Tourists'preferences ...

Olivier Cailloux, Giovanna Fancello, Alexis Tsoukiàs,...

30th December, 2016

Contents

1. Introduction

- state of the art and aim of the paper
- 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. To define the tourists values in space in respect to the correspondent tourist socio-professional class and behaviours.

2. Alghero case study

Case Study. The city of Alghero and his territory in 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 is carried out in a touristic low season period because we want to catch urban limits and opportunities in a period of reduction of urban activities and not bathing season and in order to have significant results for a tourist public policy aimed to deseasonalize this trend. **Data** We have 75 questionnaires representing **225 tourists** T described by a **set of attributes** A

$$A = \{gender, age, country, levelofstudy, profession, willingnesstopay\}$$

We know **tourists' paths** in the territory, in the 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. Let $C = \{c_1, \dots, c_7\}$ denote the set of categories of places.

Starting with these considerations, we define a **vector path** $x : C \rightarrow R^+$ as 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.

We define a preference relation $S \subseteq X \times X$ as a binary relation over the set of possible vector paths.

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. 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, ...) (SEN). Let $X^t \subseteq X$ denote the set of all vector paths that can be freely chosen by t.

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

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

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**

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:

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

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

We assume for now that the partial value functions $\{u_c\}$ are the same for each tourists $t \in T$ (or not?), but the weights may depend on 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 (TO BE COMPLETED …) .

In order to define a set of possible 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. A tourist's preference is verified if the tourist t chooses a path internal to the his choice set $x_1 \in X^t$ to another $x_2 \in X^t$. 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 three possible set of outstanding path:

- **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, 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 five outstanding paths $x_{O_1}, \dots, x_{O_5}$ that we assume tourists have considered. We assume that the tourist t prefers the path x^t he has chosen to each of the outstanding paths: $u^t(x^t) \geq u^t(x_{O_i}), 1 \leq i \leq 5$.

  - how we can define this set?

  - We started to do this with 5 outstanding paths but results demonstrate the necessity to enlarge the set of path considered.

- **Soft set2**: a set of outstanding paths that we are sure that every tourist can do and that he doesn't prefer: for example, to spend all the day in a category of place.We define eight outstanding paths $x_{O_1}, \dots, x_{O_8}$ that we assume tourists have considered.We assume that the tourist t prefers the path x^t he has chosen to each of the outstanding paths: $u^t(x^t) \geq u^t(x_{O_i}), 1 \leq i \leq 8$. This set of outstanding paths is used for learning u^t. Then we validate this results verifying if $u^t(x^t) \geq u^t(x_{O_i}), 1 \leq i \leq 5$ using the **Soft set** of outstanding path.

  - We started to do this but the software seems to have problems because gives for each tourist the same result and results are not correct.

3. Method

1. **Splitting the space** Subdivide the territory in different spaces s.

2. **Classification of spaces.** Classify each space s in a Category of places c

3. **Counting the time.** For each tourists' path we analyse the time spent for each category of place c.For each s we analyse how much time T the tourists t spent in it. The time spent in each category is given by the sum of the different ranges of time in this category of place. We consider that each tourist has 15 hours to spend in a day.

4. **Value functions**. We define the value function for each tourist with the UTA method and the DIVIZ software.

Outstanding paths	
Soft set	
Op1	Environmental local (6h),Cultural (1h) and Food services (3h), other (4h)
Op2	Environmental territorial(7h) and Food services (3h), other (5h)
Op3	Historical local (2h), Cultural (2h), Leisure (6h) and Food services (2h), other (3h)
Op4	Historical local(4h), Leisure(3h), and Food services (1h), other (7h)
Op5	Historical territorial (7h), Historical local(3h), and Food services (3h), other (2h)
Soft set2	
Op1	Environmental local (15h)
Op2	Environmental territorial(15h)
Op3	Historical local(15h)
Op4	Historical territorial (15h)
Op5	Cultural (15h)
Op6	Food services (15h)
Op7	Leisure(15h)
Op8	Other (15h)

Table 2: **Outstanding paths**

5. **Defining weights**. We define the importance that each tourist gives to visit the different spaces.

6. **Clustering to define tourists' profiles**.

4. Results

…Results

5. Conclusions

…results …

6. other data

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 are ordered by importance with an evaluation scale. Let's have for each tourists' criteria $zt \in Zt$ a value $n \in \xi$ with $\xi\{1, 2, 3, 4, 5\}$

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

We know how the tourists planned to spend their time during the holidays in Alghero. Let's have for each tourists' action $at \in At$, a value $m \in \lambda$ with $\lambda\{1, 0\}$

$$At = m\{Environmental, Surrounding, Leisure, Fun, Food, Cultural, \\ Work\}$$

A. A simple goal

Data:

T	Set of tourists (t a tourist)
Z	Set of aspects on which tourist has declared a degree of interest (here above, Zt or At)
$d^t : Z \to \{ 1, 2, 3, 4, 5 \}$	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 her trip (from leaving home to returning home)
C_V	The categories of places in Alghero *that can be visited*, thus *excluding* staying home
$x^t : C_V \to \mathbb{R}^+$	The vector path chosen by tourist t (associates a number of seconds to each category of place). By definition, $\sum_{c \in C_V} x^t(c) = i^t$.

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 $P = \sum_{t \in T_2} \|P(d^t, i^t) - x^t\|$.

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

A.1. Central prediction

We compute the centre (using L2) of $(x^t)_{t \in T_1}$ and constantly predict this.

A.2. Regression

Build some regression model between $\{1, 2, 3, 4, 5\}^Z \times \mathbb{R}^+$ and $(\mathbb{R}^+)^{C_V}$.

A.3. Guided regression

Same idea but we (somehow) use knowledge about the link between Z and C_V.

A.4. 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.

A.5. Learned transform

Same idea as Transformed space, but we learn the partial value functions using T_1.

A.6. More ideas

See article Siskos. See utilitaristic regression from Eyke.