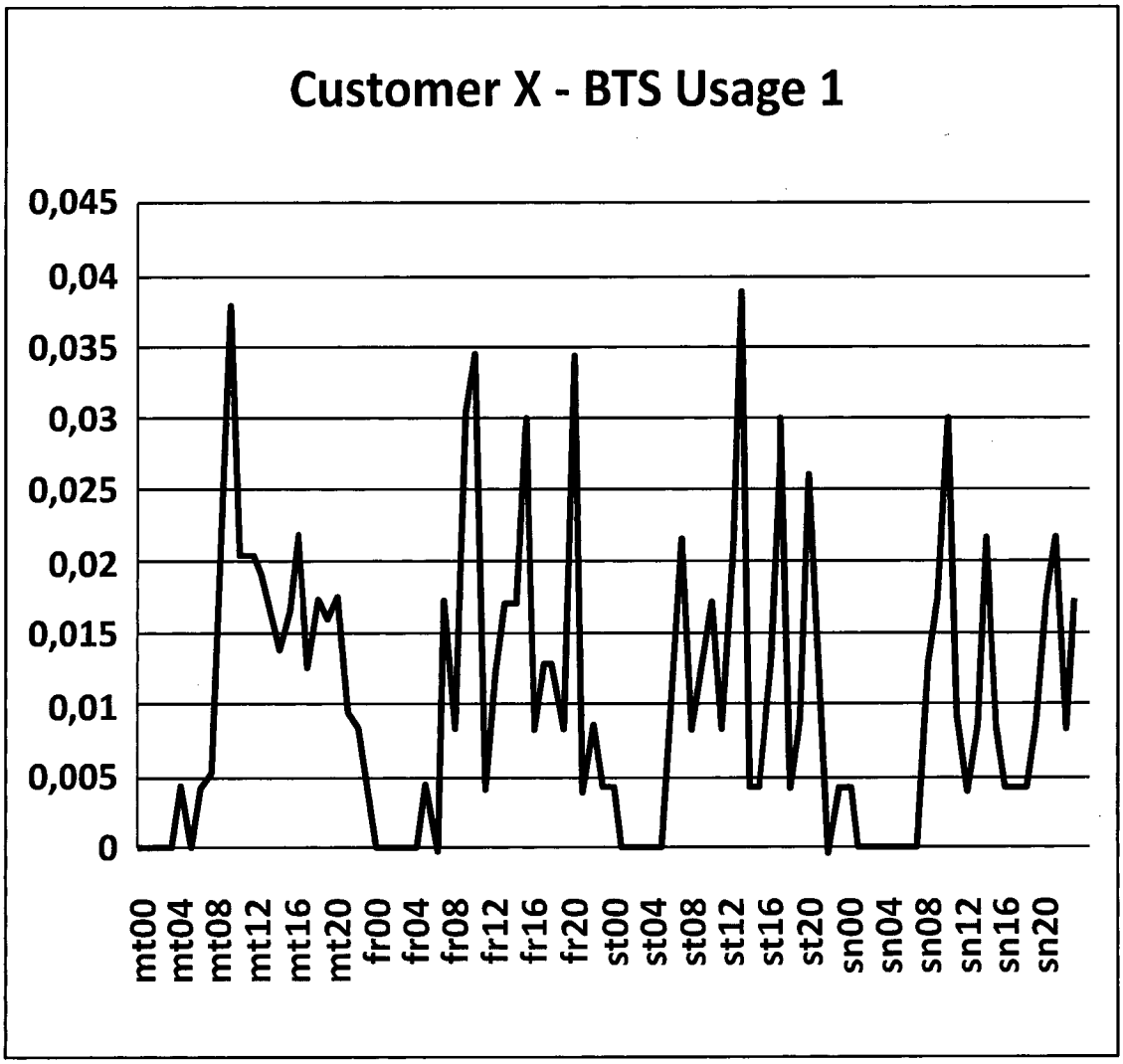

Customer X - BTS Usage 1
0,045
0,04
0,035
0,03
0,025
0,02
0,015
0,01
0,005
0
mt00
mt04
mt08
mt12
mt16
mt20
fr00
fr04
fr08
fr12
fr16
fr20
st00
st04
st08
st12
st16
st20
sn00
sn04
sn08
sn12
sn16
sn20

Figure 6a

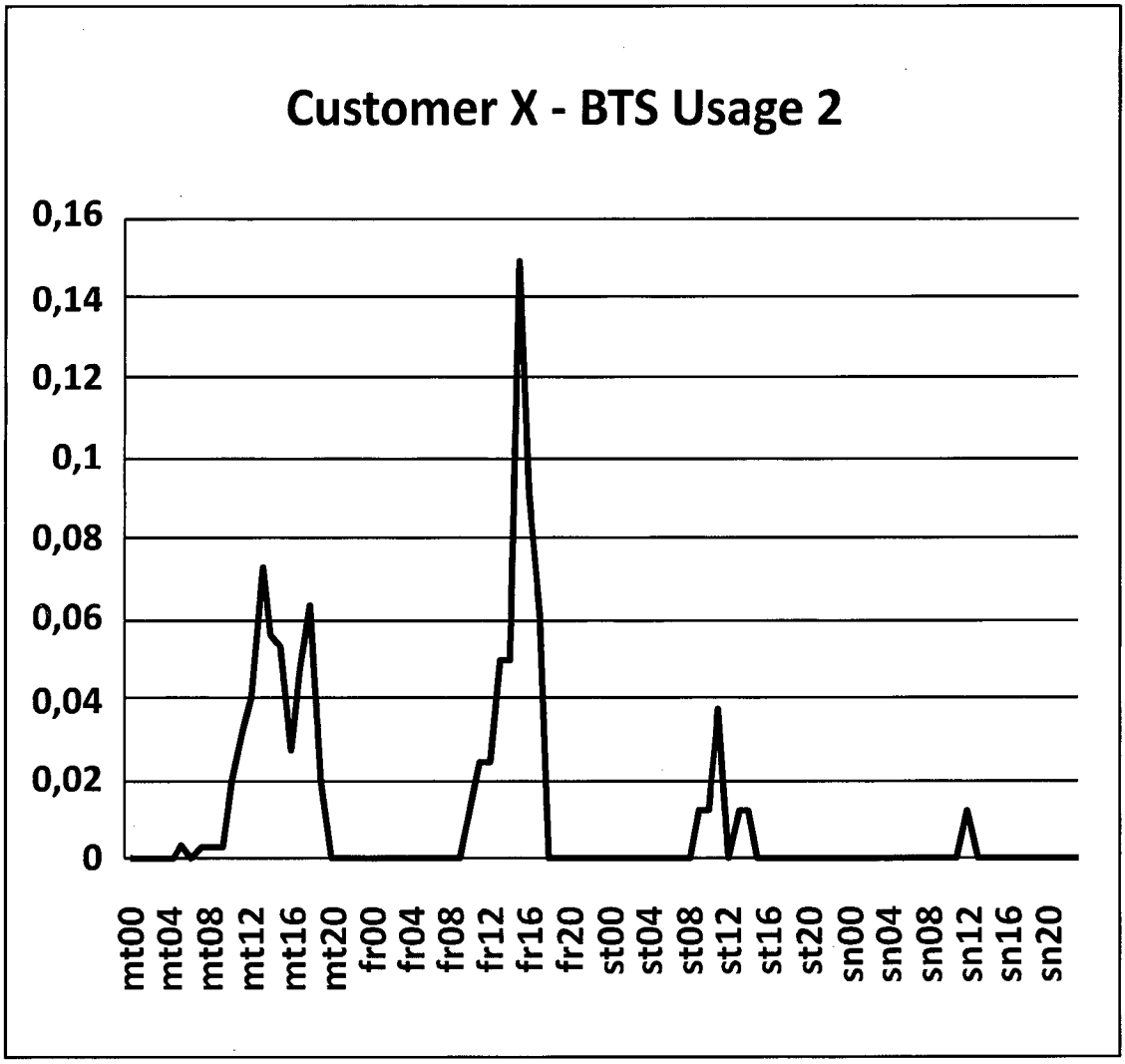

Customer X - BTS Usage 2
0,16
0,14
0,12
0,1
0,08
0,06
0,04
0,02
0
mt00
mt04
mt08
mt12
mt16
mt20
fr00
fr04
fr08
fr12
fr16
fr20
st00
st04
st08
st12
st16
st20
sn00
sn04
sn08
sn12
sn16
sn20

Figure 6b

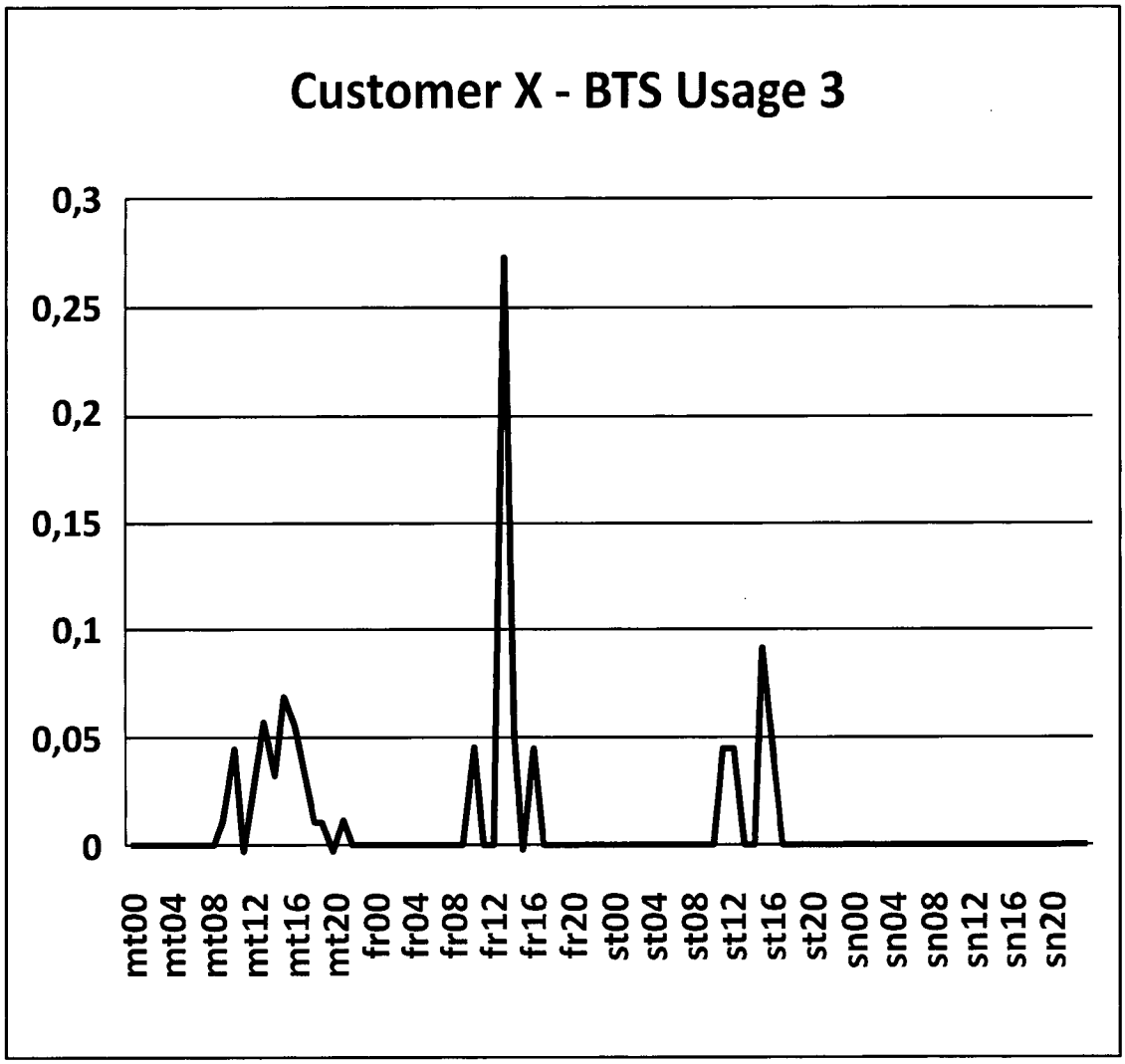
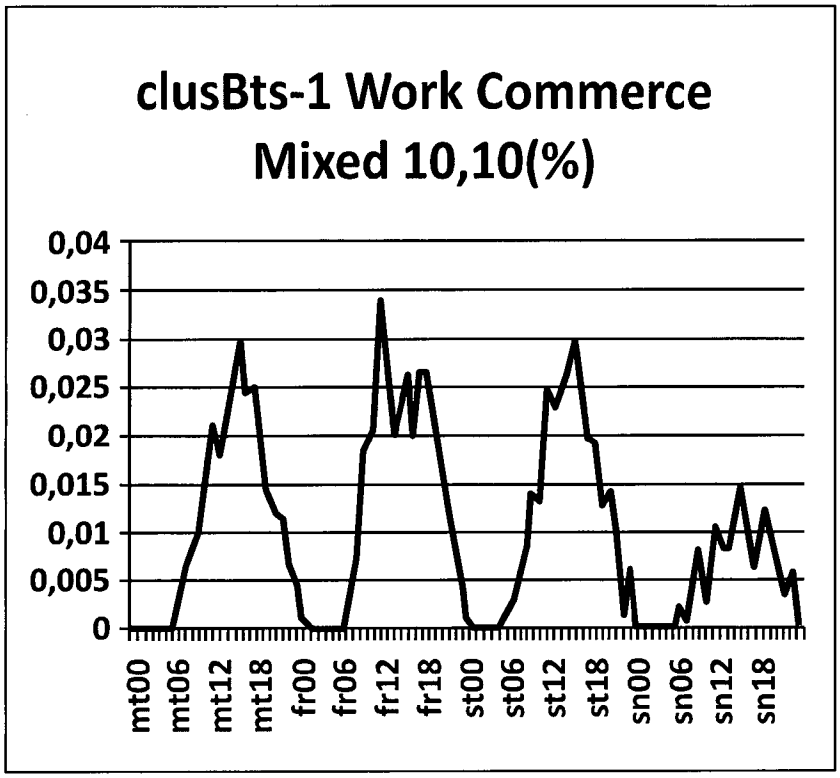

Customer X - BTS Usage 3
0,3
0,25
0,2
0,15
0,1
0,05
0
mt00
mt04
mt08
mt12
mt16
mt20
fr00
fr04
fr08
fr12
fr16
fr20
st00
st04
st08
st12
st16
st20
sn00
sn04
sn08
sn12
sn16
sn20

Figure 6c


clusBts-1 Work Commerce Mixed 10,10(%)
0,04
0,035
0,03
0,025
0,02
0,015
0,01
0,005
0
mt00 mt06 mt12 mt18 fr00 fr06 fr12 fr18 st00 st06 st12 st18 sn00 sn06 sn12 sn18

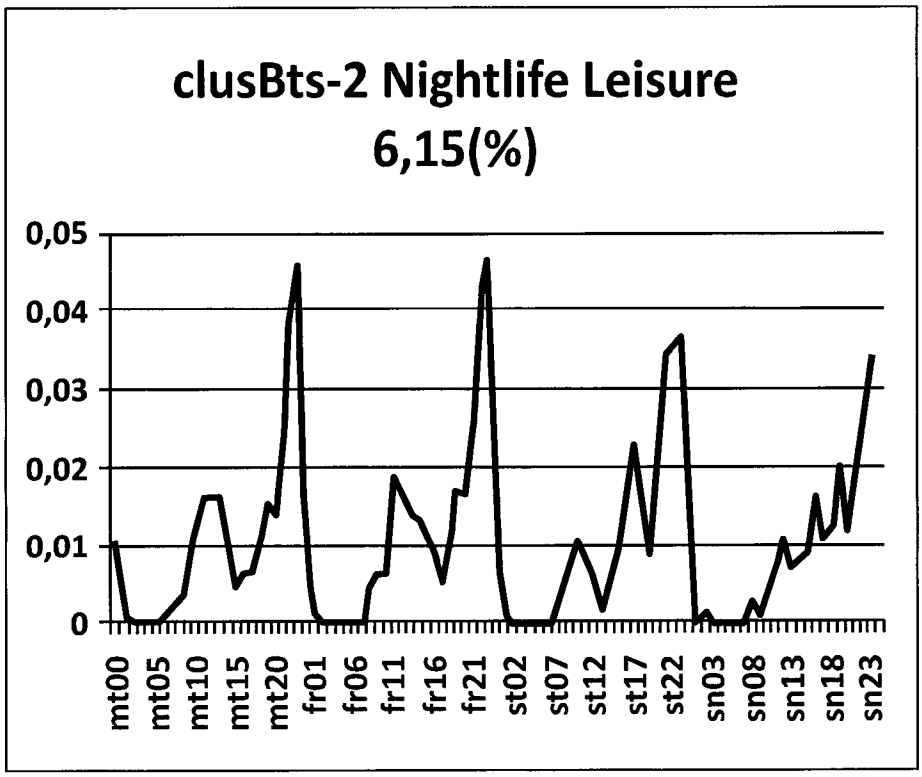

clusBts-2 Nightlife Leisure 6,15(%)
0,05
0,04
0,03
0,02
0,01
0
mt00 mt05 mt10 mt15 mt20 fr01 fr06 fr11 fr16 fr21 st02 st07 st12 st17 st22 sn03 sn08 sn13 sn18 sn23

Figure 7a

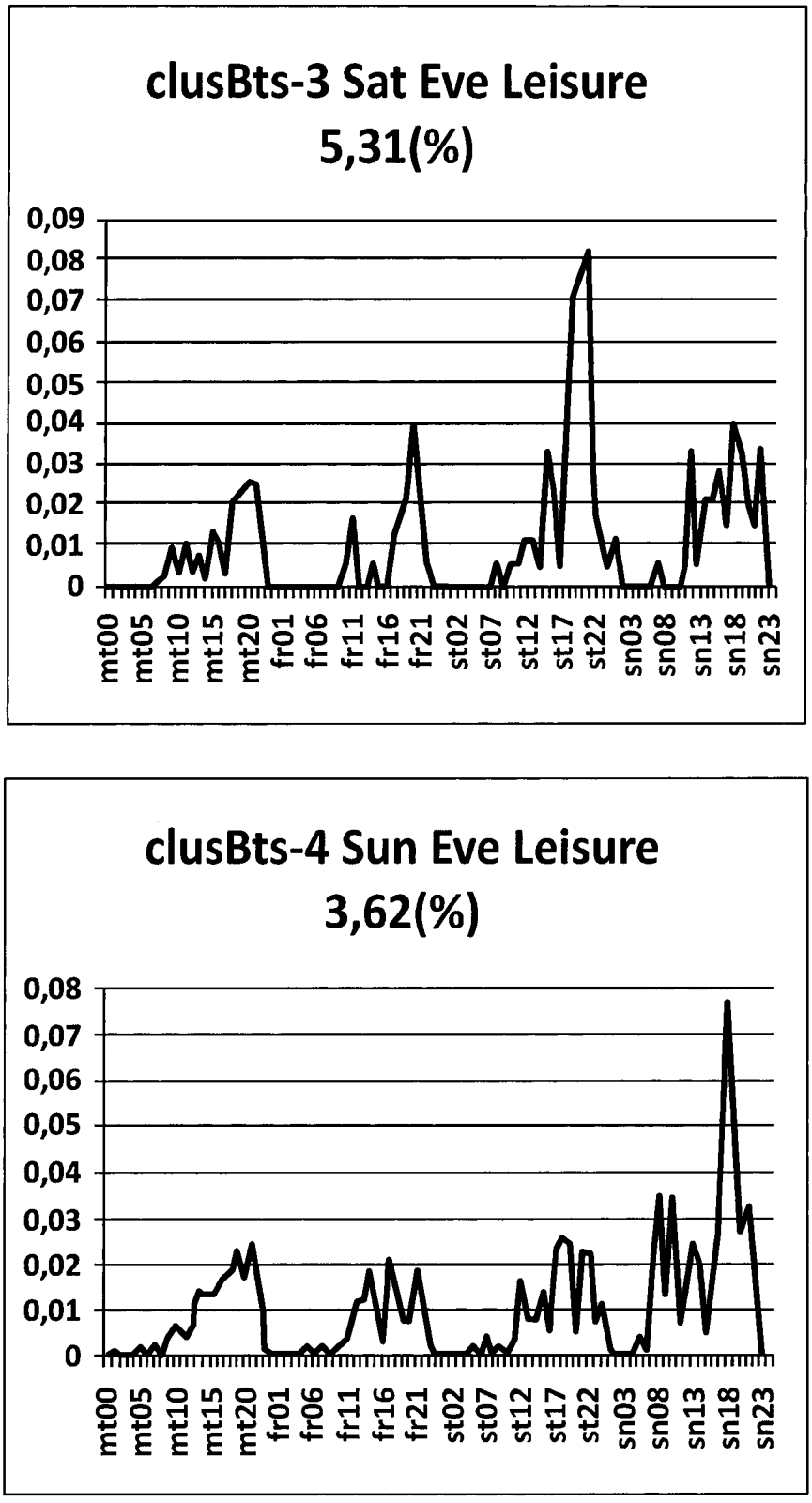
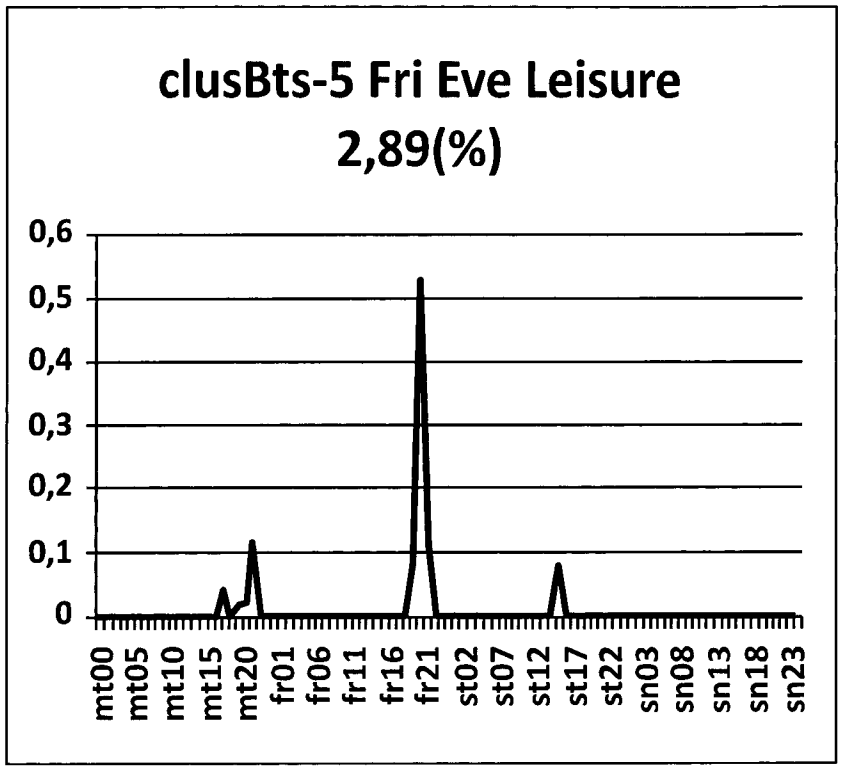

clusBts-3 Sat Eve Leisure
5,31(%)
0,09
0,08
0,07
0,06
0,05
0,04
0,03
0,02
0,01
0
mt00 mt05 mt10 mt15 mt20 fr01 fr06 fr11 fr16 fr21 st02 st07 st12 st17 st22 sn03 sn08 sn13 sn18 sn23
clusBts-4 Sun Eve Leisure
3,62(%)
0,08
0,07
0,06
0,05
0,04
0,03
0,02
0,01
0
mt00 mt05 mt10 mt15 mt20 fr01 fr06 fr11 fr16 fr21 st02 st07 st12 st17 st22 sn03 sn08 sn13 sn18 sn23

Figure 7b


clusBts-5 Fri Eve Leisure
2,89(%)
0,6
0,5
0,4
0,3
0,2
0,1
0
mt00 mt05 mt10 mt15 mt20 fr01 fr06 fr11 fr16 fr21 st02 st07 st12 st17 st22 sn03 sn08 sn13 sn18 sn23

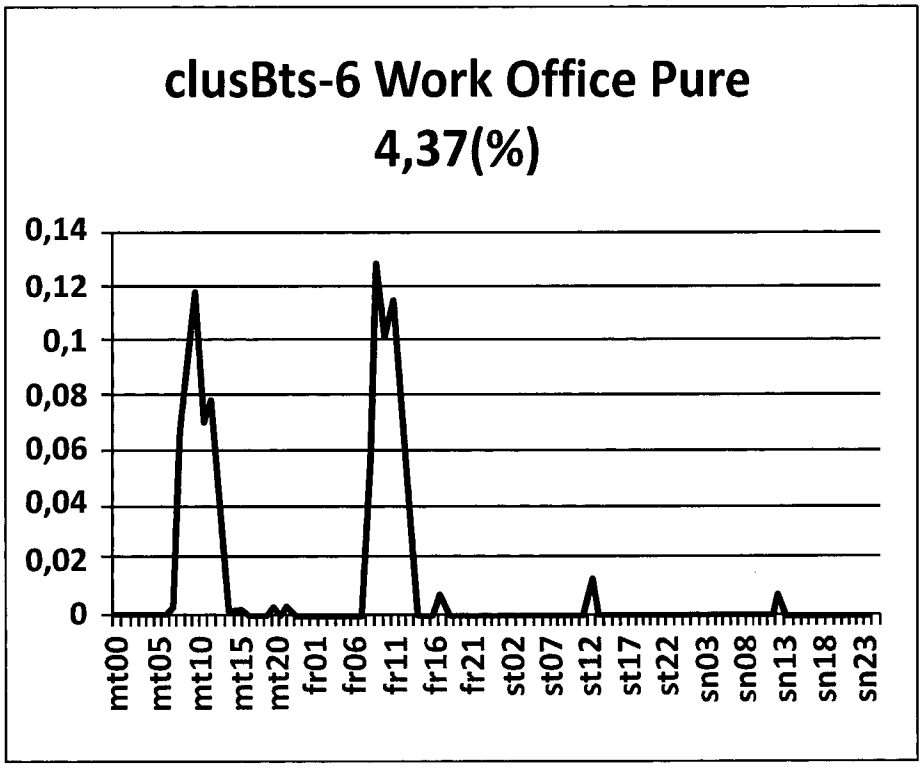
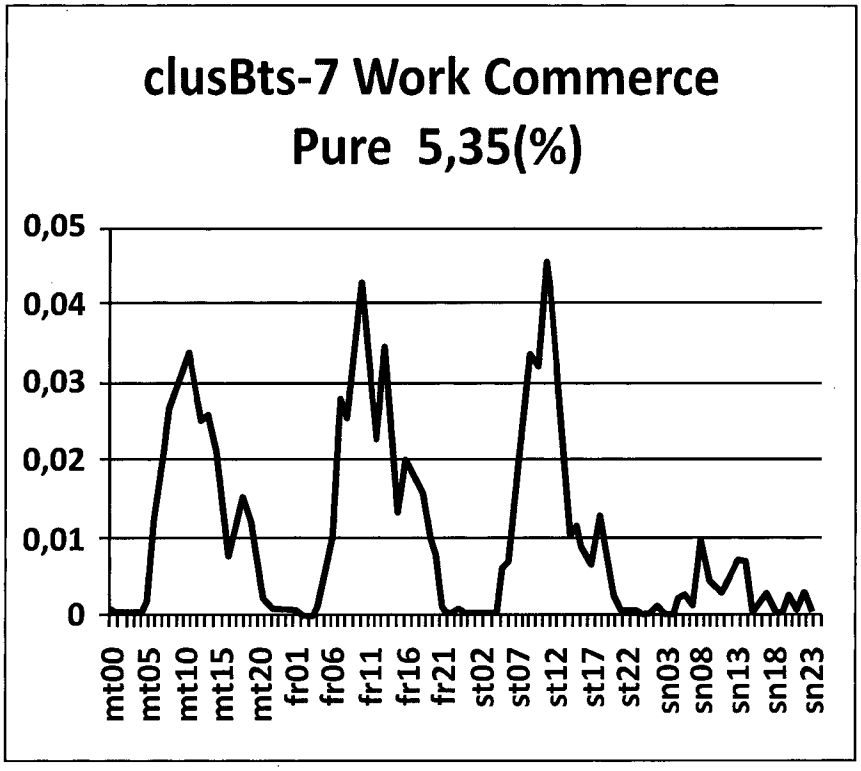

clusBts-6 Work Office Pure
4,37(%)
0,14
0,12
0,1
0,08
0,06
0,04
0,02
0
mt00 mt05 mt10 mt15 mt20 fr01 fr06 fr11 fr16 fr21 st02 st07 st12 st17 st22 sn03 sn08 sn13 sn18 sn23

Figure 7c


clusBts-7 Work Commerce
Pure 5,35(%)
0,05
0,04
0,03
0,02
0,01
0
mt00 mt05 mt10 mt15 mt20 fr01 fr06 fr11 fr16 fr21 st02 st07 st12 st17 st22 sn03 sn08 sn13 sn18 sn23

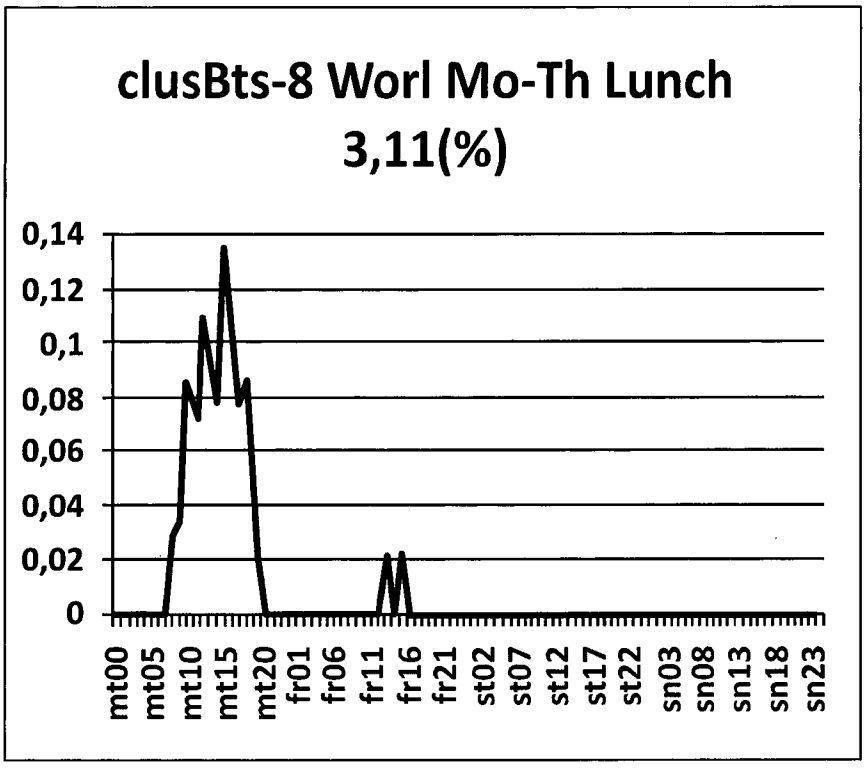
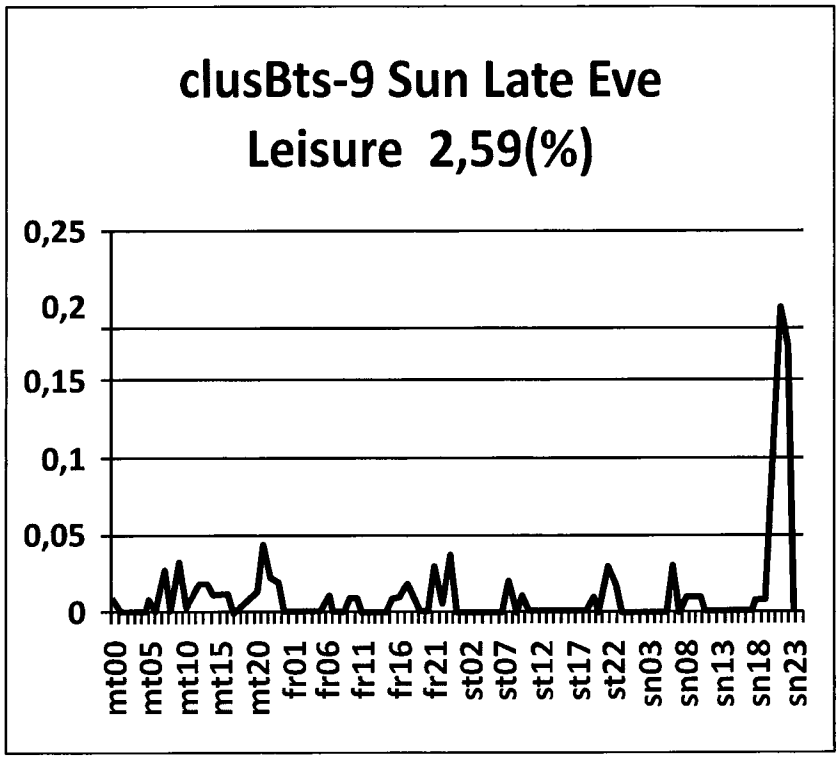

clusBts-8 Worl Mo-Th Lunch
3,11(%)
0,14
0,12
0,1
0,08
0,06
0,04
0,02
0
mt00 mt05 mt10 mt15 mt20 fr01 fr06 fr11 fr16 fr21 st02 st07 st12 st17 st22 sn03 sn08 sn13 sn18 sn23

Figure 7d


clusBts-9 Sun Late Eve
Leisure 2,59(%)
0,25
0,2
0,15
0,1
0,05
0
mt00 mt05 mt10 mt15 mt20 fr01 fr06 fr11 fr16 fr21 st02 st07 st12 st17 st22 sn03 sn08 sn13 sn18 sn23

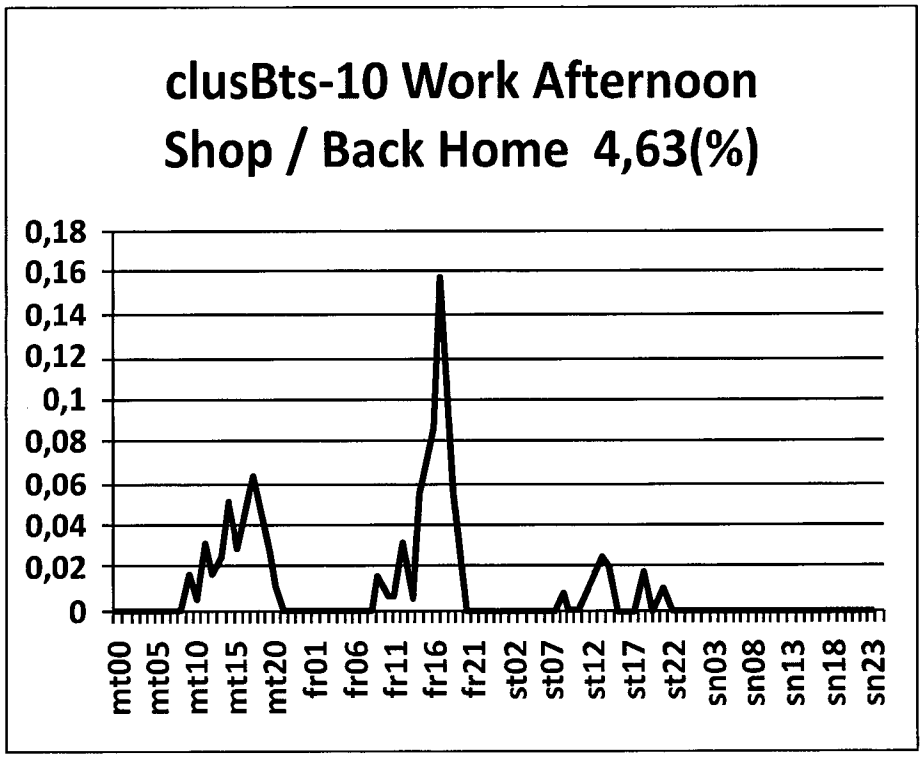
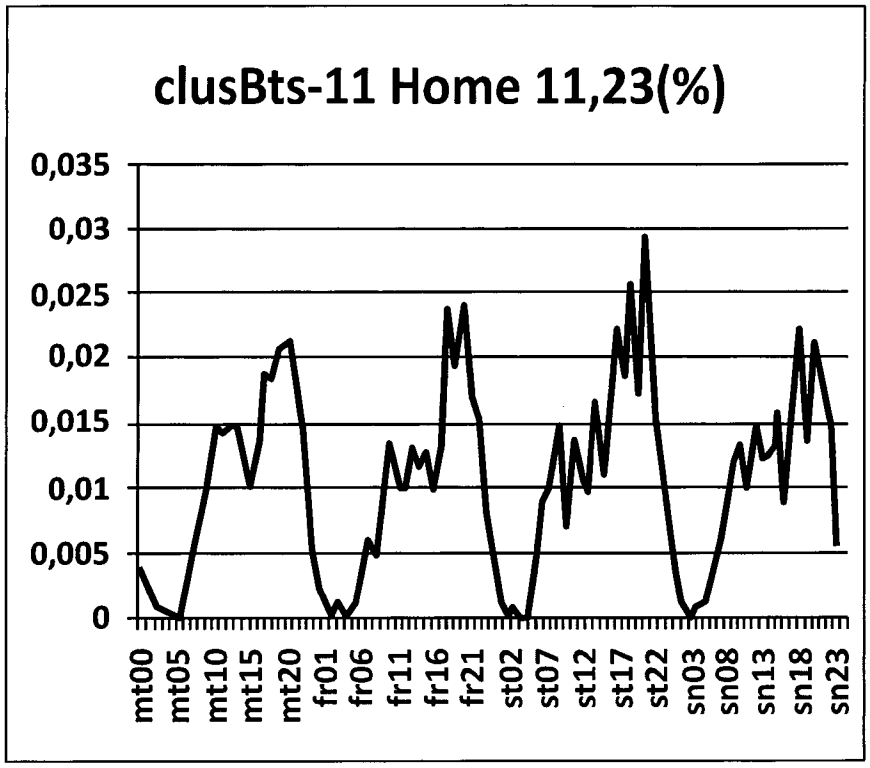

clusBts-10 Work Afternoon
Shop / Back Home 4,63(%)
0,18
0,16
0,14
0,12
0,1
0,08
0,06
0,04
0,02
0
mt00 mt05 mt10 mt15 mt20 fr01 fr06 fr11 fr16 fr21 st02 st07 st12 st17 st22 sn03 sn08 sn13 sn18 sn23

Figure 7e


clusBts-11 Home 11,23(%)
0,035
0,03
0,025
0,02
0,015
0,01
0,005
0
mt00 mt05 mt10 mt15 mt20 fr01 fr06 fr11 fr16 fr21 st02 st07 st12 st17 st22 sn03 sn08 sn13 sn18 sn23

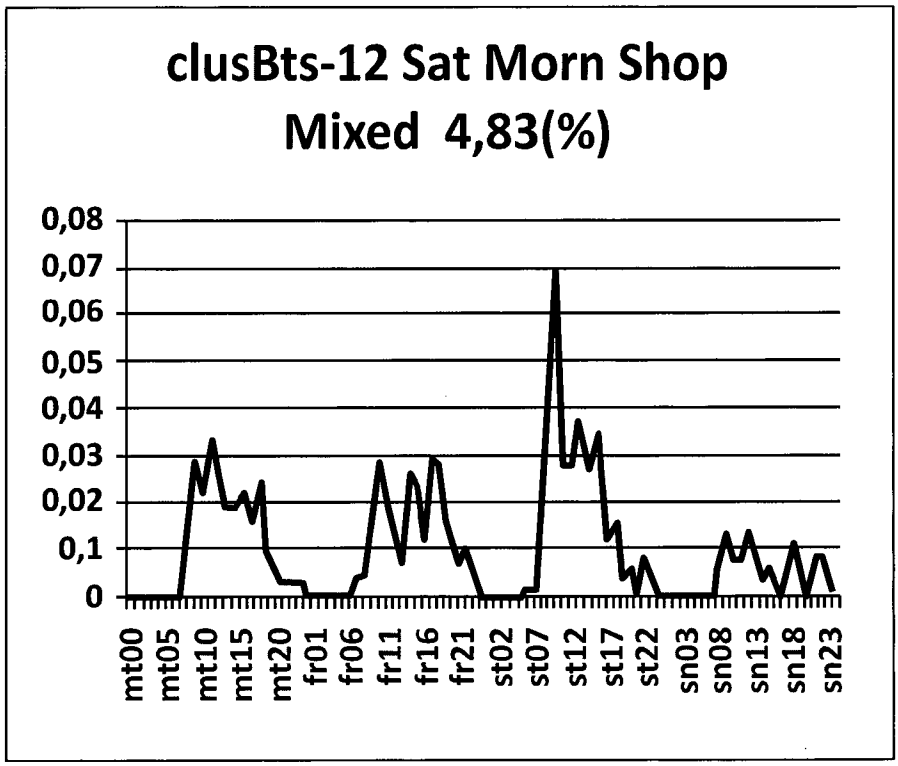
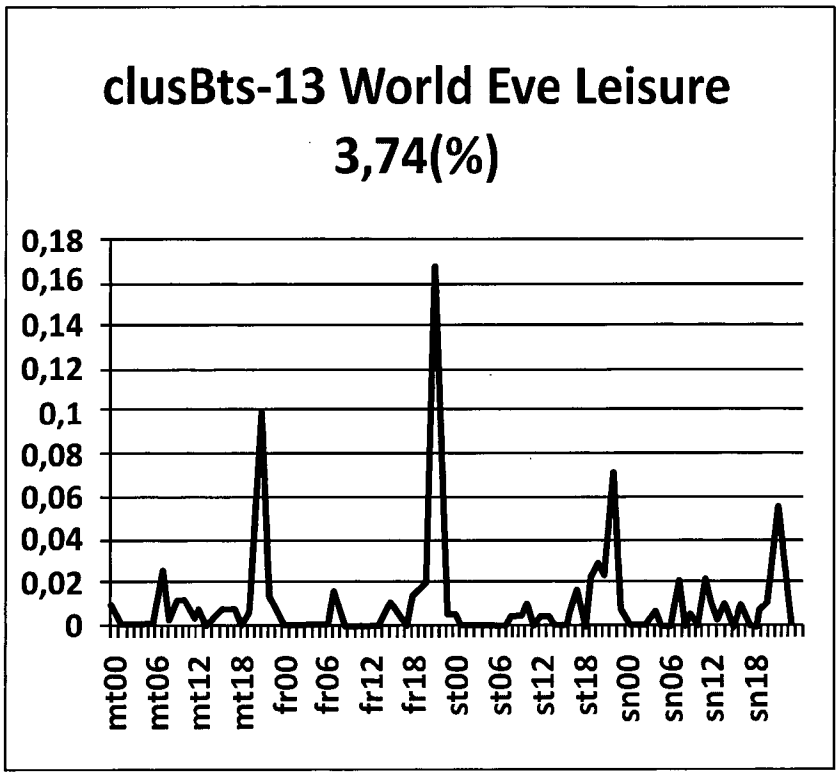

clusBts-12 Sat Morn Shop Mixed 4,83(%)
0,08
0,07
0,06
0,05
0,04
0,03
0,02
0,1
0
mt00 mt05 mt10 mt15 mt20 fr01 fr06 fr11 fr16 fr21 st02 st07 st12 st17 st22 sn03 sn08 sn13 sn18 sn23

Figure 7f


clusBts-13 World Eve Leisure
3,74(%)
0,18
0,16
0,14
0,12
0,1
0,08
0,06
0,04
0,02
0
mt00
mt06
mt12
mt18
fr00
fr06
fr12
fr18
st00
st06
st12
st18
sn00
sn06
sn12
sn18

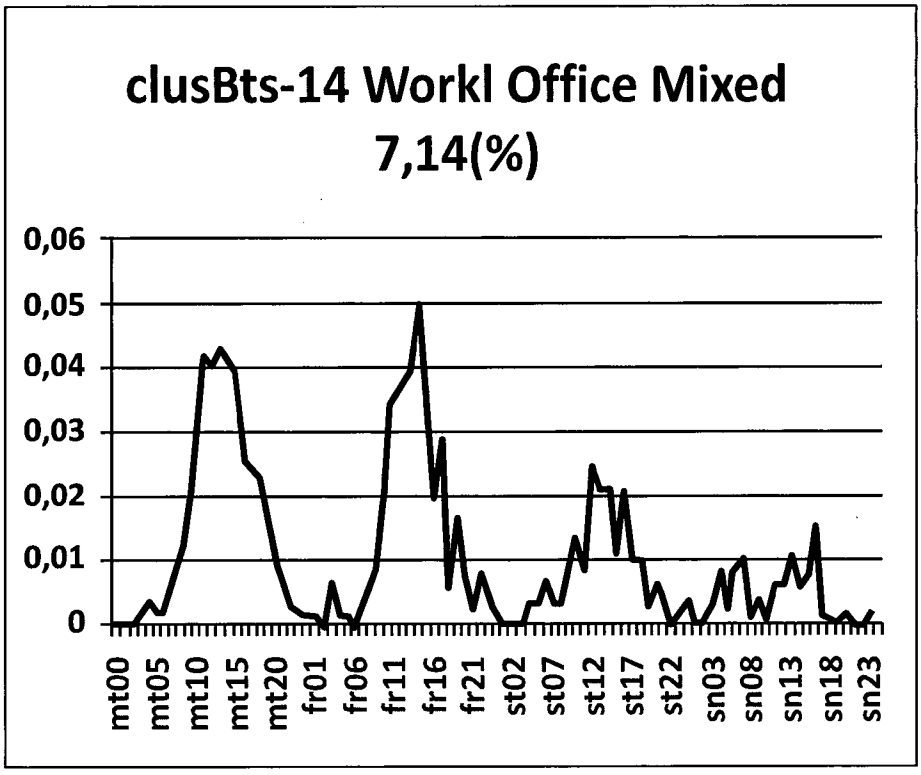
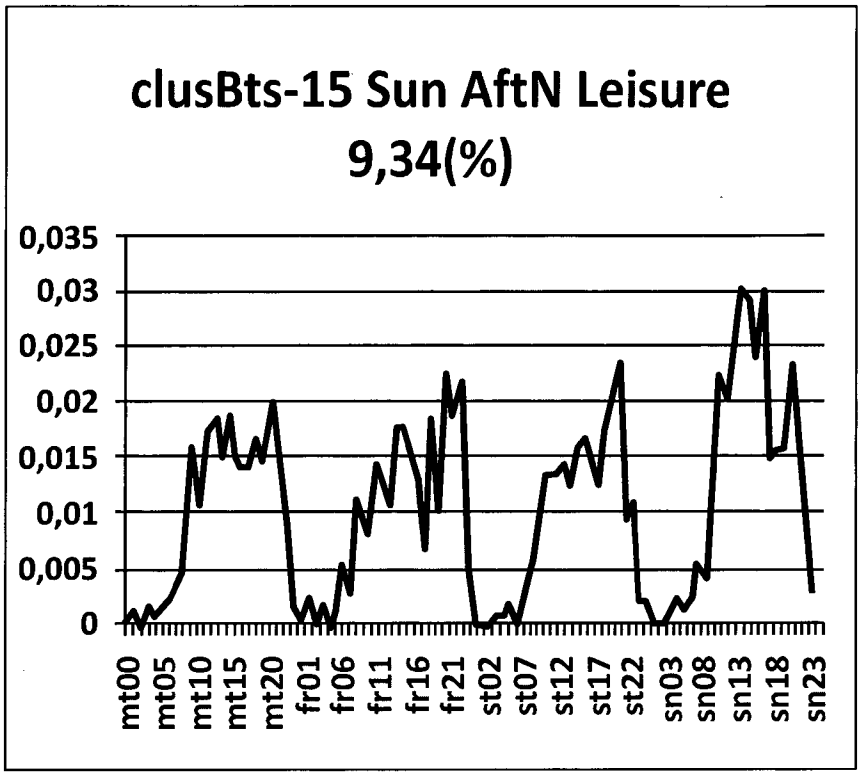

clusBts-14 Workl Office Mixed
7,14(%)
0,06
0,05
0,04
0,03
0,02
0,01
0
mt00
mt05
mt10
mt15
mt20
fr01
fr06
fr11
fr16
fr21
st02
st07
st12
st17
st22
sn03
sn08
sn13
sn18
sn23

Figure 7g


clusBts-15 Sun AftN Leisure
9,34(%)
0,035
0,03
0,025
0,02
0,015
0,01
0,005
0
mt00 mt05 mt10 mt15 mt20 fr01 fr06 fr11 fr16 fr21 st02 st07 st12 st17 st22 sn03 sn08 sn13 sn18 sn23

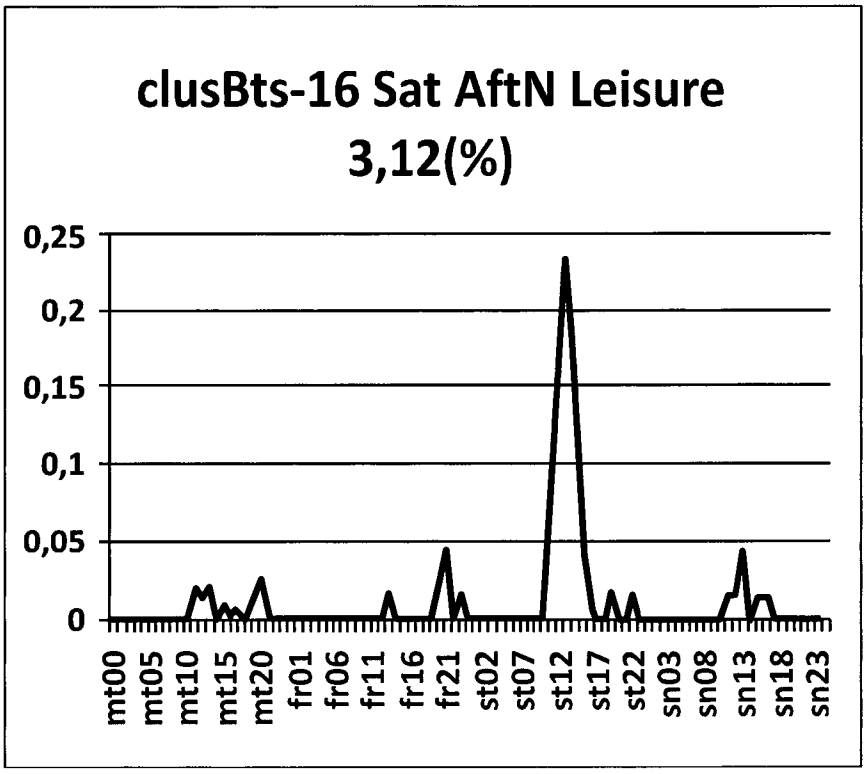
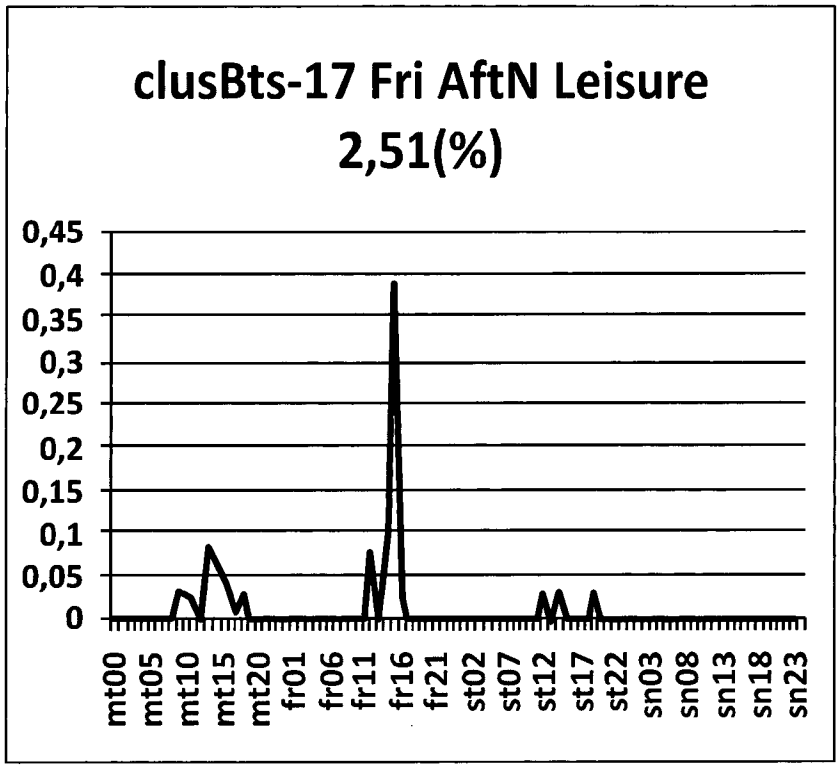

clusBts-16 Sat AftN Leisure
3,12(%)
0,25
0,2
0,15
0,1
0,05
0
mt00 mt05 mt10 mt15 mt20 fr01 fr06 fr11 fr16 fr21 st02 st07 st12 st17 st22 sn03 sn08 sn13 sn18 sn23

Figure 7h


clusBts-17 Fri AftN Leisure
2,51(%)
0,45
0,4
0,35
0,3
0,25
0,2
0,15
0,1
0,05
0
mt00 mt05 mt10 mt15 mt20 fr01 fr06 fr11 fr16 fr21 st02 st07 st12 st17 st22 sn03 sn08 sn13 sn18 sn23

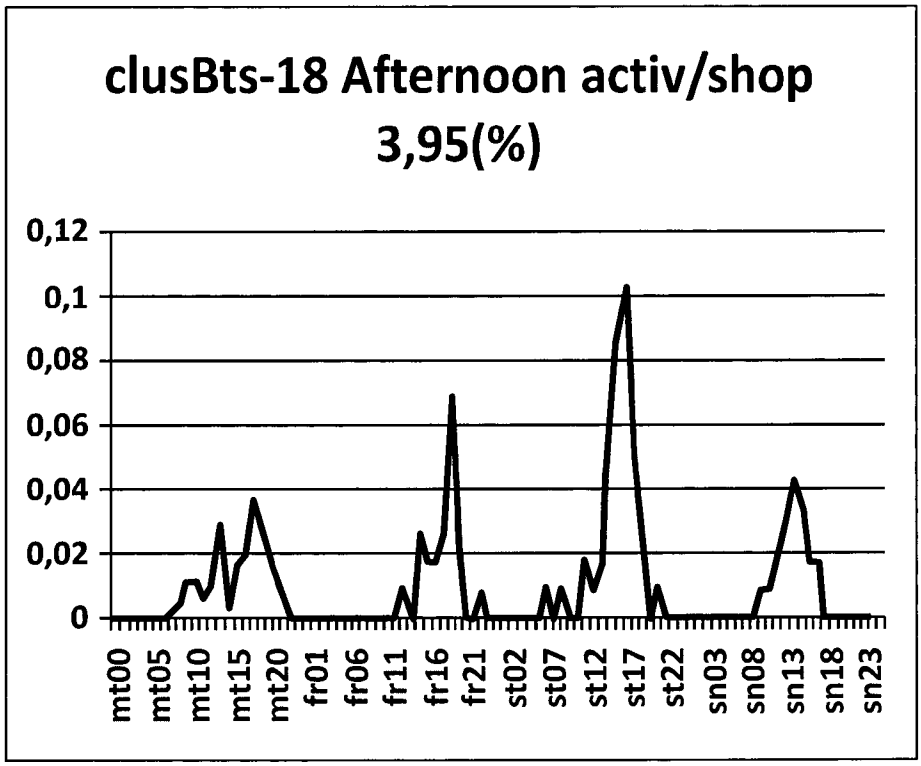
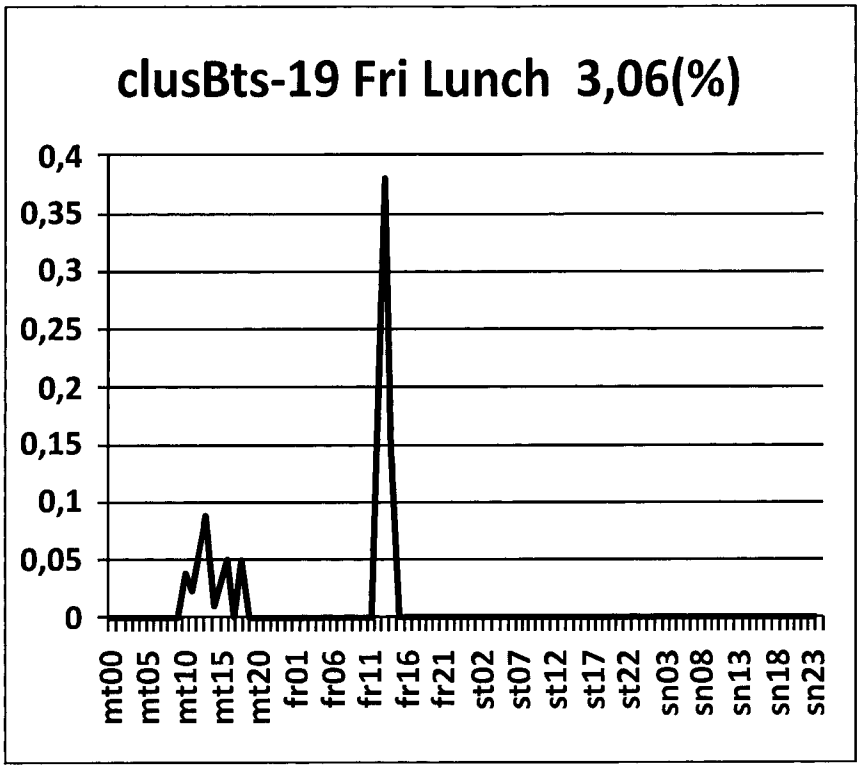

clusBts-18 Afternoon activ/shop
3,95(%)
0,12
0,1
0,08
0,06
0,04
0,02
0
mt00 mt05 mt10 mt15 mt20 fr01 fr06 fr11 fr16 fr21 st02 st07 st12 st17 st22 sn03 sn08 sn13 sn18 sn23

Figure 7i


clusBts-19 Fri Lunch 3,06(%)
0,4
0,35
0,3
0,25
0,2
0,15
0,1
0,05
0
mt00 mt05 mt10 mt15 mt20 fr01 fr06 fr11 fr16 fr21 st02 st07 st12 st17 st22 sn03 sn08 sn13 sn18 sn23

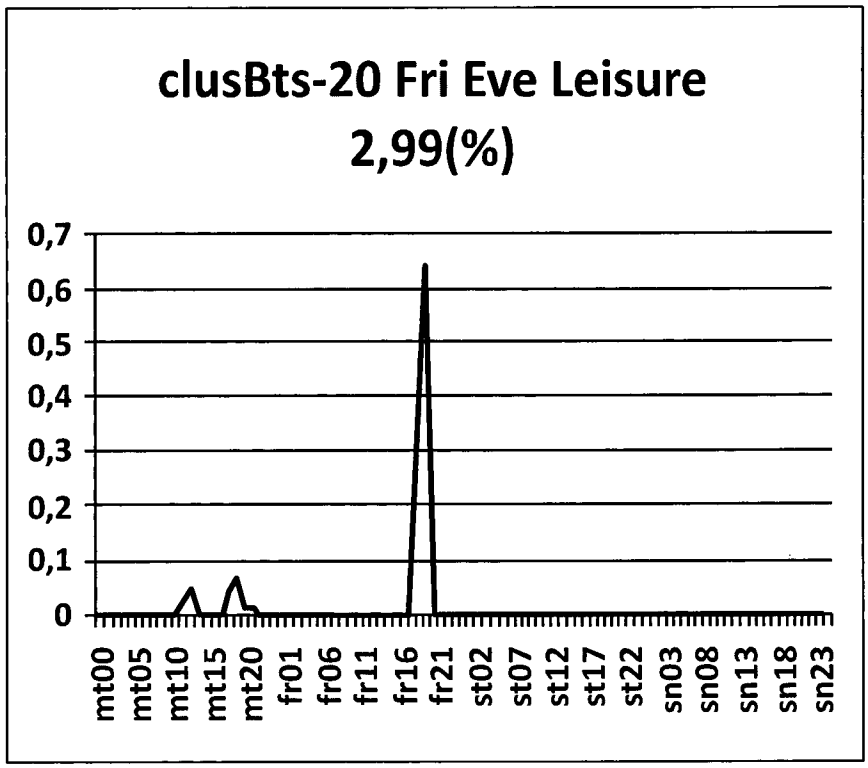

clusBts-20 Fri Eve Leisure 2,99(%)
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
mt00 mt05 mt10 mt15 mt20 fr01 fr06 fr11 fr16 fr21 st02 st07 st12 st17 st22 sn03 sn08 sn13 sn18 sn23

Figure 7j

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2013/050225

A. CLASSIFICATION OF SUBJECT MATTER
INV. H04W4/02
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
H04W

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2008/134595 A1 (PELAGO INC [US]; VENGROFF DARREN E [US]; DOWNS OLIVER B [US]) 6 November 2008 (2008-11-06) paragraphs [0003], [0014], [0015] paragraphs [0030], [0034] paragraphs [0041], [0045], [0048], [0049] paragraphs [0070], [0073], [0074] paragraphs [0023], [0026] - [0029] paragraph [0063] paragraphs [0053] - [0056] -----	1-10
Y	US 2011/151898 A1 (CHANDRA UMESH [US] ET AL) 23 June 2011 (2011-06-23) paragraphs [0026] - [0030] paragraphs [0060], [0061] -----	1-10

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
11 March 2013	15/03/2013

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Mele, Marco

2

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2003/036848 A1 (SHEHA MICHAEL A [US] ET AL) 20 February 2003 (2003-02-20) abstract	1-10
A	US 2010/197318 A1 (PETERSEN STEVEN L [US] ET AL) 5 August 2010 (2010-08-05) abstract	1-10
A	KAUFMAN L ET AL: "Partitioning Around Medoids (Program PAM)", FINDING GROUPS IN DATA: AN INTRODUCTION TO CLUSTER ANALYSIS,, vol. Chapter 2, 1 January 1990 (1990-01-01), pages 58-125, XP007921092, ISBN: 0-471-73578-7 cited in the application abstract	1-10
A	CALABRESE F ET AL: "Human mobility prediction based on individual and collective geographical preferences", 19 September 2010 (2010-09-19), INTELLIGENT TRANSPORTATION SYSTEMS (ITSC), 2010 13TH INTERNATIONAL IEEE CONFERENCE ON, IEEE, PISCATAWAY, NJ, USA, PAGE(S) 312 - 317, XP031792650, ISBN: 978-1-4244-7657-2 cited in the application the whole document	1-10
A	MATTEO ZIGNANI ET AL: "Extracting human mobility patterns from GPS-based traces", 20 October 2010 (2010-10-20), WIRELESS DAYS (WD), 2010 IFIP, IEEE, PISCATAWAY, NJ, USA, PAGE(S) 1 - 5, XP031818026, ISBN: 978-1-4244-9230-5 cited in the application the whole document	1-10
A	SANTI PHITHAKKITNUKOON ET AL: "Activity-Aware Map: Identifying Human Daily Activity Pattern Using Mobile Phone Data", HUMAN BEHAVIOR UNDERSTANDING, SPRINGER BERLIN HEIDELBERG, BERLIN, HEIDELBERG, vol. 6219, 22 August 2010 (2010-08-22), pages 14-25, XP019147862, ISBN: 978-3-642-14714-2 cited in the application abstract	1-10

2

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2013/050225

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2008134595 A1	06-11-2008	US 2009005987 A1	01-01-2009
		US 2012310736 A1	06-12-2012
		WO 2008134595 A1	06-11-2008
US 2011151898 A1	23-06-2011	CN 102783189 A	14-11-2012
		US 2011151898 A1	23-06-2011
		WO 2011076988 A1	30-06-2011
US 2003036848 A1	20-02-2003	US 2003036848 A1	20-02-2003
		US 2006229807 A1	12-10-2006
		US 2008076451 A1	27-03-2008
		US 2012011137 A1	12-01-2012
US 2010197318 A1	05-08-2010	US 2010197318 A1	05-08-2010
		US 2010197319 A1	05-08-2010
		US 2010198814 A1	05-08-2010
		US 2010198826 A1	05-08-2010
		US 2010198828 A1	05-08-2010
		US 2010198862 A1	05-08-2010
		US 2010198870 A1	05-08-2010
		US 2010198917 A1	05-08-2010
		US 2012041983 A1	16-02-2012
		US 2013017843 A1	17-01-2013