Tourists'preferences ...

Giovanna Fancello, Olivier Cailloux, Alexis Tsoukiàs

June 25, 2017

Abstract

abstract
abstract
abstract
abstract

Contents

1 Introduction

- principal aim of the paper: to define public policies legitimated and that take account of people values and diversities.
- the tourist as one of the population in the city
- the possibility to access to some urban opportunity in order to improve their urban-quality of life (in a capability theory point of view)

Main Objective. The aim of this work is to define a method able to inspect values in space as regard to people variety in order to define public policies legitimated by different categories of people, policies that take account of people values and diversities. The purpose is to personalize the territorial offer of opportunities (to do or be in space) in respect to people values, preferences and needs. Especially, the research inquires how to collect and to analyse these features in order to be useful for the policy cycle and how to synthesize them in space considering the differences among individual values. In order to do this, we propose a replicable method able to inspect values that individuals give to different spaces.

Values in space can be learned by behaviours or declared preferences. In this study we use both type of data and experiment different methods and a replicable procedure for the analysis of preferences of tourists in the Alghero territory. Among the methods we used: multiple regression model; additive value function models (UTA family); multi task regression model;...

In the following paragraphs we first describe the problem, than we explore the literature in order to select the methods more suitable for our analysis. In the following section we explain the final method proposing an application to the case study. Finally, some policy analytics recommendation are given.

2 The problem. Alghero case study

Alghero is a city in the north east coast of the Sardinian Region (ITALY), characterized by an economic dependance from the touristic season for the development of the territory. This economic development can bring to positive but also negative consequences, given by that temporal economy that risk to influence the territorial development, as well as to the services and opportunities that has to be strategically improved in order to meet the needs of different population that live the city (CITARE MARTINOTTI). The activities undertaken by the people varies, among others, according to individual characteristics and interests [8], to spatial distribution of urban and territorial activities and to quality of the accessibility in the city [2]. Therefore, a strategic public policy should consider, among others, needs and preferences of tourist population in order to strategically improve the quality of life of the territory. Especially, a policy attentive to different people values can help in developing personal urban capabilities with the improvement of that urban opportunities that are most valued by tourist population (what people value [10].

In order to help policy makers in define new public policies, this work aims to define groups of people with common values of the space. A purpose different from the statistical analysis of people in respect to the correspondent tourist

socio-professional class and behaviours. What we try to do is a clusterization of population with similar values and interests in the city. This allows to define territorial development policies focused to specific tourist population and to improve the urban opportunities of the territory. Moreover, this analysis allows to define and to propose to different populations new doings and beings (urban capabilities) that they didn't already knows. Similar people in the same territory have similar urban capabilities, but may be they are not interested in them in the same manner, as they didn't value all these capabilities as important. In this work we want to bring to light similar values in space, similar capabilities that can be offered of people of the same group.

2.1 The survey

We analyse the tourist population in Alghero during the 2014 touristic "low season" (October-November), considering that Alghero's peek period is the summer season with highest tourist concentration between July and September. Data collection was carried out in this period because we want to catch urban limits and opportunities in a period of reduction of urban activities and not bathing season and have significant information for design a public policy aimed to deseasonalize this trend. These data was collected for the project XXX, lo inseriamo nel testo o in nota?. Two type of data were collected in the survey: interviews to inspect expectations and degree of satisfaction and GPS movement tracking to explore tourists' movements in the territory.

Especially, we dispose of 75 questionnaires representing **225 tourists** T described by a **set of attributes** A

$$A = \{gender, age, country, levelofstudy, profession, willingnesstopay\} \quad (1)$$

We know the **tourists' paths** in the territory, in a space described by **Coordinates** and **Time**. Let S denote the set of possible coordinates and τ the set of possible times. A path is a set of points $P \subseteq S \times \tau$.
Finally we define a set of **Categories of places** $C \in S$ that a tourist can choose in Alghero city.

In order to have coordinates divided in categories of space we did a spatial classification analysis of coordinates data (Figure 1). Especially we:

1. **Split the space** Subdivide the territory in different places p.
2. **Classify spaces.** Classify each place p in a Category c
3. **Count the time.** For each tourists' path we analyse the time spent for any category of place c.For each t we analyse how much time the he spent in it $x^t(c)$. The time spent in any category is given by the sum of the different ranges of time in this category of place $x^t(c_v)$. We consider that each tourist has 15 hours to spend in a day.

Let $C = \{c_1, \ldots, c_7\}$ denote the set of categories of places.

We call $i^t \in \mathbb{R}^+$ the total number of seconds a tourist t has spent on his trip in the different categories of places.

Starting with these considerations:

Figure 1: Classification of paths in respect to categories of places

Categories of places	Places
Environmental elements (local)	Lido, M. Pia, ...
Environmental elements (territorial)	Grotte di Nettuno, Punta Giglio, Spiaggia del Lazzaretto, ...
Historical and archaeological elements (local)	Cattedrale,Bastioni,Historical centre, ...
Historical and archaeological elements (territorial)	Fertilia, Castelsardo, Stintino, nuraghe Palmavera, ...
Cultural Elements	Theater, Cinema, Museum, ...
Food services	Restaurants, Market,...
Leisure	Waterfront, Public Gardens, Harbor, ...
Other	Stay in the Hotel, friends' home, Route from one place to another, ...

Table 1: **Categories of places**

Definition 1 *A **vector path** $x : C \to R^+$ is a function that maps each category of places to a number of seconds. We associate to each tourist $t \in T$ a vector path x^t. Let $X = \mathbb{R}^{+C}$ denote the set of all possible vector paths.*

Among all the vector paths the tourist chose a path in respect to his preferences and values.

Definition 2 *A **preference relation** is a binary relation over the set of possible vector paths $S \subseteq X \times X$,.*

In detail, a preference relation is given when the tourist can freely choose a path to another in respect to his individual characteristics (age, gender,...) and to his personal (income,...) and spatial resources (means of transport, ...) [10].

Considering this, our objective is to associate to each tourist $t \in T$ a preference relation S_t which represents, according to our model, the way the tourist evaluates the possible paths in the Alghero territory, i.e. the way the tourist values the space. This preference relation can also reveal important elements for the design of new paths in respect to the "demand? (like marketing?)" of the tourist population in Alghero. Let $X^t \subseteq X$ denote the set of all vector paths that can be freely chosen by t.

Finally, we know the **tourist declared preferences** (why they choose to stay in Alghero and what they want to do). The criteria the tourists used to choose to visit Alghero that are ordered by importance with an evaluation scale. Let's have for each tourists' criteria $zt \in Zt$ a value $n \in \xi$ with $\xi\{1, 2, 3, 4, 5\}$

$$Zt = n\{Economy, Environment, Weather, Food, Culture, Recreation, Entertainment, Study, Work, Relax, Friendsandrelatives, others\} \quad (2)$$

2.2 Synthesis of Data

In synthesis the set of data is given by the:

T Set of tourists (t a tourist)

Z Set of aspects on which tourist has declared a degree of interest (Zt)

$d^t : Z \to \{1, 2, 3, 4, 5\}$ A function such that $d^t(z) \in \{1, 2, 3, 4, 5\}$ indicates, for tourist $t \in T$ and aspect $z \in Z$, the degree of interest on aspect z as declared by t

$i^t \in \mathbb{R}^+$ Intensity corresponding to t: the total number of seconds tourist t has spent on his trip in the different categories of space.

C_V The categories of places in Alghero *that can be visited*, thus *including* staying home.

$x^t : C_V \to \mathbb{R}^+$ The vector path chosen by tourist t that associates a number of seconds to each category of place. By definition, $\sum_{c \in C_V} x^t(c) = i^t$.

x^t E' l'insieme DI PUNTI CHE FORMANO IL PERCORSO FATTO DA UN TURISTA IN UNA DETERMINATA CATEGORIA DI LUOGO, iN REALTà IL PERCORSO è DIVERSO, NOI ABBIAMO PRESO TUTTI I PUNTI INTERNI AD UN LUOGO

2.3 A simple goal

DA PROPOSTA DI OLIVIER

Our goal is to predict x^t given d^t and an intensity: a number of seconds that indicates how long the tourist intends to visit for. We are allowed to use $T_1 \subseteq T$ as training data, and must use $T_2 \subseteq T$ as test data. Given a predictor $P : \{1, 2, 3, 4, 5\}^Z \times \mathbb{R}^+ \to \mathbb{R}^{+C_V}$, the quality of P is the sum, on the test data, of the distance, using L2 norm, between the prediction and the real path: quality of $P = \sum_{t \in T_2} \|P(d^t, i^t) \text{-} x^t\|$.

Here below we list a few possible ways of building predictors, starting from the simplest ones:

- **central prediction** We compute the centre (using L2) of $(x^t)_{t \in T_1}$ and constantly predict this.
- **Unlinked regression** Build some regression model between $\{1, 2, 3, 4, 5\}^Z \times \mathbb{R}^+$ and each dimension of $\mathbb{R}^{+C_V}$, separately.
- **Multi-task Regression**Build some regression model between $\{1, 2, 3, 4, 5\}^Z \times \mathbb{R}^+$ and $\mathbb{R}^{+C_V}$. WORK IN PROGRESS WITH ELASTICNET MIXING PARAMETERS
- **UTA+ Method**
- **Learned transformation**Same idea as Transformed space, but we learn the partial value functions using T_1.
- **Guided regression**Same idea but we (somehow) use knowledge about the link between Z and C_V.
- **Transformed space**We fix (a priori) a set of partial value functions $u_c : \mathbb{R}^+ \to [0, 1]$, one for each category of place. The vector $(u_c \circ x^t) : C_V \to [0, 1]$ represents the values associated by tourist t, given its path, to each

category of places. We learn a regression model (using normal or guided regression) between $\{1,2,3,4,5\}^Z \times \mathbb{R}^+$ and latent weights w^t, in order to maximize $w^t \cdot (u_c \circ x^t)$, the value of x^t.

- **More ideas**See article Siskos. See utilitaristic regression from Eyke.

3 Background

In order to define a method to analyse values in space for different touristic populations we can use several methods. In literature we can distinguish almost two fields of research that interested in our research.

- Analysis of tourist behaviours in space
- Analysis of people values (in respect to declared preferences or judgments)

These two fields have to be combined for our research problem, especially we can consider the first analysis of tourist behaviour as the method to have data for the analysis of people values.

3.1 Tourist behaviours in space

The most common way to study tourists' behaviours in time and space are various methods of diary (re)construction [13]. New possibilities have arisen by the development of automated tracking, above all the satellite-based (e.g. GPS) technologies for automatic position tracking with high time and spatial resolution. Even if methods for surveying and analysing spatio-temporal behaviour are, of course, becoming highly developed in transportation research and in social sciences in general, comparatively little attention was being paid to the spatial and temporal behaviour of tourists, and systematic studies taking advantage of the technological developments offered by the high precision position tracking are still relatively few [11];[13]. In recent years scholars have started to break ground in this specific domain of applied research, experimenting and exploring advantages and disadvantages of satellite-based positioning technologies for the study of tourists' spatio-temporal behaviours [12]. Different methods have been reported for tracking tourist spatio-temporal behaviour [8]:

- direct observation of tourists' activities, with interviews or remote observations;
- time-space budget techniques [9] which analyse tourists' activities within destina-tions by using diaries, questionnaires and interviews;
- video-based tracking analysis, used to track tourists movements through video footages;
- smartphones with apps using positioning systems;
- specialized GPS tracking devices
- land-based tracking systems that collect data thanks to radio technology.

Many studies report observations on a small scale, focusing on particular urban areas or activities that have a clearly defined entry and exit point, such as natural parks or historical centres [13]. Also many urban contexts have been analysed, among which Rome [4], Lago del Garda [3], Canberra and Sydney [6], Salzburg [8]. On the other side, there are studies that attempt to analyse tourist behaviours at very large scale (e.g. on the national scale, by using mobile phones data [1]. Different statistical and visualisation techniques may be employed to analyse and represent such spatial data. Frequently spatial temporal data are used only to de-scribe the tourists track in maps. Sometimes these spatial maps are combined further information collected through questionnaires. This allows to link spatial temporal behaviour to a specific tourist category of population. Furthermore, of interest are studies trying to predict visitors whereabouts, how long they will stay in a place, or carry out a specific activity [5].

3.2 People values and preferences

Value theory

3.2.1 Unlinked regression

We start building predictors with the multiple regression model.

The definition of x^t depends on the number of seconds in a category of place(HO BISOGNO DI SPIEGARE IL PROBLEMA DAL PUNTO DI VSTA FORMALE?). This means that we have two dependent elements that are influenced by the preferences of people (d^t). We cannot use a standard multiple linear regression model, for this reason, in order to directly predict x^t, we develop an unlinked multiple regression model aimed to investigate the relation among the different elements separately.

Especially, we develop a multiple regression model between $\{1,2,3,4,5\}^Z \times \mathbb{R}^+$ and each dimension of $\mathbb{R}^{+C_V}$, separately. Results (see Table 2) show low relations between the dependent variable and the independents ones. For example, the regression model between $\{1,2,3,4,5\}^Z \times \mathbb{R}^+$ and $\mathbb{R}^{+EnvLocal}$ shows a R-squared of 0.2397 and a low significance of independent variables (the *ttest*).

We also develop a regression model isolating the set of aspects $z \in Z$ that have much more significance, but also in this case we obtain low significance values.

This means that this method is not able to building predictors with this type of data.

4 UTA+ Method

As second method we present a multicriteria model that combines preference learning methods with the unsupervised machine learning in order to define the best territorial offer for each category of people. It is a sort of territorial marketing aimed to satisfy what people value as important to do [10] in the territory. What we want to inspect is not only what people do in respect to different needs and values, but to understand which are the set of things that people can do in the territory starting from their preferences and needs.

The model proposed is composed of different phases:

	Estimate	Std. Error	t value	Pr(>|t|)	signif.
(Intercept)	7226.3916	3559.1948	2.03	0.0472	*
cost	151.3463	484.6296	0.31	0.7560	
environment	-803.0865	471.5940	-1.70	0.0942	
weather	-200.6880	598.4546	-0.34	0.7386	
health	946.1080	710.4019	1.33	0.1884	
food	-1581.7536	589.8456	-2.68	0.0097	**
events	1825.5161	744.7139	2.45	0.0174	*
sport	128.9466	614.9811	0.21	0.8347	
fun	-165.3657	848.8539	-0.19	0.8463	
study	-310.8244	946.6043	-0.33	0.7439	
work	-1562.5287	715.1637	-2.18	0.0332	*
relax	-135.9579	500.6331	-0.27	0.7870	
family	714.5345	552.7549	1.29	0.2015	
Signif. codes				0 * * * 0.001 * *0.01 * 0.05.0.11	
Residual standard error				4922 on 55 degrees of freedom	
Multiple R-squared				0.2397, Adjusted R-squared: 0.07384	
F-statistic				1.445 on 12 and 55 DF, p-value: 0.1742	

Table 2: **Unlinked regression. Example by Environmental Local places**

1. **data collection** NON SO SE INSERIRLO DI NUOVO QUI
2. **Multicriteria analysis of preferences**. In order to do this we use the UTA+ algorithm.
3. **Selection of relevant factors for personalization**.

   Cluster analysis of the utility function

   Rough set analysis
4. **Cluster analysis**. Cluster analysis of the results to define the different profiles of tourists
5. **Recommendation**

4.1 Multicriteria analysis of preferences

We consider that each tourist is a decision maker that needs an aid in order to understand his priorities and values in terms of what he can do in the Alghero territory. To model this problem and define tourists' values we use the UTA algorithm. UTA was originally proposed by E.Jacquet-Lagraze and J.Siskos in 1982 [7]. Especially, we use an implementation of the UTA method, the UTA+ algorithm [**?**]. This method solves problems of multicriteria choice and ranking on a set A of alternatives.

Spiego in geneale come funziona UTA?

Utility function can express user's preferences in respect to a set of alternatives. The UTA+ method consents to give a value of each alternative starting from a ranking of alternatives. The ranking is a weak order that uses the preference (P) and indifference (I) relations only among the reference alternatives.

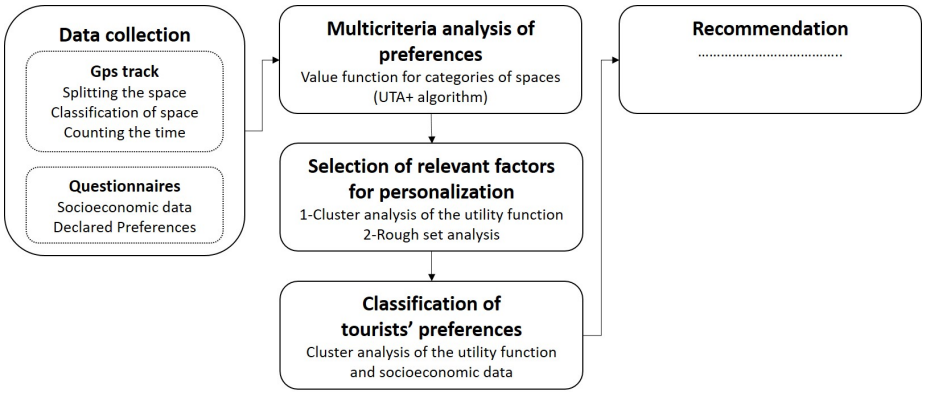

Figure 2: Schema

We represent preference relations S using additive value functions. Let u^t denote the value function of tourist t. We define S_t from u^t as follows:

$$(x_1, x_2) \in S_t \text{ iff } u^t(x_1) \geq u^t(x_2). \tag{3}$$

Our objective is thus to obtain u^t from the path data.

We assume that u^t can be represented as a weighted sum of partial value functions:

$$u^t(x) = \sum_{c \in C} u_c(x_c) w_c^t, \tag{4}$$

where $w^t \in [0, 1]$ represents the weight that the tourist t gives to the category of place c.

We assume for now that the partial value functions $\{u_c\}$ and the weights may depend on people values and are different for each tourist t.

Given a category $c \in C$, we define a partial value function $u_c : \mathbb{R}^+ \to [0, 1]$. The number $u_c(x_c)$ represents the value we assume the tourist gives to spending x_c seconds in the category c, not taking into account the partial weight w_c^t.

We define each u_c as a two linear pieces increasing function determined by the UTA method.

From now we know just the behaviours of each tourists t and their declared preferences. In order to learn the set of tourists' preferences we consider the relation among the path chosen by the tourist and a set of **outstanding paths**. The outstanding paths are defined by a combination of categories of places C and time τ they ideally spend in (with a maximum of 15 hours of visit). A tourist's preference is verified if the tourist t prefers a path internal to his choice set $x_1 \in X^t$ in respect to another $x_2 \in X^t$.

Definition 3 *We define an outstanding path as a path that a tourist can freely choose in respect to his personal characteristics and spatial and personal resources (TO BE COMPLETED . . .).*

We define two possible set of outstanding paths that are valid for all tourists:

- **Hard set**: all the tourist paths can be considered as outstanding paths. This means that a tourist t prefers his path to all the other tourists' paths.

But, as each tourist has his personal set of possible paths (in a capability framework), it is possible that $X^{t_1} \cap X^{t_2}$ or that $X^{t_1} \neq X^{t_2}$.

- but not all the people has the same characteristics and the same resources, so we need to define different outstanding paths.

- **Soft set**: a set of outstanding paths (see Table 3), not expensive and reachable on foot. This permits to consider a sort of basic set of paths (that every tourist can freely choose). We define ten outstanding paths $x_{O_1}, \ldots, x_{O_{10}}$ that we assume tourists have considered because are standard paths available for all people. We assume that the tourist t prefers the path x^t he has chosen to each of the outstanding paths: $u^t(x^t) \geq u^t(x_{O_i}), 1 \leq i \leq 10$.

Op	EL	ET	HT	HL	FS	LS	CL	OT
Op1	6	0	0	0	3	1	1	4
Op2	0	7	0	0	3	0	0	5
Op3	0	0	0	2	2	6	2	3
Op4	0	0	0	4	1	3	0	7
Op5	0	0	7	3	3	0	0	2
Op6	6	4	0	1	3	0	1	0
Op7	7	6	0	0	1	1	0	0
Op8	2	0	0	6	2	3	2	0
Op9	2	0	6	0	1	3	2	1
Op10	1	0	6	0	3	2	2	1

Table 3: **Soft set. Number of hours for category of places in the Outstanding Paths (Op).**

4.2 Calibration of the UTA+ model

In order to calibrate the UTA+ algorithm we consider:

$x^t : C_V \to \mathbb{R}^+$. The vector path chosen by tourist t that associates a number of seconds to each category of place.

Outstanding paths We assume that each tourist has considered the softset $\{1, \ldots, 10\}^{Op}$ of outstanding paths, not expensive and reachable on foot, that every tourist can freely choose.

Preferences We assume that the tourist t prefers (P) the path x^t he has chosen to any of the outstanding paths: $u^t(x^t) \geq u^t(x_{O_i}), 1 \leq i \leq 10$.

weak order The weak order considered is A P Op^1 I Op^2 I Op^3 I Op^4 I Op^5 I Op^6 I Op^7 I Op^8 I Op^9 I Op^{10}

ideal and antideal alternatives

Using these data we did different tests in order to choose the best parameters for the UTA+ algorithm INSERIAMO QUESTI TEST OPPURE LI NOMINIAMO SOLAMENTE?:

- **test1**: for each partial value, we consider as minimum and maximum values of i^t the minimum and maximum time spent by the considered alternatives.
- **test2**: for each partial value, this test considers as the minimum and maximum values of i^t the minimum and maximum time that it is possible to spend in a day by tourists $\mathbb{R}^+ \rightarrow [0, 54.000] seconds$. We consider fort his test the $\{Op1, Op2, Op3, Op4, Op5\}$ Outstanding paths.
- **test3**: for each partial value, this test considers as the minimum and maximum values of i^t the minimum and maximum time that it is possible to spend in a day by tourists $\mathbb{R}^+ \rightarrow [0, 54.000] seconds$. We consider fort his test the $\{Op1, \ldots, Op10\}$ Outstanding paths. The test includes ideal and antiideal alternatives,...).

See the appendix for other tests. In Table 5

(test)	w^{EL}	w^{ET}	w^{HL}	w^{HT}	w^{FS}	w^{LS}	w^{CL}	w^{OT}
(test1)	0.015	0	0	0.137	0.330	0.092	0	0.426
(test2)	0.269	0	0	0	0	0	0	0.731
(test3)	0	0.104	0	0.078	0.036	0.050	0.131	0.600

Table 4: **Test for the calibration of UTA+ algorithm**

Results are different in respect to the parameters chosen for the analysis. But, when we improve the set of input data results are more stable. For this reason we chose to do a the analysis of data with the parameters of the test3 by using the **soft set** of Outstanding paths. Finally, we define the value function for each tourist with the UTA+ software giving attention to the Kendalls coefficient and considering Ideal and Anti-idel alternatives. How much do I deepen this? Do I insert some utility function image??

4.3 Selection of relevant factors for personalization

This phase aims to select the relevant factors that influence the choice of a particular path by the tourists. In fact, different people in the same place can choose different paths in respect to personal and contextual factors (in a Capability Approach framework inserire lo schema del capability approach?forse meglio nell'articolo teorico?). Here we want to focus on personal factors (gender, age, profession, ...) that can influence the choice of people. In order to inspect the relation between the path chosen x_t and these factors we propose a two phase analysis. In the first step we did a cluster analysis of the only preferences (max and medium values of utilities). In the second phase we did a rough set analysis among the class of the cluster and the personal information of tourists.

4.3.1 Cluster analysis of the utility function

In this phase we classify the tourists using the preferences come to light from the UTA+ model. Especially, we did a cluster analysis of the value functions (central and maximal point) that each tourist gives to spend time in the different categories of spaces. SPIEGO I DIVERSI METODI DI CLASTER? kmeans,

PAM, Hclust? We did some tests in order to define the better number of cluster for different type of cluster analysis. Tests include the relative cluster validation and an internal validation of cluster.

cluster validation . This method aims to evaluate the clustering structure by varying different parameter values for the same algorithm.

internal validation . This method analyses the internal information of the clustering process in order to understand the goodness of the cluster in respect to internal data. For this step we use the silhouette analysis that measures how well an observation is clustered and estimates the average distance between clusters. The silhouette plot displays a measure of how close each point in one cluster is to points in the neighboring clusters

Especially, we uses the Model Based Approach with Bayesian Information Criterion (BIC), the average silhouette method and the Dindex. We can see the results of the BIC model in the Figure3. This model suggests a model VEV (ellipsoidal, equal shape) with 3 components.

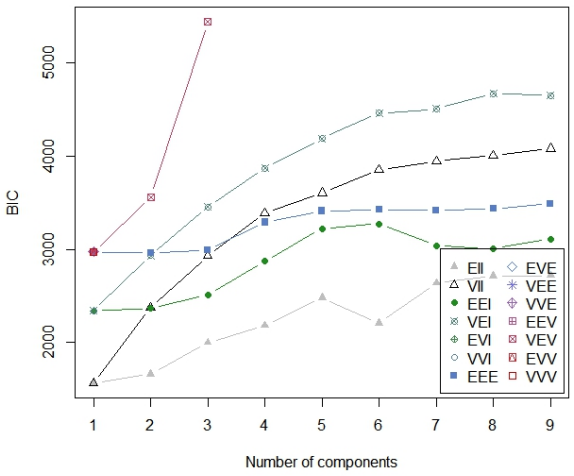

Figure 3: The Bayesian Information Criterion. A large BIC score indicates strong evidence for the corresponding model.

Also, the average silhouette method for the hierarchical cluster suggests a model with 3 components (Figure 4). It measures the quality of a clustering, determining how well each object lies within its cluster. A high average silhouette width indicates a good clustering. This method computes the average silhouette of observations for different possible numbers of cluster (k). The optimal number of clusters k is the one that maximize the average silhouette over a range of possible values for k (Kaufman and Rousseeuw [1990]).

Finally we run the R. (cran Project) Nbclust function that computes an high number of indices for cluster validation. According to the majority rule, the best number of clusters suggested by Nclust is 3, with 7 indices suggesting that model.

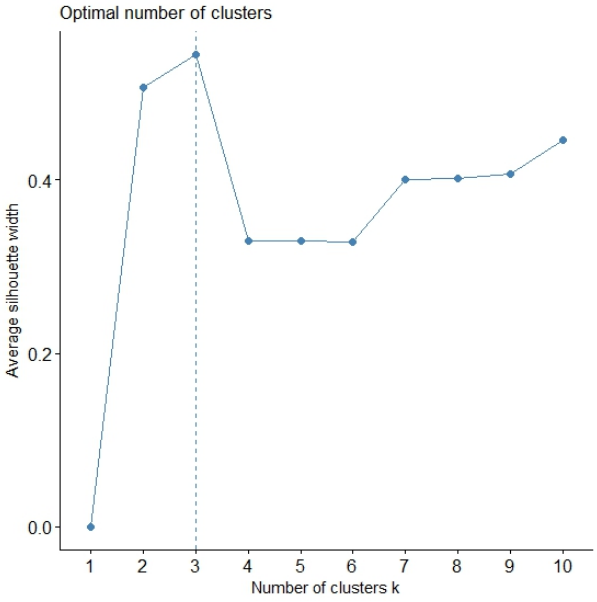

Figure 4: Average Silhouette for Hierarchical clustering

In the plot of D index (Figure5), we seek a significant knee (the significant peak in Dindex second differences plot) that corresponds to a significant increase of the value of the measure.

As results all three methods suggest a model with 3 component. Among them, we choose to use the classification suggested by the Nbclust method. INSERIAMO UNA TABELLA CON LA CLASSIFICAZIONE?

4.3.2 Rough set features selection

This phase aims to find a subset of features which have the same quality as the complete feature set. In other words, the purpose of the rough set feature selection method is to select the significant features and eliminate the not relevant ones. We use this method to find the socioeconomic features that are useful for the classification of tourists in respect to the cluster families (previous step). SPIEGO COME LAVORA LA rough SET ANALYSIS?

So, starting from the knowledge of Y we select the rough set of elements $\in A$

$$A = \{gender, age, country, levelofstudy, profession, willingnesstopay\} \quad (5)$$

The final rough set eliminated the $levelofstudy$ feature from A, creating a feature subset of 6 attributes Y:

$$A^1 = \{gender, age, country, profession, willingnesstopay\} \quad (6)$$

4.4 Cluster Analysis

In this phase we did a cluster analysis of the max and medium values of utilities (COME LE CHIAMO?) together with the socioeconomic A^1 data in order to

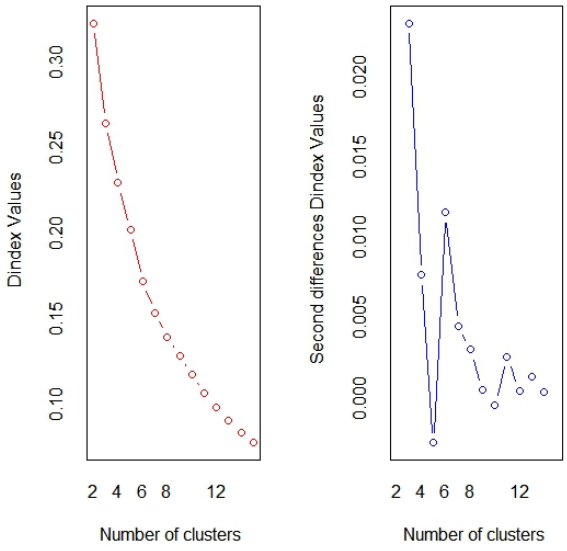

Figure 5: The Dindex

find homogeneous groups of people that value in the same manner the space. Especially, groups of people that give the same importance to spend time in particular places of the Alghero's territory.

SPIEGARE IL METODO gower USATO PER DETERMINARE LE DISTANZE DI VALORI CATEGORICI

4.4.1 Hierarchical cluster

We did a first cluster analysis with the hierarchical method using Ward classification. SPIEGARE IN COSA CONSISTE IL METODO WARD E PERCHè è FRA I MIGLIORI.

Results show a dendrogram with 5 distinct cluster as suggested by the silhouette widt X (Figure 6 and Figure 7). COMMENTARE I RISULTATI E MOSTRARE UNA SINTESI DEI CLUSTER E DELLE LORO CARATTERISTICHE. FARE GRAFICI DI SINTESI DELLE CARATTERISTICHE PERSONALI E DISEGNO DELLE FUNZIONI DI UTILITà.

The resulting clusters show different compositions of individuals and of preferences.

4.5 pam

DESCRIVIAMO ANCHE QUESTI RISULTATI? IL MODELLO MIGLIORE è QUELLO CON HIERARCHICAL ANALYSIS MA POSSIAMO MOSTRARE ACHE QUESTI RISULTATI CHE MOSTRANO UNA SUDDIVISIONE IDEALE DI 8 CLUSTER. iL METODO DI SUDDIVISIONE è DIVERSO DAL

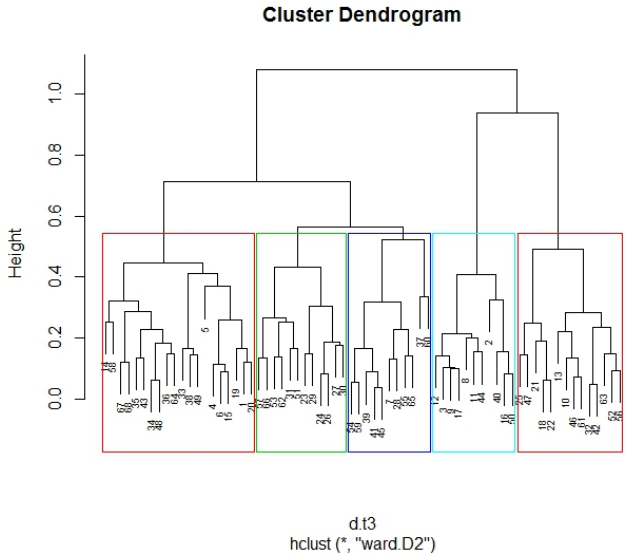

Figure 6: Dendrogram

PRECEDENTE, PER QUESTO RESTITUISCE RISULTATI DIVERSI. (see the word document)

4.6 A multicriteria recommender system?

A partire da questi dati trovare un modo per:

- vautare l'opportunità di una azione o politica rispetto alle funzioni di utilità dei miei turisti
- generare nuove politiche a partire dall'analisidella differenza tra domanda e offerta
- e se volessi cambiare i turisti?

4.7 some tests

- **test1**: Fare una analisi di regressione multipla per definire la relazione tra i risultati ottenuti con la cluster analysis e le componenti socioeconomiche (HO FATTO ALCUNI TEST, DA VERIFICARE CON OLIVIER)
- **test2**: Multirating collaborative filtering? non so se posso usarla con dati misti ma provo a studiare la questione.
- **test3**: Analisi di regressione tra le preferenze dichiarate e le utilità.

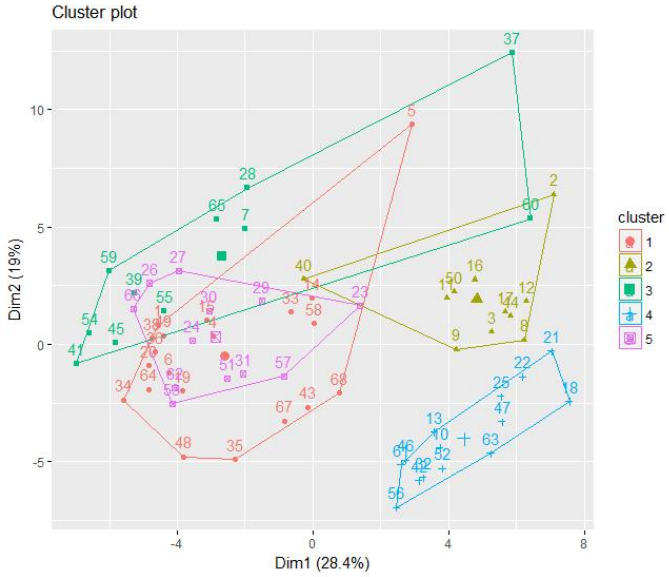

Figure 7: Dendrogram

4.8 Recommendation

References

[1] Rein Ahas, Anto Aasa, Ülar Mark, Taavi Pae, and Ain Kull. Seasonal tourism spaces in Estonia: Case study with mobile positioning data. *Tourism Management*, 28(3):898–910, 2007.

[2] I. Blečić, A. Cecchini, T. Congiu, G. Fancello, and G.A. Trunfio. Evaluating walkability: a capability-wise planning and design support system. *International Journal of Geographical Information Science*, 29(8), 2015.

[3] Andrea Bruno, Emanuela Gasca, Stefania Mauro, Giuseppe Pollichinob, Sara Levi Sacerdotti, and Fabrizio Stupino. Understanding tourist behaviour in wide areas using GPS technologies. *ENTER 2010 Conference on Information and Communication Technologies in Tourism*, 2010.

[4] Francesco Calabrese and Carlo Ratti. Real Time Rome Real Time Rome. *Networks and Communication Studies NETCOM*, 20(3-4), 2006.

[5] Prem Chhetri, Jonathan Corcoran, and Colin Arrowsmith. Investigating the Temporal Dynamics of Tourist Movement: An Application of Circular Statistics. *Tourism Analysis*, 15(1):71–88, 2010.

[6] Deborah Edwards, Tony Griffin, Bruce Hayllar, and Tracey Dickson. Using GPS to Track Tourists Spatial Behaviour in Urban Destinations. 2009.

[7] E. Jacquet-Lagreze and J. Siskos. Assessing a set of additive utility functions for multicriteria decision-making, the UTA method. *European Journal of Operational Research*, 10(2):151–164, 1982.

[8] Lenka Kellner and Roman Egger. Tracking tourist spatial-temporal behavior in urban places, a methodological overview and GPS case study. *Information and Communication Technologies in Tourism 2016*, pages 481–494, 2016.

[9] G Pearce and Douglas G Pearce. Tourist Time-Budjets. *Annals of Tourism Research*, 15(1988):106–121, 1988.

[10] Amartya Sen. Development as Freedom. *Oxford Press*, pages 1–50, 1999.

[11] Noam Shoval and Michal Isaacson. Tracking tourists in the digital age. *Annals of Tourism Research*, 34(1):141–159, 2007.

[12] Noam Shoval and Michal Isaacson. *Tourist Mobility and Advanced Tracking Technologies*. Routledge, 2010.

[13] Noam Shoval, Michal Isaacson, and Prem Chhetri. GPS, Smartphones, and the Future of Tourism Research. *The Wiley Blackwell Companion to Tourism*, pages 251–261, 2014.

5 Appendix

	Test	wEL	wET	wHT	wHL	wFS	wLS	wCL	wOT
range 0-max +20perc	t1	0,000	0,000	0,000	0,000	0,000	0,000	0,000	1,000
range automatico	t2	0,155	0,155	0,097	0,117	0,000	0,155	0,000	0,321
range(0-max+20perc)+ideal	t3	0,000	0,048	0,217	0,000	0,086	0,000	0,213	0,437
range automatico+ideal	t4	0,155	0,155	0,097	0,117	0,000	0,155	0,000	0,321
(0-54000)+ideal e anti-ideal	t5	0,160	0,145	0,218	0,000	0,000	0,226	0,000	0,252
(0-54000)+6 alternative+preferenze fra alternative	t6	0,140	0,129	0,150	0,051	0,000	0,167	0,000	0,363
(0-54000)+7 alternative+preferenze fra alternative	t7	0,106	0,097	0,086	0,119	0,197	0,137	0,000	0,259
(0-54000)+8 alternative+preferenze fra alternative	t8	0,103	0,109	0,124	0,093	0,067	0,089	0,183	0,231
(0-54000)+9 alternative+preferenze fra alternative	t9	0,105	0,106	0,115	0,098	0,104	0,105	0,128	0,240
(0-54000)+10 alternative+preferenze fra alternative	t10	0,075	0,065	0,081	0,046	0,211	0,142	0,000	0,379
(0-54000)+18 alternative+preferenze fra alternative	t11	0,118	0,118	0,118	0,118	0,118	0,118	0,118	0,172
(0-54000)-10 alternative utilità - 18alternative conferma utilità	t12	0,125	0,125	0,125	0,125	0,125	0,125	0,125	0,125
(0-54000)-10 alternative utilità - 18 alternative conferma utilità	t13	0,075	0,065	0,081	0,046	0,211	0,142	0,000	0,379

Table 5: **all Test**

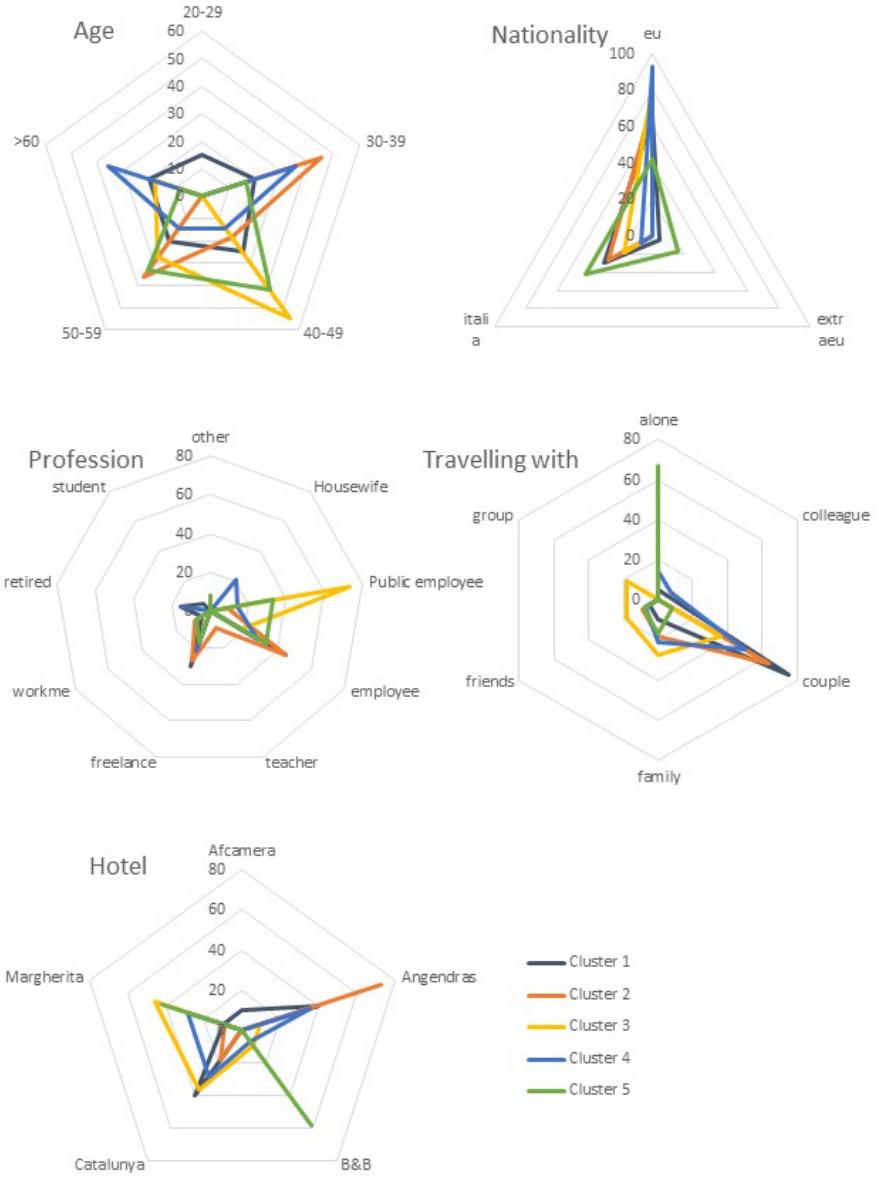

Figure 8: Personal Characteristics

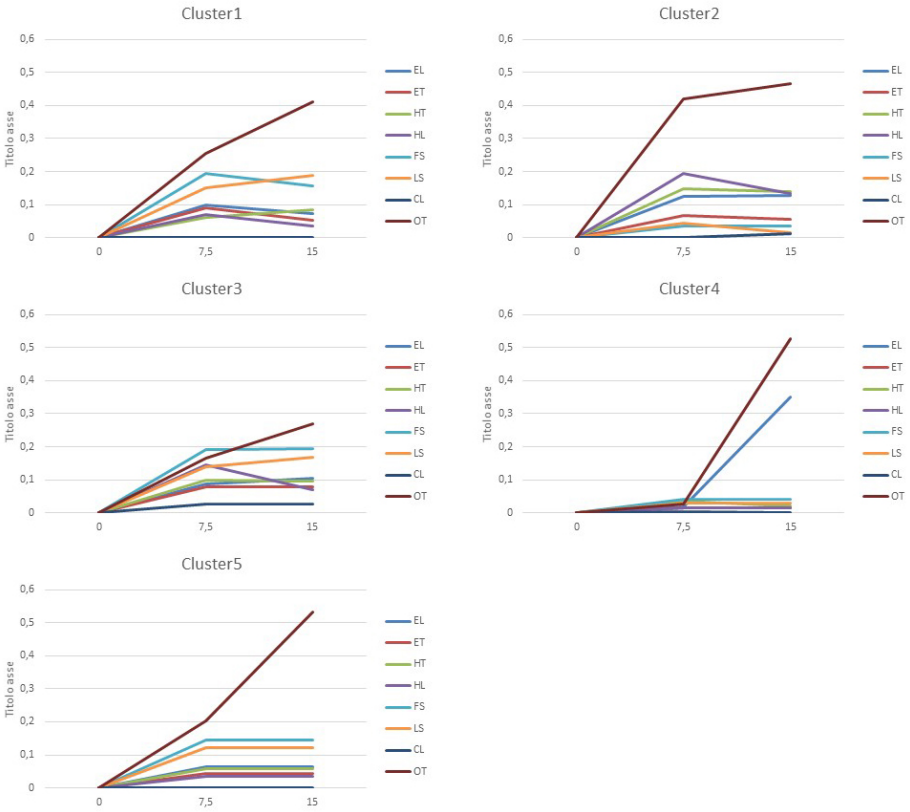

Figure 9: Utility Functions