Fuzzy Recommendations in Marketing Campaigns

S. Podapati[1], L. Lundberg[1], L. Skold[2], O. Rosander[1], and J. Sidorova[1](✉)

[1] Department of CS and Engineering,
Blekinge Institute of Technology, Karlskrona, Sweden
julia.a.sidorova@gmail.com
[2] Telenor, Stockholm, Sweden

Abstract. The population in Sweden is growing rapidly due to immigration. In this light, the issue of infrastructure upgrades to provide telecommunication services is of importance. New antennas can be installed at hot spots of user demand, which will require an investment, and/or the clientele expansion can be carried out in a planned manner to promote the exploitation of the infrastructure in the less loaded geographical zones. In this paper, we explore the second alternative. Informally speaking, the term Infrastructure-Stressing describes a user who stays in the zones of high demand, which are prone to produce service failures, if further loaded. We have studied the Infrastructure-Stressing population in the light of their correlation with geo-demographic segments. This is motivated by the fact that specific geo-demographic segments can be targeted via marketing campaigns. Fuzzy logic is applied to create an interface between big data, numeric methods for its processing, and a manager who wants a comprehensible summary.

Keywords: Intelligent data mining · Call detail records · Fuzzy membership function · Geo-demographic segments · Marketing

1 Introduction

In the era of big data a mapping is desired from multitudes of numeric data to a useful summary and insights expressed in a natural language yet with a mathematical precision [1]. Fuzzy logic bridges from mathematics to the way humans reason and the way the human world operates. Clearly, the "class of all real numbers which are much greater than 1," or "the class of beautiful women," or "the class of tall men," do not constitute classes or sets in the usual mathematical sense of these terms. Yet, "the fact remains that such imprecisely defined notions play an important role in human thinking, particularly in the domains of decision-making, abstraction and communication of information" [2]. Few works exist in business intelligence that use fuzzy logic due to certain inherent difficulties of creating such applications, and yet; despite them, such applications are possible and very useful, e.g. the reader can be referred to a review [3]. The challenges include the following. Firstly, not every problem permits trial and error calibration of threshold values. Secondly, the operators, membership

© Springer International Publishing AG 2017
M. Kirikova et al. (Eds.): ADBIS 2017, CCIS 767, pp. 246–256, 2017.
DOI: 10.1007/978-3-319-67162-8_24

functions and inference methods need to have tangible meanings, which can be very context-dependent. Thirdly, fuzzy theory is a borderline discipline with psycholinguistics, which is less objective than formal sciences (such as logic or set theory) and may require yet unavailable knowledge about human cognition. The notion of fuzzieness has distinct understandings and there are important consequences of those discrepancies, e.g. [4] is a review of theoretical models and their empirical validations. The main two types of fuzzy technology are fuzzy knowledge based systems [3] and fuzzy clustering [5]. Our idea is neither of the two, and it aims to implement the above mentioned insight by Zadeh about completing a useful summary from multitudes of data. Fuzzy logic enables us to formulate a natural language interface between big data, numeric analytics, and a manager, hiding the compexity of data and methods and providing him/her with a comprehensible summary. We summarize data using linguistic hedges (very, rather, highly) and formulate queries such as "*Tell me which neighbourhoods are safe to target, if I want more clients but my infrastructure is highly loaded*". "*Tell me, whether the infrastructure is rather loaded or highly loaded in the region.*" Our specific application is targeting different user segments to fill in the spare capacity of the network in a network-friendly manner. In [6], the notion of *Infrastructure-Stressing* (IS) Client was proposed together with the method to reveal such clients from the customer base. Informally, IS clients use the infrastructure in a stressing manner, such as always staying in the zones of high demand, where the antennas are prone to service failures, if further loaded. Being IS is not only a function of the user's qualities, but also of the infrastructure, and of the relative mobility of the rest of the population.

For marketing campaigns geodemographic segmentations (like ACORN or MOSAIC) are used, since it is known how the segments can be targeted to achieve the desired goal, as for example, the promotion of a new mobile service in certain neigbourhoods. The client's home address determines the geodemographic category. People of similar social status and lifestyle tend to live close [7, 8]. Geodemographic segmentation provides a *collective view point*, where the client is seen as a representative of the population who live nearby. However, in recent research, it has been shown that the problem of resource allocation in the zones with nearly overloaded and underloaded antennas is better handled relying on *individual modelling* based on the client's historical trajectories [9]. The authors completed a user segmentation based on clustering of user trajectories and it was demonstrated that network planning is more effective, if trajectory-based segments are used instead of geo-demographic segments. Our aim is to explore the ways to connect the individual trajectory-based view on IS customers and the geodemographic view in order to devise analytics capable to complete the efficient analysis based on the individual view point and yet be useful in marketing campaigns in which geodemographic groups are targeted. As a practical conclusion, we have compiled a ranked list of the segments according to their propensity to contain IS clients and crafted two queries:

1. Which segments contain a low or moderate number of IS clients? (target them, while the infrastructure is still rather underloaded)
2. Which segment is highly devoid of IS clients? (target them, when the customer base becomes mature and the infrastructure becomes increasingly loaded).

The simulation of the resulting fuzzy recommendations guarantees the absence of false negatives, such as, concluding that certain segments are safe to hire from, but in fact that would lead to a service failure at some place in the network.

The rest of the paper is organised as follows. Section 2 describes the data set. In Sect. 3 the proposed methodology is explained. In Sect. 4, the experiments are reported, and finally the conclusions are drawn and discussion is held in Sect. 5.

2 Data Set

The study has been conducted on anonymized geospatial and geodemographic data provided by a Scandinavian telecommunication operator. The data consist of CDRs (Call Detail Records) containing historical location data and calls made during one week in a mid-size region in Sweden with more than one thousand radio cells. Several cells can be located on the same antenna. The cell density varies in different areas and is higher in city centers, compared to rural areas. The locations of 27010 clients are registered together with which cell serves the client. The location is registered every five minutes. During the periods when the client does not generate any traffic, she does not make any impact on the infrastructure and such periods of inactivity are not included in the resource allocation analysis. Every client in the database is labeled with her MOSAIC segment. The fields of the database used in this study are:

- the cells IDs with the information about which users it served at different time points,
- the location coordinates of the cells,
- the time stamps of every event (5 min resolution),
- the MOSAIC geodemographic segment for each client, and
- the Telenor geodemographic segment for each client.

There are 14 MOSAIC segments present in the database; for their detailed description the reader is referred to [15]. The six in-house Telenor segments were developed by Telenor in collaboration with InsightOne, and, to our best knowledge, though not conceptually different from MOSAIC, they are especially crafted for telecommunication businesses.

3 A Link Between IS and Geodemographic Segments

3.1 Notation and Definitions of Fuzzy Logic

Definition (in the style of [2]). A fuzzy set A in X is characterized by a membership function $f_A(x)$, which associates with each point in X a real number in the interval $[0, 1]$, with the value of $f_A(x)$ at x representing the "grade of membership" of x in A. For the opposite quality: $f_{notA}(x) = 1 - f_A(x)$.

Fuzzy membership scores reflect the varying degree to which different cases belong to a set. Under the six value fuzzy set, there are six degrees of membership 1: fully in, [$0.9-1$): mostly but not fully in, [$0.6-0.9$): more or less in, [$0.4-0.6$): more or less out, [$0.1-0.4$): mostly but not fully out, [$0-0.1$): fully out. For a comprehensive guide of good practices in fuzzy logic analysis in social sciences the reader is referred to, for example, [10].

Linguistic Hedges:

- *Rather* will be added to a quality A, if the square root of its membership function $f_A(x)^{1/2}$ is close to 1.
- *Very* will be added to a quality A, if the square of its membership function $f_A(x)^2$ is close to 1.
- *Extremely* will be added to a quality A, if $f_A(x)^3$ is close to 1.

The principles for calculating the values of hedged membership functions, for example $f_{veryA}(x) = f_A(x)^2$, are described in [14]. Then, given the new membership function, the same principle applies: the closer to 1, the higher is the degree of membership.

3.2 Query Formulation

To keep the formulations and questions naturally sounding, the word infrastructure-friendly (IF) is used. The quality IF is defined as the opposite to IS: $f_{IF}(segment_i) = 1 - f_{IS}(segment_i)$, for some segment i. As mentioned above, within the same geodemographic segment, the clients differ with respect to the degree of being IS. When the infrastructure is not overloaded, that is, the recent historical load is still significantly smaller than the capacity, then virtually any client is welcome. As the infrastructure becomes more loaded, the operator wants to be more discriminative. We define being "loaded" for an antenna as a fuzzy variable:

$$f_{loaded}(antenna\, j) = max_{all\, t}\{load(j,t) \times capacity(antenna\, j)^{-1}\}.$$

This quality is measured in man units. Being loaded for infrastructure is defined as:

$$f_{loaded}(infrastructure) = max_{all\, antennas\, j}\{f_{loaded}(antenna\, j)\}.$$

Since being loaded is a dangerous quality, we set the strength of the system to be equal to the strength of its weakest component, and for this reason the equation above we use the max operator.

Queries:

1. Which segments to target, provided that *rather* IF users are acceptable clientele?
2. Which segments to target, provided that only *very* IF are wanted?

Depending on the load, there are different rankings of segments. If initially some segments were in the same tier, for example, "very IF segments", some of them fall out of the tier, as the hedge operator is applied and the value of the membership function is squared (for "extremely IF"). The context, when to apply Query 1 or 2, becomes clarified via calculating f_{loaded} (infrastructure) and checking the applicability of different hedges. The method to obtain fuzzy heuristics is summarized to the sequence of the following steps.

Step 1: The IS clients in the customer base are revealed with the method [6] (the algorithm is reproduced as function *reveal_IS* clients in Algorithm 1), and each client is labeled with the IS/notIS descriptor.

Step 2: The propensity of a segment to contain IS clients is defined as the frequency of IS clients among its members and it is calculated from the data:

$$f_{IS}(segment_i) = frequency_{IS}(segment_i)$$

For linguistic convenience the term Infrastructure-Friendly (IF) is introduced and is set to be opposite to IS:

$$f_{IF}(segment_i) = 1 - f_{IS}(segment_i)$$

Step 3: The ranking of segments is carried out with respect to their IF quality and the hedged values of the mebership function are calculated: for all segments i, $f_{rather\ IF}(segment_i)$, $f_{very\ IF}(segment_i)$, and $f_{extremely\ IF}(segment_i)$. Given a hedge, which also codes the severity of the context, the segments fall into the different tiers (corresponding to one of the six fuzzy values): "fully in", "mostly but not fully in", "more or less in", and so on.

Step 4: Locally for the region under analysis, the infrastructure is assessed as *loaded*, *very loaded*, or *extremely loaded*, and thus the severity of the context is assessed. A ranking from Step 3 corresponding to a particular hedge is selected (as a leap of faith further verified in the next section).

The above is depicted as a flow chart in Fig. 1, and formalized as Algorithm 1. The reasoning behind combinatorial optimization is discussed in detail in [11]. The reasoning behind the function *label_ISclients* is discussed in detail in [6].

Algorithm 1: computation of the fuzzy recommendation heuristic.

Variables:

- clientSet: set of with IDs of clients;
- *I:* the set with geodemographic segments {$segment_1$, …, $segment_k$};
- *D:* the mobility data for a region that for each user contain client's ID, client's geodemographic segment, time stamps when the client generated traffic, and which antenna served the client.
- S_i: the number of subscribers that belong to a geodemographic segment *i*;
- $\Sigma_{all\ i}\ S_{i,t,j}$: the footprint, i.e. the number of subscribers that belong to a geodemographic segment *i*, at time moment *t*, who are registered with a particular cell *j*;
- C_j: the capacity of cell *j* in terms of how many persons it can safely handle simultaneously;
- ***x***: the vector with the scaling coefficients for the geodemographic segments or other groups such as IS clients;
- x_{IS}: the coefficent for the IS segment from the vector ***x***;
- $N_{t,j}$= number of users at cell j at time t;.

Input: data set D: <$user_{ID,}$ time stamp t, cell j>.

```
label_ISclients;
for i in I{
        ratherIF[i] = false
        veryIF[i] = false
  extremelyIF[i] = false
        degreeIS = frequency(userID_IS,I)
        degreeIF = 1- degreeIS
        if (degreeIF^(1/2) ≥ 0.9) then ratherIF[i]=true
  if (degreeIF^2 ≥ 0.9) then veryIF[i]=true
  if (degreeIF^3 ≥ 0.9) then extremelyIF[i]=true
}
```

function label_ISclients{

[I. Characterize each user with respect to her relative mobility.]

```
  for each user_ID {
    trajectory_ID = cell_t1, …, cell_t2016;
    relativeTrajectory_ID = N_t1,j, …, N_t2016,j;
    sortedTrajectory_ID = sort_decreas_or.(relativeTrajectory_ID);
    topHotSpots_ID = Σ_k=1..100(5%) sortedTrajectory_ID[k];
```

```
    userTopHotSpots = <userID, topHotSpotsID>
    }
   rankedUserList = sortdecreasing_or(rankedUserList)

    [II. Initialization.]

            xstressing = 0;
            setStressingUsers = ∅.

[III. Reveal the infrastructure-stressing clients.]

  While (xstressing = 0) do {
     tentativeStressingUsers = head1%(rankedUserList);
     setFriendlyUsers = bottom1%(rankedUserList);
     otherUsers = rankedUserList – tentativeSetStressingUsers -
     setFriendlyUsers;

     [Confirm the tentative labeling via combinatorial optimization.]
     I = {stressing, medium, friendly};
       {xstressing, xmedium, xfriendly} = combinatorial_optimization(I,D);

     IF (xstressing = 0), THEN {
     [take the field userID from tentativeStressingUsers]
     tentativeSetStressingUsers = tentativeStressingUsers< userID >;
     setStressingUsers = setStressingUsers +
     tentativeSetStressingUsers<UserID>

     D = D - Dstressing
  } [end of while]

  for id in <userIDs> do {
     if (id ∊ setStressingUsers) then label(id,"IS")
     else label(id,"notIS")
  } [end loop on id in <userIDs>]
} [end reveal_ISclients]

function combinatorial_optimization(I,D){
        solve
        Maximize        Σi∊{IF,other,IS} Si xi,
             subject to:

             for all j,t, Σi∊{IF,other,IS} Si,t,j xi ≤ Cj

} returns {xIF,xother,xIS}.

Output: array ratherIF[], veryIF[], extremelyIF[].
```

3.3 Query Simulation

In the above, when deciding which context should be applied, we relied on an intuitive rule: If the infrastructure is <*hedge*> loaded, then <*hedge*> IS segments are suitable to hire clients from. For example, in the case of a *rather* loaded infrastructure, *rather* IS segments are suitable targets. Given the expected success of the campaign, e.g. the campaign can attract 300 new clients or 1500 clients, it is possible to simulate the impact of the expected result on the infrastructure. A warning is thrown, if some antenna is overloaded, i.e. when the expected footprint by the segment violates a restriction for some segment *i*, at some antenna *j*, some time moment *t*:

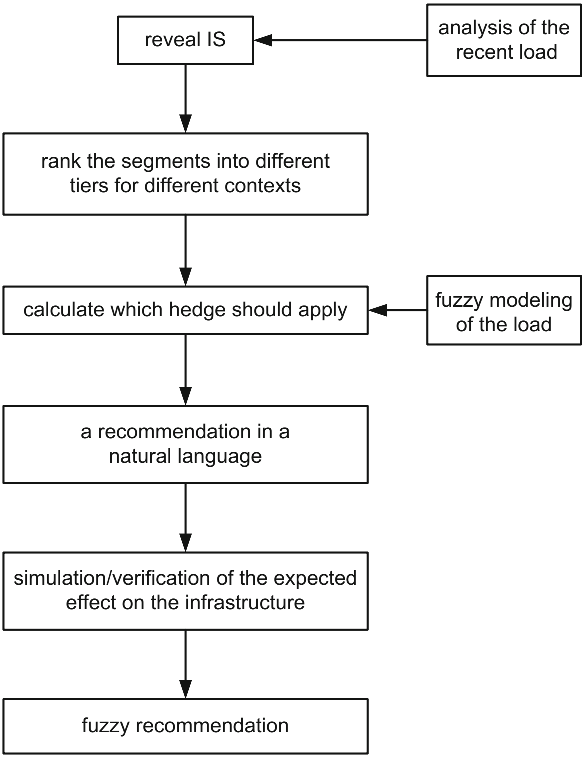

Fig. 1. The flow chart for the calculation of fuzzy recommendation for a marketing campaign.

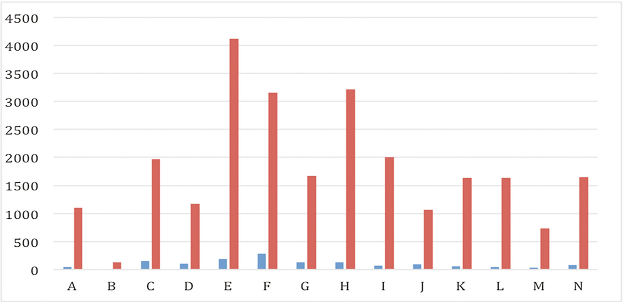

Fig. 2. The number of IS clients in different MOSAIC categories.

$$aS_{i,j,t} \leq C_j,$$

where α is a scaling coefficient:

$$a = \text{ expected number of new clients'(current number of clients)}^{-1}.$$

This is a justifiable approximation, since there is a high predictability in user trajectories within different segments, e.g. [12, 13].

4 Experiment

1. **Reveal the IS clients.** Applying the algorithm to reveal IS clients, we have added a field to the data set with the label IS or not IS as a descriptor for each client.
2. **Calculate degree of infrastructure-friendliness for each segment.** In the whole customer base, *7%* of subscribers were revealed to be IS [6]. We have obtained the distribution of the IS clients within the MOSAIC and Telenor segments and depicted them in Figs. 2 and 3, respectively. The degree of the infrastructure-friendliness is reported in Tables 1 and 2, for MOSAIC and Telenor segments, respectively.

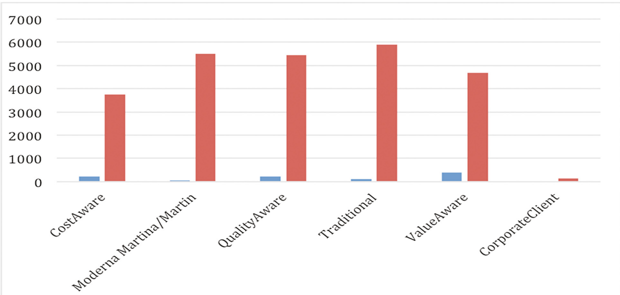

Fig. 3. The number of IS clients in different Telenor segments.

3. **Reasoning behind the queries.** Tables 1 and 2 simulate the reasoning behind the query results for different contexts (codified via a hedge) for the MOSAIC and Telenor segments, respectively. Each of the *14* MOSAIC classes qualifies as *rather IF,* which are those with $f_{IF}(i)^{1/2} > 0.9$. Once the customer base becomes larger and the spare capacity diminishes, only *very IF* will be wanted, which are those with $f_{IF}(i)^{2} > 0.9$. Out of those, only 9 segments qualify as *very* IF and five segments qualify as *extremely* IF ($f_{IF}(i)^{3} > 0.9$). The customer population was subjected to the same analysis with respect to Telenor segmentation. As follows from Table 2, each of the six Telenor segments is rather friendly, and there are four and three very and extremely IF segments, respectively.

Table 1. The reasoning behind the query results for the MOSAIC segments.

Segment	fIF(i)	$fIF(i)^{1/2}$	rather IF?	$fIF(i)^2$	very IF?	$fIF(i)^3$	extremely IF?
A	0.96	0.97	yes	0.92	yes	0.88	no
B	0.98	0.98	yes	0.96	yes	0.94	yes
C	0.93	0.96	yes	0.86	no	0.79	no
D	0.92	0.95	yes	0.84	no	0.77	no
E	0.96	0.97	yes	0.92	yes	0.88	no
F	0.92	0.95	yes	0.86	no	0.79	no
G	0.93	0.96	yes	0.86	no	0.79	no
H	0.96	0.97	yes	0.92	yes	0.88	no
I	0.97	0.98	yes	0.94	yes	0.91	yes
J	0.92	0.95	yes	0.86	no	0.79	no
K	0.97	0.98	yes	0.94	yes	0.91	yes
L	0.98	0.98	yes	0.96	yes	0.94	yes
M	0.96	0.97	yes	0.92	yes	0.88	no
N	0.95	0.97	yes	0.9	yes	0.85	no

Table 2. The reasoning behind the query results for the Telenor segments.

Segment	fIF(i)	$fIF(i)^{1/2}$	rather IF?	$fIF(i)^2$	very IF?	$fIF(i)^3$	extremely IF?
CA	0.94	0.97	yes	0.88	no	0.82	no
MM	0.99	0.89	yes	0.98	yes	0.97	yes
QA	0.96	0.92	yes	0.92	yes	0.88	no
T	0.98	0.87	yes	0.96	yes	0.94	yes
CC	0.92	0.8	yes	0.86	no	0.79	no
VA	0.97	0.91	yes	0.94	yes	0.91	yes

5 Results

When it comes to designing strategies of accommodating many more clients, being IS-prone for a segment is an important quality. We have studied the correlation between IS users and the geo-demographic segments, motivated by the fact that we can target the geo-demographic segments (MOSAIC and Telenor) in marketing campaigns. For different contexts, we have completed candidate rankings of geo-demographic segments, and, given the absence of other preferences, the top-tier segments are preferable. Which ranking out of several candidate ones is taken depends on the hedge calculated for the intensiveness of infrastructure exploitation. The simulation of the expected effect guarantees no false negatives, such as saying that certain segments are safe to hire from, but in fact that would lead to a service failure at some place and time in the network. For the implementation, please check https://sourceforge.net/projects/telenor-user-mobility/?source=navbar.

References

1. Zadeh, L.: Fuzzy logic and beyond - a new look. In: Zadeh, L., King-Sun, F., Konichi, T. (eds.) Fuzzy Sets and Their Applications to Cognitive and Decision Processes: Proceedings of the US-Japan Seminar on Fuzzy Sets and Their Applications, Held at University of California, Berkeley, California, 1–4 July 2014. Academic Press (2014)
2. Zadeh, L.A.: Fuzzy sets. Information and control **8**(3), 338–353 (1965)
3. Meyer, A., Zimmermann, H.J.: Applications of fuzzy technology in business intelligence. Int. J. Comput. Commun. Control **6**(3), 428–441 (2011)
4. Bilgiç, T., Türksen, I.B.: Measurement of membership functions: theoretical and empirical work. Fundam. fuzzy sets **1**, 195–232 (2000)
5. Tettamanzi, A., Carlesi, M., Pannese, L., Santalmasi, M.: Business intelligence for strategic marketing: predictive modelling of customer behaviour using fuzzy logic and evolutionary algorithms. In: Applications of Evolutionary Computing, pp. 233–240 (2007)
6. Sidorova, J., Skold, L., Lundberg, L.: (A) Revealing Infrastructure-Stressing Clients in the Customer Base of a Scandinavian Operator using Telenor Mobility Data and HPI Future SoC Lab Hardware Resources. Hasso Plattner Institute. A Technical report. https://www.researchgate.net/publication/312153463_Optimizing_the_Utilization_in_Cellular_Networks_using_Telenor_Mobility_Data_and_HPI_Future_SoC_Lab_Hardware_Resources
7. Haenlein, M., Kaplan, A.M.: Unprofitable customers and their management. Bus. Horiz. **52**(1), 89–97 (2009)
8. Debenham, J., Clarke, G., Stillwell, J.: Extending geodemographic classification: a new regional prototype. Environ. Planning A **35**(6), 1025–1050 (2003)
9. Sagar, S., Skold, L., Lundberg L., Sidorova, J.: Trajectory segmentation for a recommendation module of a customer relationship management system. In: The 2017 International Symposium on Advances in Smart Big Data Processing (in press). https://www.researchgate.net/publication/316657841_Trajectory_Segmentation_for_a_Recommendation_Module_of_a_Customer_Relationship_Management_System
10. Ragin, C.C., Rihoux, B.: Qualitative Comparative Analysis Using Fuzzy Sets (fsQCA) (2009)
11. Sidorova, J., Skold, L., Rosander, O., Lundberg L.: Discovering insights in telecommunication business from an interplay of geospatial and geo-demographic factors. In: The 1st International Workshop on Data Science: Methodologies and Use-Cases (DaS 2017) at 21st European Conference on Advances in Databases and Information Systems (ADBIS 2017). LNCS, 28–30 August 2017, Larnaca, Cyprus (2017, in press)
12. Song, C., Qu, Z., Blumm, N., Barabási, A.L.: Limits of predictability in human mobility. Science **327**(5968), 1018–1021 (2010)
13. Lu, X., Wetter, E., Bharti, N., Tatem, A.J., Bengtsson, L.: Approaching the limit of predictability in human mobility. Sci. Rep. **3** (2013)
14. Zadeh, L.A.: A fuzzy-set-theoretic interpretation of linguistic hedges (1972)
15. InsightOne MOSAIC lifestyle classification for Sweden, http://insightone.se/en/mosaic-lifestyle/. Accessed 15 Apr 2017