Social Specialization of Space: Clustering Households on the French Riviera

Denis Pallez, Mathieu Serrurier, Célia da Costa Pereira, Giovanni Fusco, Cristina Cao

To cite this version:

Denis Pallez, Mathieu Serrurier, Célia da Costa Pereira, Giovanni Fusco, Cristina Cao. Social Specialization of Space: Clustering Households on the French Riviera. Genetic and Evolutionary Computation Conference, Jul 2015, Madrid, Spain. 10.1145/2739482.2764665. hal-01344327

HAL Id: hal-01344327
https://hal.univ-cotedazur.fr/hal-01344327
Submitted on 19 Jul 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Social Specialization of Space: Clustering Households on the French Riviera

Denis Pallez  
Univ. Nice Sophia Antipolis,  
I3S, UMR 7271, 06900 Sophia Antipolis, France  
denis.pallez@unice.fr

Mathieu Serrurier  
University Paul Sabatier  
IRIT, UMR 5505 Toulouse, France  
serrurie@irit.fr

Célia da Costa Pereira  
Univ. Nice Sophia Antipolis,  
I3S, UMR 7271, 06900 Sophia Antipolis, France  
celia.pereira@unice.fr

Giovanni Fusco  
Univ. Nice Sophia Antipolis,  
Espace, UMR 7300, 06204 Nice, France  
Giovanni.Fusco@unice.fr

Cristina Cao  
Politecnico di Milano  
DASIU  
Milano, Italy  
cristina.cao@mail.polimi.it

ABSTRACT

The aim of this paper is to estimate the extent of social specialization of residential space within the French Riviera metropolitan area. Unlike classical approaches, where social groups are pre-defined through given characteristics of households, our approach determines clusters of households inductively. Socio-demographic characteristics of households are thus measured through 16 different indicators. Clustering is then carried out through the optimization of two distinct criteria. Simulated annealing, simple and multi-serting is then carried out through the optimization of two conflicting criteria. This is the reason why we adapt an evolutionary and a simulated annealing algorithm and apply them to a data set of households of the French Riviera extracted from the CERTU Household Mobility Survey of 2008.

Categories and Subject Descriptors

H.2.8 [Information Systems]: Database Applications—Data Mining, Spatial databases and GIS; I.5.3 [Pattern Recognition]: Clustering; J.4 [Computer Applications]: Social and Behavioral Sciences—Sociology

Keywords

NSGA-II, Simulated Annealing, Metropolitan Areas

1. INTRODUCTION

Social specialization of residential space within an urban area is the concentration of households according to some characteristics like social status, demographic characteristics, ethnicity, etc. in order to define urban subspaces. Understanding the logics of social specialization is of course crucial in order to define policies like urban planning or housing.

Geographers have traditionally used two approaches in order to study social specialization of urban space. The first is to pre-select social groups defined along lines of social status, ethnicity and position in the life cycle. Spatial concentration of these groups is later analyzed in order to detect eventual segregation patterns within the city. The second approach aims at characterizing the content of the resident population of neighborhoods without targeting a specific social group but by taking into account the whole population and all descriptors of households at the same time. Unlike these classical approaches, we want to both determine spatial concentration of specific groups of population in the urban space and to take into account all descriptors of households at the same time. This poses the problem of clustering households according to two conflicting criteria: maximum homogeneity of human content and maximum spatial difference within the metropolitan area. Classical clustering can not be directly applied due to optimization of two conflicting criteria. This is the reason why we adapt an evolutionary and a simulated annealing algorithm and apply them to a data set of households of the French Riviera extracted from the CERTU Household Mobility Survey of 2008.

2. OUR FRAMEWORK

Given a set of $n$ households $I = \{x_1, \ldots, x_i, \ldots, x_n\}$ where $x_j = \{x_{j1}, \ldots, x_{j9}\}$ represents the $j$-th household, and each $x_{ji}$ represents its $i$th feature value. We name $w_i$ the weight associated with the $i$-th attribute and $coem_j$, the weight corresponding to the $j$-th household in the sample.

In this paper we consider the clustering based on medoids: each cluster $C_{c1:k}$ is defined by a center $m_i \in I$. One common measure used for clustering is the intra-cluster distance between medoids and items and can be defined as:

$$Intra(C) = \frac{\sum_{i=1}^{k} \sum_{x_{ji} \in C} w_i \times d(m_i, x_{ji})}{\#I},$$  \hspace{1cm} (1)$$

where $d \in [0,1]$ represents the distance between two items weighted by $w_i$. Moreover, $\#I = \sum_{j=1}^{n} coem_j$. In the following the cardinality of a subset $A \subset I$ will be the sum of the $coem$s of the items it contains, i.e. $\#A = \sum_{x_{ji} \in A} coem_j$. However, the minimization of equation 1 is not sufficient to determine clusters of households that are spatially opposed within the metropolitan area. In order to ensure that socio-demographic differences among clusters also correspond to
A state responds to the medoids of the cluster. The neighborhood of choosing medoids in SA, a state is a vector of different clusters identified within the metropolitan population. Our Weighted Dissimilarity Index is a function that should move in order to spread it evenly within the metropolitan space. Our Weighted Dissimilarity Index is defined as

\[ WDI(S, C) = \sum_{j=1}^{k} \frac{\#C_j}{2\#I} \times \sum_{i=1}^{s} \frac{\#(S_i \cap C_j)}{\#C_j} - \frac{\#S_i - \#(S_i \cap C_j)}{\#I - \#C_j}. \]

where \( S = \{ S_1, \ldots, S_s \} \) be the \( q \)-partition of \( I \) in sectors. Household clustering will thus be carried out by simultaneously minimizing Equation 1 and maximizing Equation 2.

3. EXPERIMENTS

Simulated Annealing (SA) [4] and Genetic Algorithm (GA) have been adapted for optimizing a single objective combining both criteria using \( f(S, C) = \frac{WDI(S, C) - Extent(C)}{\text{Fast and Elitist MultiObjective Genetic Algorithm: Nsga-ii.}} \). For SA, a state is a vector of \( k \) distinct elements of \( I \) that correspond to the medoids of the cluster. The neighborhood of a state \( s = \{ c_1, \ldots, c_k \} \) obtained by switching one of the medoids in \( s \) with an item in \( I \). In order to be comparable with other strategies, we perform 400000 evaluation by choosing \( NBLOOP = 2000 \) and \( NBSTEP = 200 \). We fix \( t_0 = 0.1 \) and \( r = 0.995 \). \( s_0 \) is chosen randomly. Results are depicted in figure 1(a). For GA, a medoid-based representation clearly explained in [3] is used with reset mutation, tournament selection of size 10, a population size of 200 and a one point crossover. We then adapted NSGA-II [1] where objectives are considered independently.

4. DISCUSSIONS

Socio-demographic content of clusters extracted from Pareto front obtained by NSGA-II (PF17, PF12,PF48) is analyzed and presented in figure 1(b). Colors represent both household configuration and social status of the average cluster profiles in every solution. The combination of the five configurations with the four social strata produces twenty possible cases. A few additional cases account for particular profiles. Cluster labels G0-G9 are followed by the Dissimilarity Index (their weighted average gives the WDI value) and the cluster share within the total population. Intermediate Pareto front solutions are robust as WDI grows. For every configuration of households, two different social positions are normally identified. However, only PF17 identifies a cluster of couples (sometimes with children) of medium-high social status. Particular profiles with higher DI are detected by PF12 and PF48: couples of retirees of low social status and very segregated single retirees in social housing, respectively. Extreme solutions on the Pareto front are less interesting and even less pertinent when WDI is maximized at the expense of cluster internal heterogeneity.

5. CONCLUSION AND FUTURE WORK

For understanding social specialization on the French Riviera, we have adapted an NSGA-II algorithm and proposed a framework for clustering with a double optimization scope: minimizing intra-cluster distances and maximizing Weighted Dissimilarity Index seems to us the only way to empirically determine which categories of population are most opposed from a spatial point of view. The produced results seem in good agreement with expert knowledge of the social structure of the French Riviera. For the geographer, the task is now to project clustering results in geographic space in order to see what the ID values of the different clusters mean in terms of social morphology of the metropolitan area.

6. ACKNOWLEDGMENT

This work is a result of “Géo-Incertitude” project (2014-15) funded by CNRS within the research program HumaIn.

7. REFERENCES