

Tackling Travel Behaviour: An approach based on Fuzzy Cognitive Maps

Maikel León, Gonzalo Nápoles, Rafael Bello

*Central University of Las Villas (UCLV), Santa Clara, Cuba.
maikelleon@gmail.com*

Lusine Mkrtchyan

Belgian Nuclear Research Centre (SCK•CEN), Mol, Belgium.

Benoît Depaire, Koen Vanhoof

Hasselt University (UHasselt), Diepenbeek, Belgium.

Received 14 November 2012

Accepted 22 February 2013

Abstract

Although the individuals' transport behavioural modelling is a complex task, it can produce a notable social and economic impact. In this paper, Fuzzy Cognitive Maps are explored to represent the behaviour and operation of such complex systems. An automatic approach to extract mental representations from individuals and convert them into computational structures is defined. For the creation of knowledge bases the use of Knowledge Engineering is accounted and later on the data is transferred into structures based on Fuzzy Cognitive Maps. Once the maps are created, their performances get improved through the use of a Particle Swarm Optimisation algorithm as a learning method, readjusting its predicting capacity from stored scenarios, where individuals left their preferences in front of random situations. Another important result is clustering the maps for knowledge discovery. This permits to find useful groups of individuals that policymakers can use for simulating new rules and policies. After related maps are identified, to merge them as a unique structure could benefit for different usages. Therefore an aggregating procedure is elaborated for this task, constituting an alternative approach for selecting a centroid of a specific estimated group, and therefore having, in only one structure, the knowledge and behavioural acting from a collection of individuals. Learning, clustering and aggregation of Fuzzy Cognitive Maps are combined in a cascade experiment, with the intention of describing travellers' behaviour and change trends in different abstraction levels. The results of this approach will help transportation policy decision makers to understand the people's needs in a better way, consequently will help them actualising different policy formulations and implementations.

Keywords: Travel Behaviour, Fuzzy Cognitive Maps, Particle Swarm Optimisation, Clustering, Aggregation.

1. Introduction

Travel behaviour studies are important for several reasons, e.g. to decrease both travel-related energy consumption and heavy load on urban infrastructure. Since decades, many attempts have been made to influence individuals' unsustainable travel behaviour towards more sustainable forms. However, just since recently, it is manifested how these studies can be effectively and efficiently implemented if they are developed and founded on a profound understanding of

the travel basic causes, such as people's reasons and inclinations, and comprehensive information of individuals' behaviours [1]. In the process of transportation planning, the travel demand forecast is one of the most important instruments to evaluate various policy measures to influence travel supply and demand. In the past two decades, due to the increasing environmental awareness and generally accepted policy paradigm of sustainable development, transportation policy measures shifted from facilitation to reduction and control [2].

One of the objectives of those measures is to inspire better use of available transport resources in order to evade the undesirable consequences of continued growing in private mobility. Therefore, the innovative travel demand models requests a truthful representation and understanding of the travel context and the decision making process of individuals, in order to mimic their sensitivity to a wider range of transport policy measures. Several modelling approaches related with travel demand have shifted from trip and tour based models to activity based models in which the context of daily travel (i.e. the need to perform activities, household interactions, etc.) is accounted [3]. Regular activities influence travel behaviour in a determinant way. Therefore, it is necessary to capture true individual decision mechanisms and reasoning strategies in order to improve behavioural realism of these kinds of models. Mathematical and computational modelling of these aspects has supported specialists to develop solid concepts in order to comprehend these activities in a logical way, and then be able to formulate policies based on real assumptions.

One of these approaches is considering individuals' travel selections as actual decision problems, where mental representations or more specifically, the Cognitive Maps (CMs) of the decision situation are generated. Important to notice that, at an individual level, it is central to realise that the relationship between travel decisions and the spatial characteristics of the environment is established through the individuals' perception and cognition of the environment. As an individual observes the space, for instance through travel, the information is added to the individual's mental map. This CMs notion is often referred to, in theoretical frameworks of travel demand models, especially related to the representation of spatial dimensions, but many other features can be taken into account. The individuals' environment is characterised by several situations that influence on daily actions and activities, making the human brain to develop a mental representation to efficiently manage the knowledge or perception of the real world.

The research on both CMs and travel focuses primarily, in fact almost exclusively, on some route choice [4]. In contrast, two other steps such as generating trips and its distribution have been given far less attention by

cognitive mapping researchers. Records regarding individuals' decision making processes can be used as input to generate mental models. Such models treat each individual as an agent with mental qualities, for example: viewpoints, objectives, predilections, and inclinations. In [5] the authors suggest CMs modelling to extract the mental representation of individuals in the planning of trips, related to daily travels. On the other hand, [6] has centred on the location, possible destinations, and feasible alternatives for any travel mode choice. The authors in [7] claim that the CMs of people who mostly walk and use public transit may vary systematically from those who are mostly chauffeured in private vehicles, or who usually drive themselves. In addition, those with regular access to private vehicles tend not only to search larger geographic areas for work, but also to perceive job opportunities in less spatially constrained ways.

While researchers have recognised the connection between travel and spatial learning, little is known about how the existing transportation infrastructure itself shapes CMs and, in turn, affects route selection as well as other aspects of travel including trip frequency, trip purpose, destinations, and mode choice. Modelling a dynamic system (i.e. an individual's CM for a travel behaviour representation) can be hard from a computational point of view. In addition, formulating a mathematical model may be difficult, costly, and even impossible [8]. Mathematical approaches offer the advantage of quantified results but suffer from several drawbacks [9]:

- Developing the model typically requires a great deal of effort and specialised domain knowledge.
- Systems involving significant feedback propagate causal influences in complicated chains, in which case, a quantitative model may not be possible.
- Numerical data may be hard to obtain. Efforts to communicate an understanding of the system and proposed solutions must rely on natural language arguments in the absence of formal models (qualitative approach).

Evidently, there is a necessity of treating this information with new advanced approaches, with the goal of managing the uncertainty that is implicit in these models. Artificial Intelligence (AI) and soft computing in general provide new methodologies able to attack

these kinds of problems. Soft computing could be seen as a series of techniques and methods so that real practical situations could be dealt with in the same way as humans deal with them, i.e. on the basis of intelligence, common sense, consideration of analogies, approaches, etc. In this sense, soft computing is a family of problem-resolution methods headed by approximate reasoning and functional and optimization approximation methods, including search methods. In consequence, Fuzzy Cognitive Maps (FCMs) are applied in this study to simulate individuals' decision making processes [10]. The application of FCMs is not only used to understand people's travel behaviours, but also to predict changes in their actions due to some factors in their decision atmosphere, and to discover hidden patterns.

More computationally speaking, FCMs are a combination of Fuzzy Logic [11] and Neural Networks [12]; combining the heuristic and common sense rules of Fuzzy Logic with the learning heuristics of Neural Networks. They were introduced by B. Kosko, who enhanced CMs with fuzzy reasoning [13]. There are proposals using FCMs for many applications and in different scientific fields. FCMs have been applied to analyse extended graph theoretic behaviour, to make decision analysis [14] and cooperate distributed agents [15], they are used as structures for automating human problem solving skills and as behavioural models of virtual worlds [16], and in many other fields. As decision makers activate a temporary mental representation in their memory based on their previous experiences or existing knowledge or preferences, we centred in how to apply FCMs to analyse traveling decision making processes of individuals. Constructing a mental representation requires decision makers to recall, reorder and summarise relevant information in their long-term memory. It may involve translating and representing this information into other forms, such as a scheme or diagram, supporting coherent reasoning in a connected structure [17].

With this regard, FCMs constitute a promising modelling approach for the described tasks. For a travel behaviour problem this study proposes a data gathering and formalisation procedure through the use of automatic knowledge engineering, using CMs approach as modelling technique and FCMs as computational

support. Learning of FCMs is used as a predictive tool to explore the possible preferences of users given specific circumstances; the learning results of FCMs will serve as a guide for transportation policy decision makers for future plans if, for example, one of the circumstances is changing, what are the most expected and reasonable actions to be taken.

On the other hand, clustering is used as a descriptive tool to analyse different groups of users and to understand the main features of choosing a transport mode; the results from clustering of FCMs will help decision makers to plan activities considering the specific needs of different groups of users. In addition, it is reasonable to offer a procedure for aggregating different maps, in order to deal with only one structure that symbolises several maps, being a "centroid" representation of the grouped knowledge. There is also an exploration on how travellers' preferences are changing through different levels of information processing abstraction, combining the learning, clustering and aggregation approaches. Proposed methods are implemented in a computational framework, user friendly, with visual and experimental facilities. The tool is provided with general functionalities to work with FCMs, also with specific procedures and panels for modelling and analysing the main problem covered in this research.

2. Individuals' transport mode selection as decision problem

Persons frequently explore and assess alternative courses of action when confronted with a decision problem, considering for instance, personal situations, earnings, objectives, etc. Consequently, a short-term and situation-specific reduction of environment is represented in their brains [18]. This referred decision context in which people live, and the knowledge they manage are characterised by spatial features. In decision theory in general, the predominant paradigm is expected utility theory, because a decision is considered to be a choice out of specific options, depending on the chance of occurrence and a valuation of a set of alternatives [19]. There exists a contrast in approaches of decision making applied to the different types of decisions that characterise individual travel and behaviour. This is, in one hand, the repetitive nature of trips (e.g. commuting, chauffeuring kids to school, grocery shopping, etc.)

likely to render (once) conscious decisions script-based or habitual behaviour. But on the other hand, activity scheduling (including choices of destinations, travel modes and routes, etc.) likely to entail the coordination of competing goals and intentions (e.g. amongst household members) in a complex environment (e.g. traffic-jams or opening hours), similar to complex planning problems [20].

The mental representation of both types of travel decisions can be modelled as a causal network. Recently, [21] and [22] have proposed a model of the mental repertoire of fixed scripts and routines in daily activity travel. The first reference developed a theoretical framework and qualitative study, while the second one used a probability network approach. Quantitative tools to understand human behaviour and values have only moderately been developed in past years [21]. In this topic, simple and practical research has been very dynamic, and it is only recently that the benefits in real applications are incorporated. Assimilation of investigated findings into decision making for public policy has been slow, but it is going now more rapidly, basically due to recent environmental and transportation legislation and associated research investments in the majority of the industrialised world [23].

Travel behaviour indicates generally the modelling and inspection of travel demand on the basis of theories and investigative methods from a variety of scientific areas [24]. They cover, but are not restricted to, the use of time and its allocation to travel and activities, the use of time in a diversity of time contexts and stages in people's life, and the organisation and use of space at any stage of social organisation, such as the individual, the household, the community [25]. As a result, the trends of man-made environments are dynamic as are the requirements and concerns of the people involved, and the policy, actions to consider. Subsequently, processes of future development increase, and the management is dynamic and complex. Early in travel behaviour studies, it was accepted that travel is one of the means people utilise to play a part in activities [26]. Therefore, travel behaviour techniques appeared from the proposal that travel demand is resulting from the desire to participate in activities and developed into activity-based approach, increasing the field's

traditional research boundaries and perspectives. In this way, investigators can comprehend the complex pattern of substitution between time at home and out of home, the use of stationary and mobile telecommunications' technologies to balance and substitute for travel, shifts between travel for work and travel for spare time, the effects of rising pace of life on travel, time allocation within a day and across days as well as weekends and its effects on predictability of travel demand, the effects of labour force composition shifts on travel, and so forth [27].

Travel-related decision making by people and groups occurs, within a spatial context. Decisions on where to locate activity centres, where to live, how to act together with other people and groups have a strong spatial component, but they are not restricted to this dimension. In current research, an FCMs model is primarily conceived as an individual mental model of decision problems, involved in daily activities and travel scheduling and embedding spatial and non-spatial cognitive factors. Travel mode choice is related to situational, attribute and benefit choice factors. This way, a decision network is constructed; an automation process is conceived both for data gathering and modelling of FCMs, being able to create an FCM structure representing each individual's mental representation.

3. Necessity of an effective knowledge engineering

While faced with a complex choice problem like an activity-travel option, people generate mental representations that allow them to understand the choice situation and assess the alternative courses of action. Much of the qualitative studies done in transportation research are carried out in order to improve the design of quantitative survey instruments. Examples are preliminary focus groups or personal interviews to test the wording of concepts and questions and to test whether all important factors have been included in the main survey design [28]. However, designers of large-scale surveys typically try to avoid open-ended questions or issues requiring much probing by the interviewer or reflection by the respondents [29]. As a result, the only qualitative information is provided by the survey itself, and the focus of such surveys are limited to the point that they rarely provide new insights

about the dynamic processes underlying the choice behaviour of interest. With very small sample sizes, any interesting new information is typically in the form of anecdotal evidence and it is difficult to generalise to a wider population.

Capturing the data is a first step in order to later understand or predict travel behaviour, therefore is one of the main challenges for transportation modelling. It is believed that travel and transportation models based on human behavioural characteristics will perform better when estimating the effect of various policy measures and that their use will lead to more realistic and thus more adequate predictions [30]. An effective knowledge engineering tactic could certainly benefit a correct data gathering task, ensuring less effort for the users and more quality in the results. Formalising knowledge for AI processing is a fundamental task. In the Knowledge Engineering (KE) process, after selecting the variables that better characterise a problem and the interaction among them, it is necessary to construct abstract structures to manage the organisation of the referred stored information.

The KE is defined as the group of principles, methods and tools that allow applying the scientific knowledge and experience to the use of the knowledge and their sources, by means of useful constructions for the human. It faces the problem of building computational systems with dexterity, aspiring first to acquire the knowledge from different sources and, in particular, to conclude the knowledge of experts and then to organise it in an effective implementation [31]. Therefore, KE is the process to design and make Knowledge Based Systems (KBS) operative; it is the topic concerning AI acquisition, conceptualisation, representation and knowledge application [32].

The fundamental problems in the construction of the KBS are [33]:

- Knowledge Acquisition (KA): How to transfer the human knowledge to an effective abstract representation, denominated conceptualisation.
- Knowledge Representation: How to represent the knowledge in terms of information structures that a computer can later process.

- Inferences Generation: How to use those information structures to generate useful information in the context of a specific case.

There are different ways of KA, some possible examples are:

- The expert interacts with the knowledge engineer to build the Knowledge Base (KB):
 - $[Expert] \rightarrow [Knowledge\ Engineer] \rightarrow [KB]$
- The expert can interact more directly with the ES through an intelligent publishing program, qualified with sophisticated dialogues and knowledge about the KB structure:
 - $[Expert] \rightarrow [Intelligent\ Program] \rightarrow [KB]^*$
- The KB can be built partially by an induction program starting from cases defined in books or past experiences:
 - $[Books] \rightarrow [Induction\ Program] \rightarrow [KB]$
- An acquisition method representing the most advanced techniques is the direct learning from books:
 - $[Books] \rightarrow [Data\ Processing] \rightarrow [KB]$

General requirements exist for the automation of the KA and they should be considered before attempting this automation, such as independence of the domain and direct use of the experts without middlemen, multiple accesses to sources of such knowledge as text, interviews with experts and the experts' observations [34]. The automated methods for the KA could also include analogy, learning like apprentice, learning based on induction cases and analysis of decision trees, discovery, learning based on explanations, neural nets, modification of rules and tools, helps for both the modelling and acquisition of knowledge that have been successful applied [35]. They seem to depend on intermediary representations that constitute a problem modelling language that help to fill the hole between the experts and the program implementations.

3.1. Automatic perspective

Diverse reasons have been considered for the construction of the Automated Knowledge Engineering (AKE), for example, the cost of software and the hardware for ES declined, favouring the development of AKE implementations. This has increased the demand

* Next section describes the implementation followed to conduct this type of knowledge acquisition.

of KBS, greater than the quantity of AKE. The movement toward an extensive human activity, as the KE, is contrary to all the industry tendencies, in particular the software industry [36]. The Knowledge Engineer's role, as a middleman between the expert and the technology, sometimes is questioned. Not only because it increases the costs but also for their effectiveness, that is to say, it can cause loose of knowledge or the KB that is build can be influenced subjectively. The automated KA keeps in mind what measures belong together, also the description of the application domain that the expert has and the existent description in the KB, together with how to integrate the new information that the expert offers to the KB. The AKE, if possible, should be independent of the experts' domain, to be directly applicable for the experts without middleman to ascend diverse sources of knowledge, including texts, interviews with the experts, and other features [37].

Diverse methods for implementation of AKE exist, some of the most known are [38]:

- Generation of rules starting from a database: All data fields, except the last one, correspond to the attributes or conditions, and the final one corresponds to the conclusion. Each article of the base becomes a production rule.
- Dialogue with the experts: The AKE should guide the expert, but with a set of flexibilities.
- Learning for similarity: Given a group of objects which represent correct examples and opposite ones, the AKE generalises a description that covers the positive examples and not the negatives. The positive examples generalise and the negatives specialise the objects (the concepts can be described as rules).
- Numeric parameters adjustment of specific parts of the knowledge: Could be an improvement in the coefficient of expressions that belongs to the production rules.

Several of the existent methods to acquire the knowledge in an automatically mode work with a fixed representation language, that could be developed in many cases by the own designer. The training data (examples) for these methods can contain non-prospective errors using the knowledge domain to guide the learning. Some methods of automated learning are

not strong enough to select the appropriate generalisation of the data [39].

For the data capture in the knowledge engineering process, an AKE implementation have been used, where the user is able to select groups of variables depending on some categories, which characterise what they usually take into account in a daily travel activity. There are diverse dialogues, trying to guide the user, but never in a strict way or order. To achieve an effective knowledge acquisition, a dynamically appearance according to the behaviour of the experts is required. This makes possible to capture and formalise in the KB, the relationships among the variables involved in the individuals' decisions and the degrees of causal influences. Thus, there is a mitigation of the "bottlenecks" problem, typically present in the knowledge based systems construction [40].

For storing the information we used a cognitive model, which is an abstract structure containing all information related with visualisation in the decision making process. The following design is used in our approach:

- Personal information about the individual: useful for demographic analyses, etc.
- Cognitive subsets: interaction among types of variables in the decision making process expressed by triplets in the form 'situation'[†] - 'benefit'[‡] - 'attribute'[§].
- Expert criteria: all cognitive subsets are captured by using artificial scenarios. Situational variables are assigned with random states, and the respondent specifies the utility of those conditions in terms of using 'bus', 'car' or 'bike'.
- Causal influences among variables: the users (also called experts due that they provide their knowledge in the conducted study) evaluate the causal relations among the variables they had selected.
- Benefits importance: experts assign an importance level to all benefit variables.

[†] Factors that are not influenced by decisions.

[‡] Individuals' pursued goals or needs.

[§] Observable characteristics of alternatives in a choice set.

Decision alternatives, situational aspects, attributes, and benefits are considered together in an evaluation process, prior to making a choice [22].

Software products' success definitely depends on its architectural design and the flexibility of the user interface, especially when the software product is an AKE implementation. The first stage aims at eliciting considered aspects, constructs, beliefs and their interconnections in the decision process using probing questions. They are represented as cognitive subsets, consisting of the interconnected situations, benefits, and attributes.

3.2. System reliability

In the conducted case study, 223 persons living between 5 and 10 kilometres from the city centre of Hasselt (Belgium) were asked to use the software, and the results were satisfactory given that the 99% of individuals (see Figure 1) were able to interact completely along with the AKE implementation, generating their own CM about a shopping activity scenario that was presented. Users from different academic levels, age, gender, living neighbourhoods, incomes, etc. participated in the survey; they specified useful information for further analysis.

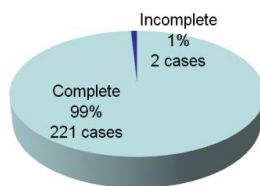


Fig. 1. Percentage of complete generated KBs.

From the 221 files, socio-demographic factors are extracted in order to characterise the composition of participants. Some statistics information will be referenced from the conducted research in [41]. For example, 43% of respondents are male (95) while 57% are female (126), both groups are sufficiently represented. Ages between 18 and 70 are represented; Annex 4 shows the sample divided by age. It is also important to mention that 62% of the respondents obtained a higher education degree. About respondents' overall mobility, most of them state that their household has 1 or 2 cars. Few households possess 3 or more cars. Only 1 respondent states that his/her household did not possess a car. They were also asked to indicate what other transport options they have, besides car. More than 90% of them indicate they have access to a bicycle.

A few respondents have access to other means of motorised transport like a moped or motorcycle.

Respondents were also asked where they park when they go to the city centre by car. By far the largest part of respondents usually parks on a free parking space when they go to the city centre by car. The situation variables that are most indicated are 'time available', 'precipitation', 'baggage' and 'parking space availability'. The most important benefits are 'efficiency' and 'freedom'. The attributes that are indicated by most respondents are 'flexibility', 'travel time', 'accessibility', 'easiness for parking' and 'treatment of bags'. The triplet 'precipitation' (situation) – 'shelter provision' (attribute) – 'physical comfort' (benefit) is by far indicated by most respondents.

Furthermore, the situation variable 'time available' is linked by many respondents to the benefit 'efficiency' and the attributes 'travel time', 'flexibility', 'easiness for parking' and 'direct travel'. Also, the situation variable 'baggage' is quite often linked to the benefit 'physical comfort' and the attributes 'treatment of bags' and 'physical effort'. The situation variable 'parking space availability' is often linked to the benefit 'efficiency' and the attribute 'easiness for parking', 'accessibility' and 'travel time'.

Even though these most important considerations are, in general, important to most socio-demographic groups, there are some differences among different subgroups of the sample. These differences can be used to target intervention strategies and campaigns in a segmented way, which is generally considered to be a more effective approach [41]. The 221 files, outcomes of the data gathering process, serve as KBs for the forthcoming analysis of the travellers' mental representations. Hereinafter, we will often refer to this quantity of KBs in next sections, involving this information for machine learning, clustering and aggregation procedures.

4. Fuzzy Cognitive Maps as modelling technique

In the referred knowledge acquiring process, the developed AKE codes all information concerning the mental maps of the experts in KBs. However, the stored knowledge is worthless if it cannot be represented in

terms of computational structures for further analysis, automated or not, but basically for simulations or predictions that are useful to science and application areas. The construction of FCMs from the KBs generated by the AKE helps to solve this problem [42]. Among different possible AI knowledge representation forms, FCMs are explored in this study. First an overview about the technique is presented; afterwards an automatic construction is described.

The CMs theory was used notoriously by Axelrod [43] who focused on the policy domain studies. This is one of the first references where a mathematical perspective was introduced into the referred theory. Since then, many researchers have used CMs in various fields where the problems are ill structured or not well-defined. The reader can refer to [44] and [45] for details about CM early usages, and to [46], [47], [48] and [9] for more recent studies. Several applications were developed, reported in [49], [50], [51], [52], [53], [54], [55], [56], [57], [58], [59], [60], [61], etc.

In general, a CM has two types of elements: *concepts* and *causal beliefs*. The former are variables while the latter are relationships among variables. Causal relationships can be either positive or negative, as specified by a '+', respectively a '-', as a sign on the arrow connecting two variables. The variable that causes a change is called a *cause* variable and the one that undergoes the effect of the change is called an *effect* variable [62]. If the relationship is positive, an increase or decrease of the cause variable causes the effect variable to change in the same direction (e.g., an increase in the cause variable causes increase of the effect variable). In the case of a negative relationship, the change of the effect variable is in the opposite direction (e.g., an increase in the cause variable causes decrease of the effect variable).

Figure 2 shows an example related to the travel behaviour problem addressed previously through a simple CM. This example shows some of the concepts to decide which transport mode is more appropriate for a specific task: using bike or car? Here only the change directions are shown; for example the increase/decrease of 'number of bags' causes the decrease/increase of 'bike comfort': therefore the sign is negative. On the other hand, if we consider the causal relationship

between 'bike infrastructure' and 'bike comfort', the link is positive as the change in cause node changes the effect node in the same direction. However, CMs, whatever their types, are not easy to define only expressing a sign in the relationships. Usually CMs are constructed by gathering information from experts/users and generally, they are more likely to express values in qualitative terms or with a specific degree of causality [63].

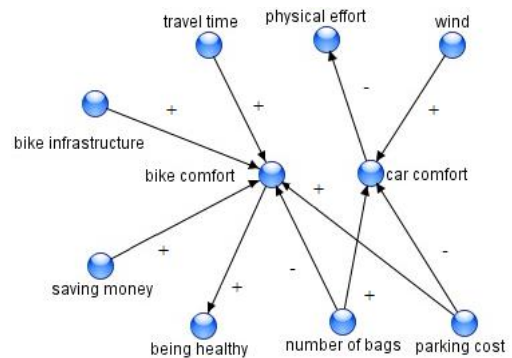


Fig. 2. Example of a CM.

To this end, it may be more appropriate to use FCMs, suggested by Kosko [13]. Actually, FCMs are weighted CMs where the weights are associated with fuzzy sets [64]. So, the degree of the relationship among concepts in an FCM is either a linguistic term, such as: often, extremely, some, etc.; or a degree of activation/causality in [-1, 1].

Figure 3 shows the corresponding FCM illustrated in Figure 2 where the causal relationships are expressed by using fuzzy linguistic terms. For example, if we consider again the relationship between 'bike infrastructure' and 'bike comfort', the increase/decrease of cause variable will cause high increase/decrease in effect variable. The nature of the collected data in this study permitted to model the maps in such a way that all links have only positive signs.

Note that each concept can represent a characteristic of the system; in general it stands for events, actions, goals, values, trends of the system that is modelled, etc. In addition, each concept is characterised by a number that represents its value and it results from the renovation of the real value of the system's variable [65].

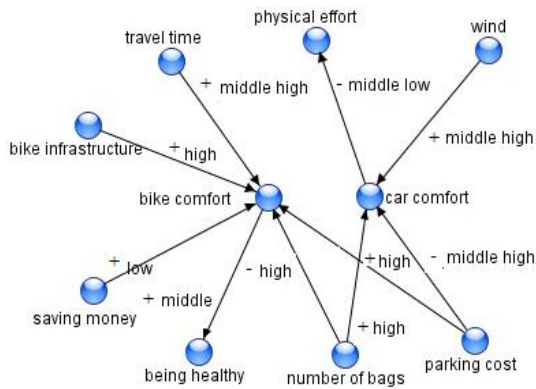


Fig. 3. Example of an FCM.

When a degree of activation is used in the relations modelling, beyond the graphical representation of the FCM there is a precise mathematical model consisting in a $1 \times n$ state vector A which includes the values of the n concepts and, a $n \times n$ weight matrix (adjacency matrix) W which gathers the weights W_{ij} of the interconnections among the n concepts [66]. The value of each concept is influenced by the values of the connected concepts with the appropriate weights and by its previous value. So the value A_i for each concept C_i can be calculated as expressed in (1):

$$A_i = f\left(\sum_{j \neq i}^n [A_j \times W_{ji}]\right) \quad (1)$$

where A_j is the activation level of concept C_j and W_{ji} is the weight of the interconnection between C_j and C_i , it is to say, the value of A_i depends of the weighted sum of its input concepts, and f is a threshold or normalisation function [67]. The nonlinear function, f , can be a simple thresholding operation with a threshold value T resulting for example binary concept values. To produce continuous concept values, a continuous-output transformation function may be used [68]. The most widely used function is the sigmoid function. As for our application the links of FCMs have only positive signs, we choose the normalisation function given in expression (2) that fits our task the best.

$$f(x) = \frac{1}{1+e^{-c(x-0.5)}} \quad (2)$$

Among several ways of developing CMs and FCMs, the most common used methods are extracting knowledge from questionnaires, extracting knowledge from written texts, conducting interviews, or drawing maps from data [69]. Note that these methods can be used also in combinations, such as questionnaires with interviews. To have more reliable results, more than one expert must participate in an FCM drawing or in the knowledge acquiring process to construct it. So, FCMs are a powerful methodology that can be used for modelling systems, avoiding many of the knowledge extraction problems which are usually present, for example, in rule based systems. For this study we developed a dedicated software tool, and the details are described in [70]. Although it was conceived for general purposes, special options are included and some specific methods are developed to deal with data requirements used in this study.

An important advantage of FCMs over other approaches like Petri Nets (PNs) or Bayesian Networks (BNs) is related with the possibility of developing a group map based on a set of individual maps. Petri Nets [71] are another graphical and mathematical modelling tool consisting of places, transitions, and arcs that connect them, and can be used as a visual-communication aid similar to flow charts, block diagrams, and networks. As a mathematical instrument, it is possible to set up state equations, algebraic equations, and other mathematical models governing the performance of systems. It is well known that the drawing process of PNs by non-experts in the technique is almost impossible; in addition, it is not well established how to combine different PNs that describe the same system [72]. BNs [73] are a powerful tool for graphically representing the relationships among variables and for dealing with uncertainties, but request demanding effort caused by the net specification (structure and parameters) and its expensive algorithm of probabilities propagation. It is not evident for a non-expert in BNs its construction, and even more difficult how to compare or combine them [74].

4.1. Automatic construction from KBs and proposed topology

The AKE implementation codes into files a “flat” mental representation but it could be interesting to visualise it and to process this information in order to provide another level of construct, where it is possible

to discover relevant hidden patterns. Consequently, the goal of this section is to describe how to build FCM structures automatically from the generated KBs.

To express the stored knowledge in terms of components that allow building FCMs automatically, we identify three primary structures:

- Cognitive subsets: represent triplets of concepts composed by ‘situation’ - ‘benefit’ - ‘attribute’, defining the sub-structures of the map. The concepts for the decisions (‘Bus’, ‘Car’ and ‘Bike’) and for the final utility (‘Utility’) are implicit concepts and mandatory for every cognitive subset. For example, for the cognitive subset ‘Precipitation’ - ‘Stay dry’ - ‘Physical comfort’, the topology of the resulting map is shown in Figure 4.
- Concept causal influences: values in the interval [0, 1] representing the causal influences among the variables of the cognitive subsets.
- Weights of the benefit variables (Ratings): represent a list of all weighted benefits with a value in the interval [0, 1]. This value represents the impact of a benefit concept in the map.

One of the most important features of this topology is the fact that the concepts C_6 and C_7 (see Figure 4) are considered vector variables and not scalar ones, thus, when running the inference process only one of the three decisions is active (property exclusivity in the decision node), so the final result of these concepts will be a three-dimensional vector (making an inference for each decision). This idea allows calculating the utility of each decision under specific circumstances. In an imaginary case where an expert considers modelling his/her considerations according to own experience, the situational variable ‘Precipitation’ will be the cornerstone of his/her decisions, the benefit variable ‘Physical comfort’ describes the benefit to select ‘Precipitation’ variable and ‘Stay dry’ characterises the relationship between the two variables. Later, a list of imaginary scenarios (involving the variables that were previously chosen by an expert) is loaded to know the expert criteria under specific circumstances. For example, if a ‘No rain’ state of the ‘Precipitation’ variable (situation) is presented, the respondent preferred ‘bike’ as first option, with less preference considers taking a ‘bus’ and finally a private transport (‘car’); because the benefit he/she wants to gain is

‘Physical comfort’ with the intention to ‘Stay dry’ (attribute).

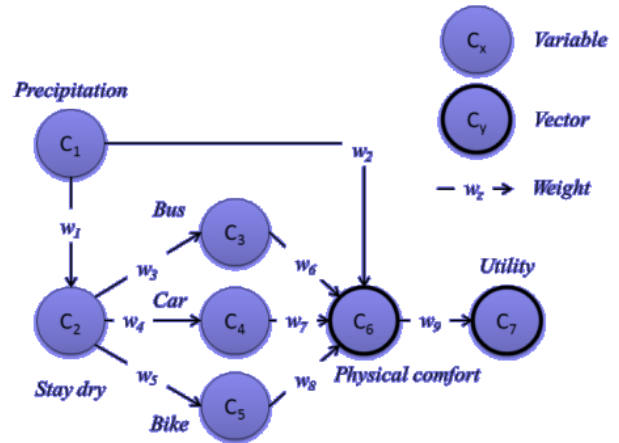


Fig. 4. General topology of a minimum FCM.

The benefits of converting the “flat” mental maps’ representations which are stored on the KBs into structures of FCMs with an automatic method are listed below [75]:

- uniformity in the visual structures,
- convenience for further processing, as there is an established configuration of terms and data, and
- flexibility to researchers for advance studies and analysis.

5. A Particle Swarm Optimisation method for learning of Fuzzy Cognitive Maps

Due to several problems associated with the development of FCMs, many researchers investigated automated or semi-automated computational methods for learning of FCMs using mainly historical data. Semi-automated methods require a relatively limited human intervention, whereas fully automated approaches are able to compute a model of FCMs solely based on historical data. Research on learning models of FCMs from data have resulted in many alternative approaches. One group of methods aims at providing a supplement tool that helps experts to develop accurate models based on their knowledge about the modelled system [76]. Algorithms from the other group are oriented toward eliminating human factors from the entire development process, only historical data is necessary to establish models of FCMs. In general, the algorithms can be categorised into two groups based on

the learning paradigm that is used, i.e., Hebbian-based learners (characterized by a high computationally expensive) and methods based on evolutionary algorithms [77].

Particle Swarm Optimisation (PSO) belongs to the class of Swarm Intelligence algorithms. PSO is a bio-inspired metaheuristic that simulates the social behaviour observed in groups or swarms of biological individuals [78]. This stochastic technique starts from the principle that intelligence does not lie in individuals but in the collective, allowing for the solution of complex optimisation problems from a distributed point of view, without a centralised control of a specific individual. Each organism (particle) adjusts its position by using a combination of an attraction to the best solution that they individually have found, and an attraction to the best solutions that any particle has found, imitating those with a better performance. Thus, the particle swarm overflies the search space detecting promising regions (see Annex 8).

Some examples of population based algorithm in the related literature are:

- Ant colonies [79],
- Bird flocks [80],
- Fish schools [81].

Some basics of PSO are the following:

- Each particle (possible solution) has a quality measure indicating how good it is.
- Each particle has a position and a speed in the search space.
- The position of a particle is interpreted as a possible solution to the problem.
- Each particle interacts with a number of neighbours, learning from the interaction, adjusting position and speed, attracted by the best position of the neighbourhood, and for the best founded by the swarm.

Summing up, the particles are interpreted as possible solutions for the optimisation problem and are represented as points in n-dimensional solution space. In the case of standard PSO, each particle (X_i) has its own velocity (V_i) bounded by a maximum value (V_{max}), a memory of the best position it has obtained and

knowledge of the best solution found in its neighbourhood.

In the search process the particles adjust their position according to the following equations (3) and (4):

$$V_i^{(k+1)} = V_i^{(k)} + c_1 r_1 (X_{pbest_i} - X_i^{(k)}) + c_2 r_2 (X_{gbest} - X_i^{(k)}) \quad (3)$$

$$X_i^{(k+1)} = X_i^{(k)} + V_i^{(k+1)} \quad (4)$$

where k indexes the current generation, c_1 and c_2 are positive constants, r_1 and r_2 are random numbers with uniform distribution on the interval $[0, 1]$, X_{pbest_i} is the best previous position for the particle X_i , while X_{gbest} represents the best global found by the swarm. An important parameter that modifies the PSO algorithm is the *inertia weight* (ω) added by [82] to replace V_{max} . The incorporation of this parameter guarantees the balance between the capacities of local and global search; a higher weight value will facilitate the exploration, while a low weight facilitates the exploitation. The wrong choice of this parameter value will affect the algorithm convergence speed, so it is recommended to adjust it dynamically as shown in the following equation (5):

$$\omega_k = \omega_{max} - \frac{\omega_{max} - \omega_{min}}{N} k \quad (5)$$

where k is the current cycle, N corresponds to the number of generations, while ω_{min} and ω_{max} match the end points of the interval $[\omega_{min}, \omega_{max}]$ on which the k^{th} inertia weight is defined. Other parameter that changes the PSO algorithm is the *constriction coefficient* (χ) introduced by [83]. This parameter ensures that the algorithm converges to avoid the explosion of the particle swarm and it can be expressed in terms of c_1 and c_2 as follows in (6):

$$\chi = \frac{2}{\left| 2 - \varphi - \sqrt{\varphi^2 - 4\varphi} \right|} \quad \text{and } \varphi = c_1 + c_2, \varphi > 4 \quad (6)$$

A comparison study of the two methods demonstrated that the constricted PSO algorithm is in fact a special case of the algorithm with inertia weight in which the values for the parameters have been determined

analytically [84]. In our proposal both factors are applied to the equations (3) as follow in (7):

$$V_i^{(k+1)} = \chi \left(\begin{aligned} &\omega_{k+1} V_i^{(k)} + c_1 r_1 (Xpbest_i - X_i^{(k)}) + \\ &+ c_2 r_2 (Xgbest - X_i^{(k)}) \end{aligned} \right) \quad (7)$$

These modifications returns improved performance over the original PSO algorithm [85].

5.1. Proposed method to improve the causal matrix of an FCM

A PSO approach can be used to improve (*training, learning*) the accuracy of FCM structures based on collected data. Some methods have been proposed with this goal, such as [86], [87] and [88]. The map generated from the AKE implementation needs to be adjusted using the information from stored scenarios (cases). As far as for FCMs, this method improves the quality of resulting FCM models by minimising an *objective* or *heuristic function*. The function incorporates human knowledge by adding constraints the map must satisfy, without affecting the physical meaning defined by experts.

Figure 5 shows the main idea of the proposal to use PSO in learning of FCMs. The flow chart illustrates the application of PSO to train the weight matrix, trying to find a better configuration for a convergence or finds the expected results. PSO is applied straight forwardly using an objective function defined by the user. First the concepts and relation are defined, and the construction of an FCM is made, and then it is possible to make simulations and obtain outputs due to the inference process. If the output values are not adequate or the expected ones, easily known by the execution of the heuristic function, it is necessary a learning process (in this case, through the use of the PSO metaheuristic); resulting then a new weight matrix configuration.

The optimal case is when the newly found causal matrix offers the expected output (same values of the inference and expert decisions), minimising the sum of the absolute differences between desired and actual output concepts' values. Another element requiring special attention to solve the problem using the PSO metaheuristic is the representation of the causal matrix as a particle. In case of FCMs, each particle of the swarm is a weight matrix [89], encoded as a vector. The

matrix w_{ij} is expressed as a vector with dimension $i*j$, containing all rows of the matrix, one following the other.

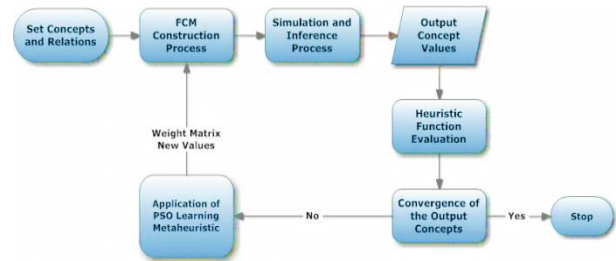


Fig. 5. Idea of using PSO for readjusting an FCM.

According to all these considerations, the percentage (P) defined by $P_h(x)$ of well classified cases (c) and the accumulated error $E_h(x)$ of the classification for heuristic h and map x are calculated by the following expressions shown in (8) and (9):

$$P_h(x) = \frac{[\max h(W_x)/c] - [h(W_x)/c]}{\max h(W_x)} * 100 \quad (8)$$

$$E_h(x) = h(W_x) \bmod c \quad (9)$$

As it is known, the optimisation capability of the PSO algorithm is quite sensitive to parameters [90]. Through experimentation some of them were deducted to perform the task the best, but in general the tactic itself offered a satisfactory solution, this according to the very complex explored search space. The scenarios' evaluation must represent, in a certain way, the expert exigency level, that is to say, the users' expected utility degree that allows taking the best decision. So, to find the best causal matrix which minimises the differences between the expert and the computed inference, according to heuristic h , and to adjust the causal influences among the concepts of the map, we apply a PSO algorithm adapted to the characteristic of the current optimisation problem.

Several criteria can be considered to minimise the desired function h , four possible heuristics are going to be presented and its corresponding meaning (useful definitions at the end of the formulations):

- h_1 : Considers all cases as correctly classified, only evaluating the classification error according to the difference between the utilities given by the expert

and the inferred one, not regarding the order among them.

$$h_1(x) = \sum_{i=1}^n \sum_{k=1}^m |ExpDecision_i(x)_k - InfDecision_i(x)_k|$$

- h_2 : Considers one scenario well classified when the first decision obtained by the inference is the same as the one given by the expert, not accounting the correct order of the other decisions.

$$h_2(x) = \sum_{i=1}^n r_i(x),$$

$$r_i(x) = \begin{cases} 0, & ExpDecision_i(x)_1 = InfDecision_i(x)_1 \\ q > 0, & \text{in other case} \end{cases}$$

- h_3 : Considers one scenario well classified when the order of all decisions given by expert is the same as the one found by the map inference.

$$h_3(x) = \sum_{i=1}^n r_i(x),$$

$$r_i(x) = \begin{cases} 0, & ExpDecision_i(x) = InfDecision_i(x) \\ q > 0, & \text{in other case} \end{cases}$$

- h_4 : Considers one scenario well classified when the order of all decisions given by expert is the same one found by the map inference but also quantify the quantitative values of all decisions. This is the most restrictive heuristic, the optimal case is when the new found causal matrix completely offers the expected output:

$$h_4(x) = \sum_{i=1}^n r_i(x),$$

$$r_i(x) = \begin{cases} A, & ExpDecision_i(x) = InfDecision_i(x) \\ q > 0, & \text{in other case} \end{cases},$$

$$A = \sum_{k=1}^m |ExpDecision_i(x)_k - InfDecision_i(x)_k|$$

In the previous functions, n is the number of scenarios (evaluation data for predicting and comparing), m is the number of transport mode decisions (in this research is three, benefit of using: ‘bus’, ‘bike’ and ‘car’). Thus, the k^{th} decision of i^{th} scenario according to expert criteria (*ExpDecision*) is compared with homologous decision inferred by the map x (*InfDecision*). We had used $q= 30$ (penalisation quota) as this value was best fitting the data. In the definition of the optimisation problem it must be considered that the value of the parameter q must be a positive number, because this is a penalisation to the function h (that needs to be minimised) when the computed inference by map x does not coincide with

expert criteria for specific evaluated scenario and the causal matrix w . There were presented six possible distance measures among other approaches that could be explored. The heuristic h_4 is the most difficult function to be minimised, because it tries to obtain a map that offers the most suitable output for the analysed cases. The next section describes the effectiveness of the proposed method and a comparison with other approaches.

5.2. Validating the learning method

After the construction of the 221 maps (section 3.5), only using the variables and default values for relations, only 24% of the original maps were able to predict 100% of scenarios using h_4 (most restrictive function) and sigmoidal with $c=9$ normalisation function. Using the stored scenarios as training data**, after a learning process, the 76% of maps were able to predict 100% of scenarios (see Figure 6). The improvement in the classification skills of the maps is significant; the value of the averaged heuristic function (distance) reported a decrease in its value of 83.26 units.

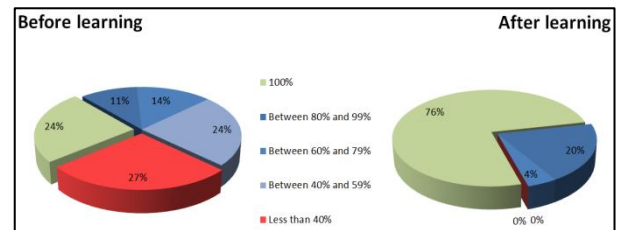


Fig. 6. Qualitative view of the scenarios' predictions.

The evidences allow concluding that the learning method improved the performances of the initial maps. But also a comparison is needed with other techniques to justify its use. As mentioned before, in [22] a study was performed with the same used data, nevertheless the results will not be taken into comparison as there was not any application of machine learning to readjust parameters from data. The results of the applied Inference Diagrams were significantly lower in comparison to a simple Decision Tree results. The ‘car’ prediction performed 76.8%, ‘bus’ prediction at 69.3% while ‘bike’ only a poor 51.7%. So, it is not really fair to compare with the results obtained in our proposal. It is important to mention that, the applied Decision Tree classifier can only use a categorical variable as the

** Preliminary study, next section introduces statistical analysis.

dependent variable to predict. While the followed approach in this research, modelled the three transportation options as an order of preferences.

In order to validate the performance of an FCM against other classical approaches such as Multilayer Perceptron (MLP) Neural Network, ID3 Decision Tree, or Naive Bayes (NB) classifier, the same knowledge had been modelled with these referred techniques. Many others classifiers could be considered, the intention was to deal with methods based on different approaches (probabilistic, rules, etc.). Waikato Environment for Knowledge Analysis (WEKA^{††}) has been used for constructing these models. The prediction capability had been measured in the forecast of the first possible decision and in three decisions given by the experts (order of preference using ‘bus’, ‘bike’ and ‘car’). In Table 1 the data organisation for the statistical experiment is listed, through a population comparison. The idea consists in analysing a possible significant difference among the techniques using their classification per cent (CP), obtained in a cross-validation (C-V) process with 10-folds. By columns we have the four considered techniques in the comparison, in the rows the CP that each technique offered according to the information from each expert. So CP_{FCM (1)} means the CP that the FCM obtained for the data provided by Expert 1. Similar for the others cells.

Table 1. Data organisation for processing.

	FCM	MLP	ID3	NB
Expert 1	CP _{FCM (1)}	CP _{MLP (1)}	CP _{ID3 (1)}	CP _{NB (1)}
Expert 2	CP _{FCM (2)}	CP _{MLP (2)}	CP _{ID3 (2)}	CP _{NB (2)}
...
Expert 221	CP _{FCM (221)}	CP _{MLP (221)}	CP _{ID3 (221)}	CP _{NB (221)}

After applying a Kolmogorov-Smirnov test [91] and having a non-normal distribution in our data, we apply non parametric Friedman test [92], where a significance less than 0.05 suggests rejecting the main hypothesis (no significant differences among groups), therefore we can conclude that there exists a significant difference. It is important to mention that as the data reflects classification percentage, a higher value in the mean rank corresponds to a better performance. Looking to the mean ranks, the best value is given to FCM.

^{††} <http://www.cs.waikato.ac.nz/ml/weka/>

However, it is not possible yet to affirm that our technique performs better than the others. Using a Wilcoxon test [93] for related samples it is possible to analyse per pairs, and in all cases the main hypothesis (no significant difference between pairs) of the test is rejected and it is confirmed that there exists a significant difference. Using a Nemenyi test [94] and calculating the critical distance (CD) among classifiers, it was corroborated that FCM definitely offers better results.

Finally, Table 2 contains the average percentages. First the learning scenarios serve for training, then for calculating optimistic estimation (resubstitution technique, empirical error) of the convergence. The resubstitution test is absolutely necessary because it reflects the self-consistency of the method, a prediction algorithm certainly cannot be deemed as a good one if its self-consistency is poor. Later, the testing scenarios were used to obtain a pessimist estimation (C-V, real error) of the convergence through a C-V process with 10-folds.

Table 2. Classification percentage per Technique (FCM, MLP, ID3, NB), Experiment (First or Three Decisions) and Model (Optimistic or Pessimistic).

FCM	MLP	ID3	NB
FIRST DECISION			
<i>Optimistic Model</i>			
99.47	97.38	94.26	95.63
<i>Pessimistic Model</i>			
93.74	92.06	89.39	91.37
THREE DECISIONS			
<i>Optimistic Model</i>			
96.27	94.38	87.29	93.12
<i>Pessimistic Model</i>			
88.72	82.40	77.59	80.25

A C-V test for an independent testing data set is needed because it can reflect the effectiveness of the method in future practical applications. An analysis about the advantages that our method has over the others is productive, thus we clarify that FCMs had a higher accuracy than other classical well-known techniques when predicting the best decision, in this case the transport mode selection or the whole preferences of a transport mode order. In one hand, the existing

knowledge in the elaborated topology definitely provides benefit to the classification structure, together with the effectiveness of the PSO learning method, which also contributes in this task, consequently the learned FCMs gained in the performed comparison. On the other hand, FCMs not only performed better, but also the most important is its capacity of presenting visual comprehensible information and this combined with the referred classification skills, makes them a good approach for these kinds of tasks.

6. Clustering method for Fuzzy Cognitive Maps

Analysing clusters is related to finding specific groups in data, by calculating a degree of associations among its objects. Therefore, the association between two items is maximal when they are in the same group and minimal if not. This technique has commonly been used in many applications and basically in all kinds of knowledge areas and domains [95]. Accordingly, the objective of this section is to develop an approach to cluster the FCMs in order to identify types of individuals, based on their mental representation structures, and to analyse how and why their reasoning mechanisms take into account specific variables for gaining an objective. There are diverse varieties of methods for clustering, such as joining (the tree-clustering or the agglomerative hierarchical method), two-way joining (block clustering), and k-means clustering, etc. Some software products have incorporated these techniques to allow studies and analysis. While there are many examples of the use of CMs or FCMs in different application fields, there are only a few attempts for clustering of CMs. Earlier studies of finding the similarity between CMs are done in [96] and in [97].

The first study is restricted only to a similarity algorithm without further use of it for cluster analysis. The second study extends the similarity measurement of the first approach and, in addition, a cluster analysis is provided. However there are some drawbacks of this study. For cluster analysis *Ward linkage* method is applied which commonly uses the *squared Euclidean distance* for the similarity matrix calculation. In addition, the optimal number of clusters is not discussed; neither any validation method is applied for cluster analysis. An interesting and recent study is done in [98].

However, while the clustering based on a map structure is comprehensive enough to be used for different application domains, the clustering based on the map *content* is rather case specific and is restricted to principal component analysis. Moreover, neither the weights of the links nor the links' signs are considered; thus, the functionalities of CMs are not fully discussed in cluster analysis. Another study of clustering is done in [99]. This approach differs from previous ones as it clusters not FCMs into different groups but the nodes of an FCM. A similar approach of clustering the nodes into hierarchical structure is proposed in [100]. These last two studies are of less interest for this proposal because first we already have structured representation of the nodes in our dataset, and, second, we are more interested in clustering the FCMs (not in clustering the nodes).

6.1. The distance matrix of FCMs

In general CMs/FCMs can be compared in two dimensions: comparing the content and the structure of each map. The content difference is associated with the differences in elements in both maps and the differences of the relationships among those elements. The structural difference, on the other hand, is associated with the varying complexity degrees of the maps' structure. In our application we have hierarchical maps for all users therefore we will focus only on the content difference analysis.

Three types of differences between two individuals (maps) can be considered:

- Existence or non-existence of elements: thus one expert considers a specific element as important for the given domain; the other has the opposite opinion. In this case the adjacency matrix for the CM of the first expert contains the element/elements while the other matrix does not contain.
- Existence or non-existence of beliefs: thus one expert considers that there is a casual relationship between two concepts, while the other has the opposite opinion. In this case two experts should agree upon the fact that the nodes are important for the given domain, but have opposite opinions towards the causal link.
- Different values for identical beliefs: thus two experts agree that there is a relationship between

two nodes, but one expert holds the belief more strongly than the other. In adjacency matrices this difference is expressed by non-identical non-zero values for the cell showing the causal link between two nodes.

In [97] the authors suggest an improvement similarity measurement algorithm for FCMs described in [96] mentioning that the algorithm does not consider the missing values properly as well as it lacks of generalisability. We will take into account only the comment about generalisability, and will adjust the algorithm to be applicable in our study. Notice that by generalisability, we mean that in [96] the number of linguistic terms is fixed and the comparison formula cannot handle different number of linguistic terms. More specifically, the study fixed the number of linguistic terms to 7, assigning maximum strength to 3, minimum strength to -3, and the similarity measurement algorithm cannot be applied for the cases with more/less linguistic terms.

Therefore, for our task, we use the following distance ratio (DR), expressed in (8) and (9).

$$DR(u_A, u_B) = \frac{\sum_{i=1}^p \sum_{j=1}^p (|a_{ij}^* - b_{ij}^*|)}{2p_c^2 + 2p_c(p_{u_A} + p_{u_B}) + p_{u_A}^2 + p_{u_B}^2 - (2p_c + p_{u_A} + p_{u_B})} \quad (8)$$

$$m_{ij}^* = \begin{cases} 1, & \text{if } m_{ij} \neq 0, \text{ and } i \text{ or } j \in P_c \\ m_{ij}, & \text{otherwise} \end{cases} \quad (9)$$

In (8), a_{ij} and b_{ij} are the adjacency matrices of the first and second map respectively, p is the total number of possible nodes, P_c is the set of common nodes for both maps, p_c is the number of such nodes, p_{u_A} / p_{u_B} is the number of nodes unique to user u_A / u_B respectively. In (9) m_{ij} is the value of the i^{th} row and j^{th} column in the zero augmented adjacency matrix.

6.2. Hierarchical clustering approach

Using cluster analysis as a descriptive tool can group the travellers and allows us analysing each group by finding similarity in thinking of people when making a decision about some transport mode. The understanding of travellers' behaviour tendencies will help policymakers in more realistic assessment when some concepts are changing over the time. As we have already mentioned in the previous sections, we have 221 users who were asked to choose the most important

variables of some transport mode choices. In addition different scenarios have been developed and asked to the users in order to provide their choice according to the circumstances in a scenario.

The nodes in FCMs are divided into three groups: benefit nodes, situation nodes and attribute nodes. A traveller indicates the main benefits of choosing a particular transport mode depending on the situations and the attributes related with that situations. For example, considering the set {'situation', 'attribute', 'travel mode', 'benefit'} with the following maps' nodes in each set respectively {'car availability'}, {'flexibility', 'independency', 'travel time', 'speed'}, {'car', 'bike', 'bus'}, {'convenient', 'freedom'}. Here the 'car availability' is a node showing the situation or context which is related with attributes such as 'flexibility', 'independency', 'travel time', 'speed'. The main benefits of choosing one out of three transport modes are 'convenient' and 'freedom'.

The expected results of cluster analysis will help transportation policy decision makers to understand which group of people has the inclination on which benefits, and what the situations leading to the chosen benefits are. Also, in discovering specific patterns in the groups which are not easily visible. Cluster analysis is an unsupervised learning method to examine the dataset by dividing it into groups so that the similarity within the clusters and dissimilarity among different clusters are maximised [101]. Cluster analysis methods have been applied in different fields such as engineering, social science, medical sciences, economics, etc. For more detailed description of clustering techniques the interested reader can refer to [102] and to [95]. In the previous subsection we presented the distance ratio algorithm to find the distance between two FCMs. There are two widely used clustering techniques based on similarity or distance measurements: the hierarchical approach and the partitionial approach (e.g., K-means).

Hierarchical clustering gives some advantages over other clustering methods because of revealing outputs in a graded form. Also, it does not require the number of clusters to be previously specified. Another advantage of hierarchical cluster comes from its simplicity due of having a low complexity if comparing with other algorithms. Hierarchical clustering algorithms produce a

nested series of partitions for merging or splitting clusters based on the similarity [103]. Partitional clustering algorithms identify the partition that optimises a clustering criterion. The preference of one approach over another depends on the task on hand and on the main goal of the study. As in our case we do not know the number of clusters in advance and we do not have time complexity issue for the maps in our dataset we prefer using hierarchical clustering. There are also some hybrid approaches discussed in the literature, as in [104] and in [105]. In this study we explore only hierarchical clustering algorithms as they suit the best our task.

The most widely used linkage methods are: single, complete and ward linkage methods [102]. The ward method is not efficient for our study as we do not use the Euclidian distance for similarity measurement. To cluster FCMs the single-linkage method takes the distance between two clusters as the minimum of the distances between all pairs of maps from the two clusters. On the other hand, the complete-linkage algorithm calculates the distance between two clusters as the maximum of all pairwise distances between maps in the two clusters. Single-linkage suffers from a chaining effect producing elongated clusters. The study in [102] shows that the complete-linkage produces more compact and more useful hierarchies in many applications than the single-linkage algorithm. Besides to decide which algorithm suits the data of the study best, we calculated the Cophenetic Coefficient (CC) for single, complete, weighted, average, ward and centroid linkage methods.

Note that the CC shows how strong the linking of maps in the cluster tree is correlated with the distances between the maps in the distance vector. This coefficient usually is used to compare different linkage methods. The closer the CC to “1” the more accurately the clustering solution reflects the data. For our dataset, the results of CC calculation show that the complete and weighted methods are the best for our data as they gave maximum values for CC. However, the weighted method also suffers from chaining effect, and the best option that we used for our cluster analysis is the complete linkage.

6.3. Finding the optimum number of clusters

After clustering FCMs there are several important questions to be considered: How good is the clustering? What is the optimum number of clusters? Or what are the main patterns to be explored in the clusters? There are several cluster validity indexes that lead to the decision of the optimal number of clusters to be determined [106]. We will explore two of them, the *Silhouette Index* (SI) and *Davies-Bouldin Index* (DBI), see [107] and [108].

Accordingly, in this section we discuss the optimal number of the clusters and we also propose the adjustment of the well-known DBI. First we find the optimal number of clusters; second, we use the *concept of central map* to analyse the clusters separately. We also compare in this section the results of the two indexes (SI and the proposed extension of DBI to be used for FCMs). For each sample in a cluster the SI will assign a confidence value showing how good the sample has been classified. The SI for i^{th} sample in cluster X_j ($j=1, \dots, c$) is defined in (10):

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (10)$$

where c is the number of clusters, $a(i)$ is the average distance between i^{th} sample and all other samples included in X_j , $b(i)$ is the minimum average distance among all samples in cluster X_k ($k=1, \dots, c; k \neq j$). The closer $s(i)$ to “1” the better the i^{th} sample assigned to its cluster. Showing the heterogeneity and isolation properties, the cluster SI is defined as follows in (11).

$$S(j) = \frac{1}{m} \sum_{i=1}^m s(i) \quad (11)$$

To find the optimal number of clusters the average value of SI is calculated for different number of clusters, and the one with the maximum value is taken as the optimum number. As SI, DBI also aims to find the optimal number of clusters. DBI is defined as follows in (12):

$$DBI = \frac{1}{c} \sum_{i=1}^c \max_{i \neq j} \left\{ \frac{\Delta(X_i) + \Delta(X_j)}{\delta(X_i, X_j)} \right\} \quad (12)$$

where $\Delta X_i/\Delta X_j$ is the average distance of the samples in the i^{th}/j^{th} cluster to the centre of the cluster, $\delta(X_i, X_j)$ is the distance between the centres of the i^{th} and j^{th} clusters. The cluster configuration minimising DBI is taken as the optimal number of discovered clusters. As we are clustering FCMs, it was needed to adjust the DBI to be applied for FCMs. We propose the concept of a central map so that we are able to use (22): hence we first derive a central map for each cluster, then calculate the distance of each map in a cluster from its central map, the distance among the central maps of different clusters and derive BDI index defined above.

To find the central map first we do some observations about the structure of the maps that we have for our task. Note that all maps in our dataset have the same hierarchical structure, some nodes are only transmitters ('situation' nodes), some are only receivers, and some others are both transmitters and receivers. A node is included in the central map if it exists in more than half of maps in a cluster.

The weights of the links are calculated as the average value of the weights from all maps that contain both nodes that comprise the link. Once we have all central maps of all clusters we calculate BDI for different number of clusters to find the optimum number. Note that once we identify the number of clusters, we use central maps for further analysis of each cluster. In the next section we show the results explained here and above sections about clustering of FCMs.

6.4. Cluster estimation and validation

In this section we explain the results of clustering the 221 FCMs, the implementation was prepared in Matlab^{††} environment. The dendrogram in Figure 7 illustrates the arrangement of the clusters produced by (8) for the distance matrix and (9) by complete linkage for inter-cluster distance.

Note that we also tested other linkage methods on our data set. However, the results were not satisfactory as the resulting clusters were not compact as some of the clusters had only few maps while the others were bigger in size. To find the optimal number of clusters, we calculated the SI for 10 clusters. Moreover, we derived

also DBI for again 10 clusters first finding the central maps for each cluster configuration. After finding the optimal number of clusters, we first analyse the 'situation' - 'benefit' - 'attribute' sets for each cluster; afterwards, we analyse the demographic features of the clusters. For that we use central maps that we derived for DBI: thus, we have six maps representing each cluster.

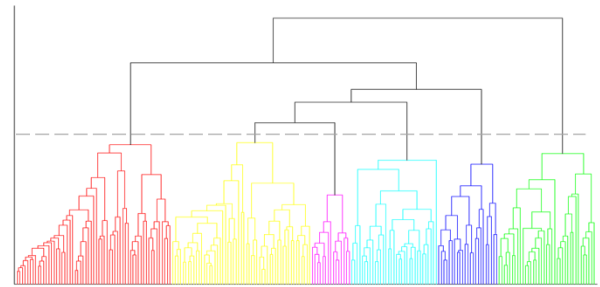


Fig. 7. Clustering results' interpretation with a dendrogram.

For demographic analysis we take into consideration the users' age, gender, income, household size, education level, occupation, parking type (paid or free), bike ownership, number of owned cars and having a bus card. The results of the chi-square test indicate that some of these variables namely the income, occupation, bus card, parking availability and number of cars are dependent variables. The optimum Number of Clusters (NC) with both SI and DBI is found equal to six (see Figure 8) with 15, 53, 33, 23, 59 and 38 maps in each cluster respectively.

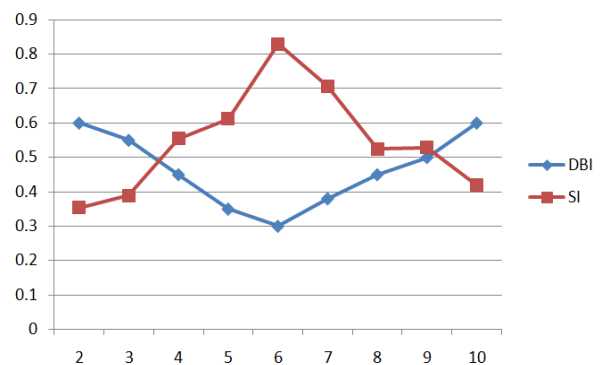


Fig. 8. SI and DBI values for different NC.

As this study focuses on clustering the maps generated from individuals' cognitive representations, the hierarchical clustering technique allows us to understand the hierarchical relationships among them. It also allows us to group people while recalling their

^{††} <http://www.mathworks.com/>

identification numbers, enabling to trace back to their cognitive subset and personal data to perform subsequent analyses. Later on the paper, we explain each of six clusters in more details in terms of maps structure, nodes, links as well as demographic features of the users in each cluster.

7. Aggregating procedure for Fuzzy Cognitive Maps

There are several characteristics making FCMs peculiar among other techniques, for example, as B. Kosko recently expressed, its feedback structure distinguishes it from the earlier forward-only acyclic CMs and from modern AI expert-system search trees. Such tree structures are not dynamic systems because they lack edge cycles or closed inference loops, nor are trees closed under combination [109]. Moreover, combining several trees does not produce a new tree in general because cycles or loops tend to occur as the number of combined trees increases. But combining FCMs always produces a new FCM. The combined FCM naturally averages^{§§} the FCMs and their corresponding causal descriptions as well as much of their dynamics (basic example on Figure 9).

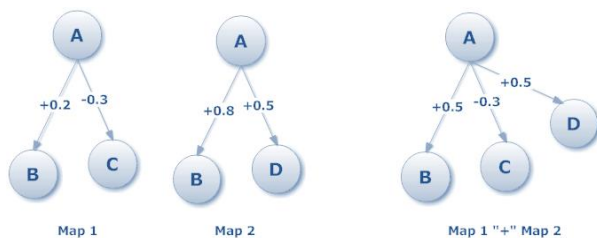


Fig. 9. Combining two basic FCMs.

Definitely to combine multiple FCMs into a single one, in order to obtain a multi-knowledge representation of a system as a whole reports several benefits. We propose a new idea to aggregate FCMs; the performance of the proposed algorithm is examined and experimental results show its ability to combine FCMs into an effective structure.

The fusion or aggregation of FCMs is considered an advantage over other methods, where combining the structures becomes a problematic issue for several

reasons. For example, the user can combine any number of FCMs, merging them into a single FCM by the simple artifice of “adding” their scaled and augmented (zero-padded) adjacency edge matrices [110]. The strong law of large numbers then ensures that the sample average of even quantised or rounded-off FCMs will converge with probability “1” to the underlying but often unknown FCM that generates these matrix realisations. Accordingly, the edge-matrix combination of FCMs improves with sample size. The knowledge representation of FCMs likewise tends to improve as the user combines more FCMs from an ever larger pool of domain experts [109].

Developing a reliable model of FCMs for a given system is a challenging task. Models established by a single expert are vulnerable to bias and inaccuracy. Credibility of modelling with FCMs can be improved by aggregating maps from multiple experts [111]. Thus, FCMs allow for a relatively simple knowledge aggregation, obtained from several experts. The aggregation of FCMs aims at improving the reliability of the final model which is less susceptible to potentially erroneous beliefs of a single expert [112]. It is common that experts evaluate different set of concepts.

Consequently, the sizes of the corresponding matrices may not be the same and/or the corresponding rows/columns may refer to different concepts. There are a couple of procedures for combining multiple FCMs into a single final model [113]. They involve simple matrix operations, such as summations and multiplications by a number, which are computed using individual connection matrices developed by different experts.

In such a case, the first step towards the aggregation of maps is to equalise their sizes. Each connection matrix is augmented, if necessary, by including any missing concept(s), when compared to any other map, through the addition of extra rows and columns in the connection matrix filled with zeros. If the total number of distinct concepts for all input FCMs equals N , then each individual connection matrix is augmented to the matrix of $N \times N$ size [114].

^{§§} The average operator is presented as example and for a basic aggregation procedure, but depending on the distribution of the data, other operators could be considered.

To aggregate a set of FCMs given by experts with different credibility the proposed maps are multiplied with a nonnegative “credibility” weight. So the combination of these different FCMs will produce an augmented FCM. Methods for the aggregation of FCMs based on the dynamic properties of the individual maps have not been investigated deeply. In order to fill this gap we introduce an approach to combine FCMs which takes into account simulations of the input models.

Some other reported methods, like [115] and [116] have extended classical ones, in order to accommodate for the credibility factors of individual maps, by replacing the ordinary average with a weighted average. The weights w_i are assigned to experts proportionally to their reliability and take values from range [0,1]. Hence, experts with higher credibility have a higher influence on the structure of the aggregated map than those with lower credibility. The applicability of this method is limited by difficulties in estimating the credibility coefficients.

This section presents a method to combine different maps; the idea is simple, if a relation of a specific map is present while conforming the new one, its value is taken into account together with the already existing ones, but if it is not present yet, it is added as a new one to the coming out map. The performance of maps were studied by several experiments, these measures can be considered as a quality index when combining different maps. This criterion allows us to use the accuracy percentage as an idea of how good the map is. Thus, instead of a simple average of the map links, we obtain a weighted average using the credibility of each map.

There is not any preference if selecting their accuracy index before or after the readjustment of their classification skill, that will only depend on the purpose of the task. There is also another idea in order to have a resulting map which better represents its contributors. This time the hint is not on the structure of the map itself but in its semantic meaning. As mentioned, each map is composed by ‘situation’, ‘attribute’ and ‘benefit’ variables obtained from the knowledge engineering step, and could be that not all types of variables causes the same impact when creating the final aggregated map.

In the proposed model, besides a credibility index per map, a new parameter is introduced, restricting the inclusion of concepts for the final map, based on the amount of times a concept was present in the maps to be aggregated. So, the user, before executing the procedure for producing a new map from several ones, could specify the desired appearance percentage for each type of variables (‘situation’, ‘attribute’, ‘benefit’). As a consequence the variables that appear only in few maps are not going to be present in the final map, because they did not pass a specified threshold. By default the model takes into account all set of variables.

In this section we also discuss which configuration performs better. For the experiment, an aggregated map in each of the four discovered clusters after learning (these clusters are going to be further described) is created, by different configurations (average, weighted average, etc.). After that step, the classification accuracy (in %) of each new map is measured for all the scenarios from the contributors’ maps, using a C-V model and h_4 as the function to be minimised. Table 3 summarises the comparison, in the rows are the four clusters, and by columns different configuration of the experiment.

Table 3. Accuracy of aggregated maps.

Cluster	A	B	C	D	E	F	G	H
1	70	73	74	76	73	77	75	78
2	59	61	63	64	63	66	66	69
3	66	67	67	70	69	71	69	73
4	55	57	60	62	61	63	60	65

Where:

- A: Averages all links from selected maps; credibility is “1” by default value for all maps. All variables are considered.
- B: Weighted average. Each map has a credibility index taken when their classification accuracy was studied.
- C: Like in A, but situational variables must be in more than in the 30% of maps in order to be considered (S: 30%). Same for attribute and benefit variables.
- D: Like in B, but S: 30%, A: 30%, B: 30%.
- E: Like A, but S: 50%, A: 50%, B: 50%.
- F: Like in B, but S: 50%, A: 50%, B: 50%.

- G: Like A, but S: 50%, A: 70%, B: 30%.
- H: Like in B, but S: 50%, A: 70%, B: 30%.

So, the configuration described in A, for cluster 1, reports only 70% of accuracy, while the configuration in H reports 78%. Same idea to understand the results described in the other cells of the table. Some reported methods in literature, already referenced; use expert credibility values, but not any accuracy measure, and neither a threshold for adding specific types of nodes. Although in this study only an assessment is given, and not any statistical comparison or similar evaluation, it can be considered as a result. Summing up, an automated credibility of maps based in their classification skills and user’s criteria about importance of variables is considered for aggregating FCMs into a single structure, constituting then a promising centroid construction due to its predicting capabilities. Also, it is important to notice from the described experiment, that an approach based on weighted average of links, giving more relevance to the use of ‘benefit’, ‘situational’, and ‘attribute’ variables, in this descending order, reported a better quality of the map representing its contributors. We can conclude that the semantic meaning of variables in the problem truly gives a direct influence to the final procedure to be followed.

8. Decision making with different levels of abstraction

In this section we analyse the results of clustering FCMs before and after learning [117]: thus we investigate traveller’s preferences and similarities when they choose only the nodes and the weighted links (first case) and when in addition they evaluate also different scenarios (second case). Figure 10 shows the general approach overview and results of using clustering and learning of FCMs to analyse users’ preferences in different stages of reasoning.

As already mentioned before, we used the learning of FCMs as a predictive tool and the clustering as a descriptive tool, to analyse the preferences of travellers under specific situations corresponding to different scenarios. In addition, we can use clustering and learning results to analyse the users’ decision making preferences with different levels of abstraction. Particularly, if we analyse clustering the maps before learning, we will obtain the groups of people that are

similar to each other in their initial mental representation. However, if we go further and provide more information from users about different scenarios and ask them to evaluate their behaviour in different circumstances, we will acquire deeper knowledge representation.

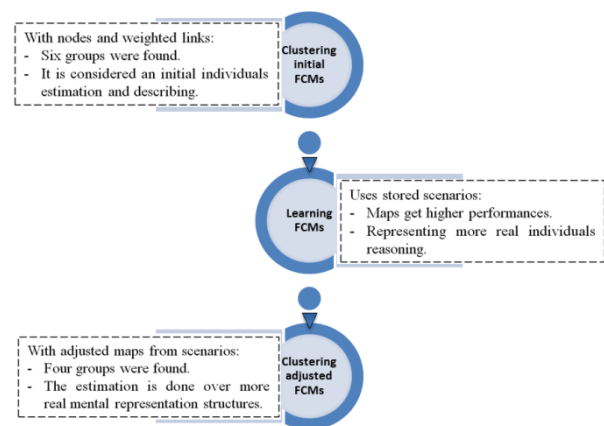


Fig. 10. Decision making analysis procedure.

8.1. Clustering before the readjustment of FCMs

Figure 11 shows one central map of the six obtained clusters (the width of the relations corresponds to the frequency of links evaluation by group members). The central map is a representation of aggregating a group of maps into one that must represent the whole collection. For simplicity of the map visualisation we omit the three decision nodes and the final utility node.

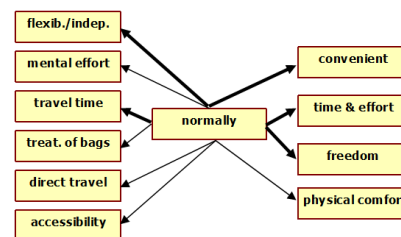


Fig. 11. Central map of one of the clusters (before learning).

As already stated in the previous section, we obtained six different clusters when we took the user provided maps, while after learning of maps and clustering them, we got only four clusters. Before the readjustment of the FCMs the travellers are profiled as follows:

- In the first group (Figure 11) we have young travellers (less than 30 years old), with medium income (from 2000 to 4000 euros) having more than two cars and a small household size (living

alone or in two). The main benefits combining the travellers in this group are to have convenience, to pay attention on the required time and effort, their desire to be free while traveling, and physical comfort. The most important attributes for them are the travel time as well as the flexibility and independence. They take into account the travel time, the treatment of bags; mental effort needed and the possibility of direct travelling, etc.

- In the second group we have mostly retired people (more than 60 years old); having only one car, with household more than two persons and who mainly use paid parking. This group has similar preferences as the previous one; paying more attention on physical comfort, easiness of parking, physical effort, and the situation of precipitation, the shelter availability of the transport mode to be used as well as the reliability under the chosen situation variables.
- In the third cluster we have the travellers who are older than 40 years. They are either employed or retired, with high income (more than 4000 euros), using mainly paid parking. The preferences of this group are very similar to the second group with the difference of choosing as a benefit: being healthy, convenience and paying no attention on reliability. Actually, this group and the previous one are in the same cluster after maps' learning.
- The fourth group is different from previous ones. This group mainly contains students or young employers with medium or high income, who mainly use free parking. There is one benefit variable unique to this group: having fun. Another difference with respect to the other groups is the set of situation variables. Namely, this group considers the situations related with precipitation, number or size of the goods purchased, as well as the available time.
- The fifth group includes the travellers with low education, preferring not to provide information about their income or declaring themselves as unemployed. This group has a unique benefit variable that is the assurance and certainty related with their decision. The unique situation that the group considers and other groups do not, is the availability of parking and some attributes that the group considers as important are: reliability,

accessibility, travel time, transport mode preference, etc.

- The last group is the only group that gives importance to safety and security, saving money, and the traveling cost. Here travellers have low education, low income (less than 2000 euros), mainly students and unemployed or retired with small household.

These were the preferences of users when they choose the variables and the links among the variables that they consider important for travel decisions. However, these preferences are changing when the users are provided with more information in terms of different possible scenarios.

8.2. Clustering after the readjustment of FCMs

Indeed, in different levels of information, there is a difference in the knowledge and experience that people tend to act. In the beginning it is more reactive response, then with deeper understanding and analysing, individuals give more situational awareness leading to more rational decisions. For example, in our case study we found that in the beginning a group of travellers consider the cost and saving money as important factors for traveling decisions. However, with awareness of different circumstances and scenarios, their preferences for the mentioned factors are decreasing. Therefore, the travellers with these preferences (the sixth cluster described above) are distributed among other clusters: thus only four cluster groups have been identified with adjusted maps after applying the PSO learning algorithm (see Figure 12), with 25, 64, 42 and 90 maps in each cluster. As a consideration remark, we should notice that for policymakers it is important to distinguish in which case they have to refer to the initial clusters and when to the clustering of the adjusted maps.

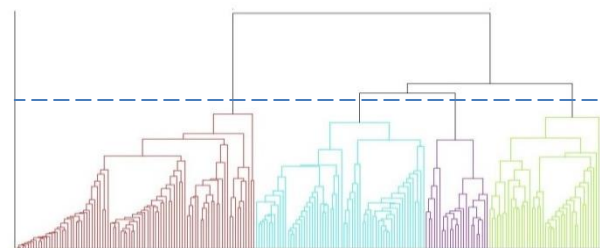


Fig. 12. Dendrogram of the clustering after the learning process.

Clustering of learned maps takes into account different scenarios; therefore it gives a more comprehensive view of travellers' subconscious decision making preferences. Consequently, policymakers can refer to these clustering results when the task is related with a complex policy testing. On the other hand, the clustering of non-learned maps can be used when testing new policies related with simpler measures.

The most important benefits in both clustering experiments: 'freedom', 'physical comfort', 'time and effort' and 'convenience'. The relevant situations in both cases: 'normally', 'precipitation' and 'time available'. The most important attributes: 'flexibility/independence', 'accessibility', 'treatment of bags', 'travel time', 'easiness for parking' and 'shelter'. In the section through a cascading of the technique applications, an integration of learning, clustering and aggregating was presented. The implementation in the case study allows the reader for a profounder understanding of the intention of each of the applied methods. From this practical perspective the process of knowledge discovering outcomes with useful information that can be used for future regulations and testing of scenarios.

9. Conclusions

Increasing concerns over the continuous growth in private transport use and the progressively insufferable externalities have produced growing attention in how transport planning strategies might at least moderate the weights in growth of personal mobility and support the doctrines of sustainable development. In this paper we have argued how cognitive mapping research has the potential to address the enduring focus on accessibility in transportation research. While accessibility has traditionally been conceived simply as proximity (or cost) from one location to others, cognitive mapping research shows that physical distances are only one factor shaping how individuals make choices in their spatial context. Many differences, including prior modal travel experiences, cultural preferences, and spatial abilities, outline the cognitive map and, in that way, the cognitive proximity and accessibility of possible desired destinations.

The automated methods used in the Knowledge Engineering, in occasions, can end up being more

competent than the humans to acquire and to refine specific types of knowledge. They can reduce the high cost significantly in human resources that it brings the construction of Knowledge Based Systems. In this research, the satisfactory use of the Automated Knowledge Engineering to extract mental representations has been revealed as an interesting way of automatizing the cognitive map formalisation. Considering this, an Automated Knowledge Engineer implementation was developed to acquire individuals' mental representations about travel behaviour; it is an advantageous tactic to homogeneously, rapidly, and efficiently collect individuals' records as a flat mental representation.

In the conducted study in the city of Hasselt, the participants' socio-demographic and travel behaviour information are collected, the developed implementation elicits individuals' mental representations based on their decision making procedure, revealing all considered cognitive subsets that compose a whole reasoning mechanism.

From the previously generated Knowledge Bases, it is possible to directly build structures based on Fuzzy Cognitive Maps on behalf of each individual way of reasoning, taking into account all factors and also the relationships among them. A computational deduction simulation mechanism is established, in order to understand the behaviour of people when in front of a transportation decision problem. In this study we proposed Fuzzy Cognitive Maps as a modelling tool to analyse the behaviour of complex systems, where it is very difficult to describe the entire system by a precise mathematical model. Consequently, it is easier and more practical to represent the decision making model in a graphical way, analysing the main concepts that affect the travellers' choice of a specific transport mode such as car, bike or bus.

Also, we explored an evolutionary technique as a machine learning method in order to readjust the causal matrix representing each map. We showed the benefits of the application of the learning method inspired by the Particle Swarm Optimisation metaheuristic, obtaining an improvement over the original modelled knowledge structures. It is a weight adaptation methodology that has been introduced to fine-tune the causal links of

Fuzzy Cognitive Maps. Accompanied with the good knowledge of a given system or a process, it can contribute towards the establishment of a robust modelling technique. In comparison with other applicable approaches, the proposed one exhibits a better performance, statistically proved. Specifically the second phase in the two steps model has been beneficial, where after the initial modelling of the system, a reconfiguration of parameters provides a cognitive model able to predict better the user's preferences. Experimental results based on simulations verified the effectiveness, validity and advantageous behaviour of the proposed algorithm. The area dealing the learning of Fuzzy Cognitive Maps is still very promising because the obtained maps are directly interpretable by humans and are useful to extract information from data about the relations among concepts or variables.

In addition, a clustering methodology has been introduced. The results can be used by policymakers to analyse the preferences of different groups of travellers in terms of their preferred benefit, situation and attribute variables. We used the concept of a central map as a representative for each group. Some of the suggestions from this study are related with several improvements which according to the travellers will increase their will to use public transportation more frequently. Namely, the travel time, the easiness of parking, the direct travel, the treatment of bags, the independence, the shelter/staying dry, direct travel, the physical effort, etc. These were the most important (with different strength) situation variables for almost all groups: thus improving those factors will certainly increase the use of public transportation for all kinds of travellers. Note that, the cost which is an important decision factor in many life-situations was mentioned as important only by one group of travellers. Mainly saving money is not the main concern while choosing a certain travel mode but other external factors which can be improved by policymakers. With an increasing concern of air pollution, fuel and car dependency, these kinds of studies will contribute to the development of a better infrastructure for city transportation facilities.

A procedure for aggregation of maps is described, considering a credibility degree for each map, depending of its real performance when assessing user's

scenarios in the machine learning tests. The idea of obtaining one map representing a specific group results convenient because the prototype map reflexes its contributors and it is easier to deal with one structure than with a set of them. This advantage of Fuzzy Cognitive Maps over other techniques was well exploited and could be used by policymakers in various applications and purposes.

A cascade experiment was performed, obtaining different cluster results and central maps at two different stages, before and after applying the learning algorithm. The comparison shows that travellers change their preferences while providing more information of different scenarios. These results provide an idea of how it seems the individuals have more differences for travel preferences, when deeper analysing inside their decision making mechanism these differences decreased.

We used clustering and learning of FCMs to dig deeper inside travellers' minds while making a decision of a transport mode and to offer policymakers a framework and real data to deal with, in order to study and simulate individuals' behaviour and to produce important knowledge to be used in the development of city infrastructure and demographic planning.

References

1. Gutiérrez, J., Análisis de los efectos de las infraestructuras de transporte sobre la accesibilidad y la cohesión regional. Estudios de Construcción y Transportes, Ministerio de Fomento, España. 2006.
2. Janssens, D., E. Hannes, and G. Wets, Tracking Down the Effects of Travel Demand Policies. 2008.
3. Hannes, E., D. Janssens, and G. Wets, Proximity is a State of Mind: Exploring Mental Maps in Daily Activity Travel Behaviour. 2006.
4. Chorus, C., An Empirical Study into the Influence of Travel Behavior on Stated and Revealed Mental Maps, in 88th Annual Meeting of the Transportation Research Board. 2009.
5. León, M., R. Bello, and K. Vanhoof, Cognitive Maps in Transport Behavior, in Proceedings of the 2009 Eighth Mexican International Conference on Artificial Intelligence, IEEE Computer Society. pp. 179–184. 2009.
6. Dijst, M., Spatial policy and passenger transportation. Journal of Housing and the Built Environment. Vol. 12. pp. 91–111. 1997.

7. Bradley, M., Process Data for Understanding and Modelling Travel Behavior. *Travel Survey Methods: Quality and Future Directions*. Elsevier Science. pp. 491–510. 2006.
8. Torra, V. and Y. Narukawa, Modeling Decisions. *Information Fusion and Aggregation Operators*. Springer-Verlag Berlin Heidelberg. 2007.
9. Aguilar, J., A Survey about Fuzzy Cognitive Maps Papers. *International Journal of Computational Cognition*. Vol. 3. 2005.
10. León, M., R. Bello, and K. Vanhoof, Considering Artificial Intelligence Techniques to perform Adaptable Knowledge Structures. *World Scientific Proceedings Series on Computer Engineering and Information Science - Vol. 2. Intelligent Decision Making Systems*. pp 88–93. 2009.
11. Zadeh, L.A., K.-S. Fu, and K. Tanaka, *Fuzzy Sets and their Applications to Cognitive and Decision Processes*: Academic Press. 1975.
12. Jang, J.-S.R., ANFIS: Adaptive-Network-Based Fuzzy Inference System. *IEEE Trans. on Systems, Man and Cybernetics*. Vol. 23, No. 3. pp. 665–685. 1993.
13. Kosko, B., Fuzzy Cognitive Maps. *International Journal of Man-Machine Studies*. Vol. 24. pp. 65–75. 1986.
14. Stylios, C.D., et al., Fuzzy cognitive map architectures for medical decision support systems. *Applied Soft Computing*. Vol. 8. pp. 1243–1251. 2008.
15. Xirogiannis, G., M. Glykas, and C. Staikouras, Fuzzy Cognitive Maps as a Back End to Knowledge-based Systems in Geographically Dispersed Financial Organizations. *Knowledge and Process Management*. Vol. 11. pp. 137–154. 2004.
16. Contreras, J., et al., Realistic Ecosystem Modelling with Fuzzy Cognitive Maps. *International Journal of Computational Intelligence Research*. pp. 139–144. 2007.
17. León, M., et al., A Revision and Experience using Cognitive Mapping and Knowledge Engineering in Travel Behavior Sciences. “POLIBITS” Research journal on Computer Science and Computer Engineering with Applications, 2010.
18. Arbid, M.A., *The Handbook of Brain Theory and Neural Networks*. The MIT Press. 2003.
19. Tzafestas, S.G., A.N. Venetsanopoulos, and S. Terzaki, *Fuzzy Reasoning in Information, Decision and Control Systems*. Kluwer Academic. 1994.
20. Golledge, R.G. and T. Garling, Spatial Behavior in Transportation Modeling and Planning. Chp. 3. *Transportation and Engineering Handbook*. CRC Press. Vol. 46, pp. 0–37. 2001.
21. Hannes, E., Mental Maps and daily travel: Qualitative exploration and modelling framework. (PhD dissertation). Hasselt University. 2010.
22. Kusumastuti, D., Scrutinizing fun-shopping travel decisions: Modelling individuals’ mental representations. (PhD dissertation). Hasselt University. 2011.
23. Fries, R., M. Chowdhury, and J. Brummond, *Transportation Infrastructure Security Utilizing Intelligent Transportation Systems*. John Wiley & Sons, Inc. 2009.
24. Hannes, E., León Espinosa, M., Kusumastuti, D., Janssens, D., Vanhoof, K., & Wets, G., Modelling Multiple Meanings of Mental Maps, in 12th International Conference on Travel Behaviour Research. IATBR, Jaipur, India. 2009.
25. Waskan, J.A., *Models and Cognition: Prediction and Explanation in Everyday Life and in Science*. MIT Press. 2006.
26. Ghosh, S. and T.S. Lee, *Intelligent Transportation Systems: New Principles and Architectures* Mechanical Engineering Handbook Series. CRC Press. 2000.
27. Aldian, A. and M.A.P. Taylor, Fuzzy Multicriteria Analysis for Inter-City Travel Demand Modelling. *Journal of the Eastern Asia Society for Transportation Studies*. Vol. 5. 2003.
28. Horeni, O., et al., Design of a Computer-Assisted Instrument for Measuring Mental Representations Underlying Activity-Travel Choices, in 8th International Conference on Survey Methods in Transport Annecy. 2008.
29. Arentze, T., Modeling and Measuring Individuals Mental Representations of Complex Spatio-Temporal Decision Problems. *Environment and Behavior*. Vol. 40, No. 6. pp. 843–869. 2008.
30. Janssens, D., et al., Improving the Performance of a Multi-Agent Rule-Based Model for Activity Pattern Decisions Using Bayesian Networks. *TRB Annual Meeting*. 2003.
31. Davis, R., *Knowledge-based systems in Artificial Intelligence: 2 Case Studies*. McGraw-Hill, Inc. New York, NY, USA. 1982.
32. Forsythe, D., *Engineering Knowledge: The Construction of Knowledge in Artificial Intelligence*. Social Studies of Science. Vol. 23. pp. 445–477. 1993.
33. Woolf, B.P., *Building Intelligent Interactive Tutors: Student-centered strategies or revolutionizing e-learning*. Elsevier Inc. 2009.
34. Wagner, C., Breaking the Knowledge Acquisition Bottleneck Through Conversational Knowledge Management. *Information Resources Management Journal*, Vol. No. 1. 2008.
35. Castillo, E., J.M. Gutiérrez, and A.S. Hadi, *Expert Systems and Probabilistic Network Models*: New York: Springer Verlag. 1996.
36. Hoppenbrouwers, S.J.B.A. and P.J.F. Lucas, Attacking the Knowledge Acquisition Bottleneck through Games-For-Modelling, in AISB’09 workshop “AI and Games”. 2009.

37. Gomes, M.E., A human factors evaluation using tools for automated knowledge engineering. NAECON IEEE Aerospace and Electronics Conference. Vol. 2 p. pp. 661–664. 1993.
38. Chandana, S., R.V. Mayorga, and C.W. Chan, Automated Knowledge Engineering International Journal of Computer and Information Engineering. Vol. 2, No. 6. pp. 370–379. 2008.
39. Abidi, S., Y.-N. Cheah, and J. Curran, A Knowledge Creation Info-Structure to Acquire and Crystallize the Tacit Knowledge of Health-Care Experts. IEEE Transaction on Information Technology in Biomedicine. Vol. 9, No. 2. pp. 193–204. 2005.
40. Mishra, B.K., Automated knowledge engineering using rough set approach. ICM2CS International Conference on Methods and Models in Computer Science. pp. 14–17. 2010.
41. De Ceunynck, T., The description of individuals' cognitive subsets in fun shopping activities by making use of association rules algorithms: Case study in Hasselt, Belgium. Master's thesis. Hasselt University. 2010.
42. León, M., et al., Mapas Cognitivos Difusos aplicados a un problema de Comportamiento de Viajes. III Taller Internacional de Descubrimiento de Conocimiento, Gestión del Conocimiento y Toma de Decisiones. Eureka Iberoamérica. Universidad de Cantabria, Santander, España. 2011.
43. Axelrod, R., Structure of Decision: The Cognitive Maps of political Elites, Princeton University. 1976.
44. Eden, C., Cognitive Mapping: a review. European Journal of Operational Research. Vol. 36. pp. 1–13. 1988.
45. Eden, C., On the Nature of Cognitive Maps. Journal of Management Studies. Vol. 29. pp. 261–265. 1992.
46. Wei, Z., L. Lu, and Z. Yanchun, Using fuzzy cognitive time maps for modeling and evaluating trust dynamics in the virtual enterprises. Expert Systems with Applications. Elsevier Ltd. pp. 1583–1592. 2008.
47. Beena, P. and R. Ganguli, Structural damage detection using fuzzy cognitive maps and Hebbian learning. Applied Soft Computing. Vol. 11. pp. 1014–1020. 2011.
48. Tsadiras, A.K., Using Fuzzy Cognitive Maps for E-Commerce Strategic Planning. 2007.
49. Peña, A., H. Sossa, and F. Gutierrez, Ontology Agent Based Rule Base Fuzzy Cognitive Maps KES-AMSTA. Springer-Verlag Berlin Heidelberg. pp. 328–337. 2007.
50. Sadiq, R., Y. Kleiner, and B. Rajani, Estimating Risk of Contaminant Intrusion in Distribution Networks Using Fuzzy Rule-Based Modeling. NATO Advanced Research Workshop on Computational Models of Risks to Infrastructure. pp. 318–327. Primosten, Croatis. 2006.
51. Carvalho, J.P., M. Carola, and J.A.B. Tomé, Forest Fire Modelling using Rule-Based Fuzzy Cognitive Maps and Voronoi Based Cellular Automata NAFIPS 2006 2006 Annual Meeting of the North American Fuzzy Information Processing Society. pp. 217–222. 2006.
52. Peña, A. and H. Sossa, Negotiated Learning by Fuzzy Cognitive Maps, in IASTED International Conference. Grindelwald, Switzerland. 2005.
53. Grant, D., Using Fuzzy Cognitive Maps to Assess Mis Organizational Change Impact, in 38th Hawaii International Conference on System Sciences. 2005.
54. Sadiq, R., Y. Kleiner, and B.B. Rajani, Fuzzy cognitive maps for decision support to maintain water quality in ageing water mains. 4th International Conference on Decision-Making in Urban and Civil Engineering, pp. 1–10. Porto, Portugal. 2004.
55. Balder, D. Fuzzy Cognitive Maps and their uses as Knowledge Mapping Systems and Decision Support Systems. 2004.
56. Laureano-Cruces, A.L., J. Ramírez-Rodríguez, and A. Tern-Gilmore Evaluation of the Teaching-Learning Process with Fuzzy Cognitive Maps. pp. 922-931. 2004.
57. Rodríguez-Repiso, L., R. Setchi, and J.L. Salmeron, Modelling IT Projects Success with Fuzzy Cognitive Maps. Expert Systems with Applications, 2006.
58. Mateou, N.H., M. Moiseos, and A.S. Andreou, Multi-objective evolutionary fuzzy cognitive maps for decision support. IEEE Computer Society. pp. 824–830. 2005.
59. Groumpos, P.P., N. Christova, and C. Stylios, Implementation of Fuzzy Cognitive Maps for Production Planning of Plant Control Systems. MED. 2003.
60. Siraj, A., S.M. Bridges, and R.B. Vaughn, Fuzzy Cognitive Maps for Decision Support In An Intelligent Intrusion Detection System, in Joint 9th International Fuzzy Systems Association World Congress and the 20th North American Fuzzy Information Processing Society International Conference on Fuzziness and Soft Computing in the New Millennium, Vancouver, Canada. 2001.
61. Kardaras, D. and G. Mentzas, Using Fuzzy Cognitive Maps to Model and Analyse Business Performance Assessment. Advances in Industrial Engineering Applications and Practice II. pp. 63–68. 1997.
62. Peláez, C.E. and J.B. Bowles, Using Fuzzy Cognitive Maps as a System Model for Failure Modes Effects Analysis. Information Sciences. 1996.
63. Stylios, C.D., V.C. Georgopoulos, and P.P. Groumpos, The Use of Fuzzy Cognitive Maps in Modeling Systems. 5th IEEE Mediterranean Conference on Control and Systems. 1997.
64. Kosko, B., Fuzzy Thinking: Hyperion. 1993.
65. Aguilar, J., A Dynamic Fuzzy-Cognitive-Map Approach based on Random Neural Networks. International Journal of Computational Cognition. pp. 91–107. 2003.

66. Kandasamy, W.B.V., F. Smarandache, and K. Ilanthenral, Elementary Fuzzy Matrix Theory and Fuzzy Models for Social Scientists. Automaton. 2007.
67. Stylios, C.D. and P.P. Groumpos, Mathematical Formulation of Fuzzy Cognitive Maps, in 7th Mediterranean Conference on Control and Automation, Haifa, Israel. 1999.
68. Tsadiras, A.K., Inference using Binary, Trivalent and Sigmoid Fuzzy Cognitive Maps in Information Sciences, pp. 3880–3894. 2008.
69. Schneidera, M., et al., Automatic construction of FCMs. Fuzzy Sets and Systems. Vol. 9. pp. 161–172. 1998.
70. León, M., et al., Fuzzy cognitive maps for modeling complex systems. Advances on Artificial Intelligence, Part I. Lecture Notes in Artificial Intelligence. Springer-Verlag Berlin Heidelberg. Vol. 6437, pp. 166–174. 2010.
71. David, R. and H. Alla, eds. Discrete, Continuous, and Hybrid Petri Nets. Springer. 2010.
72. Vidal, J., M. Lama, and A. Bugarín, Toward the use of Petri nets for the formalization of OWL-S choreographies. Knowledge and Information Systems. pp. 1–37. 2011.
73. Wang, T. and J. Yang, A heuristic method for learning Bayesian networks using discrete particle swarm optimization. Knowledge and Information Systems. Vol. 24, No. 2. pp. 269–281. 2010.
74. Xi, R., et al., Compression and aggregation of Bayesian estimates for data intensive computing. Knowledge and Information Systems. pp. 1–22. 2011.
75. León, M., et al., Uso de Mapas Cognitivos Difusos para modelar Representaciones Mentales que caracterizan el Comportamiento de Viajes, in Segundo Taller Cubano Eureka. 2010.
76. Koulouriotis, D., et al., Efficiently Modeling and Controlling Complex Dynamic Systems Using Evolutionary Fuzzy Cognitive Maps. International Journal of Computational Cognition. pp. 41–65. 2003.
77. Koulouriotis, D., et al., Efficiently modeling and controlling complex dynamic systems using evolutionary fuzzy cognitive maps. The ABC of Computational Pragmatics. Vol. 1. pp. 41–65. 2003.
78. Kennedy, J. and R. Eberhart, Particle Swarm Optimization, in IEEE International Conference on Neural Networks, Vol. 4. pp. 1942–1948. Australia. 1995.
79. Mohan, B.C. and R. Baskaran, A survey: Ant Colony Optimization based recent research and implementation on several engineering domain. Expert Systems with Applications, Vol. 39, No. 4. pp. 4618–4627. [dx.doi.org/10.1016/j.eswa.2011.09.076](https://doi.org/10.1016/j.eswa.2011.09.076). 2012.
80. Saka, E. and O. Nasraoui, A recommender system based on the collaborative behavior of bird flocks. CollaborateCom 6th International Conference on Collaborative Computing: Networking, Applications and Worksharing. IEEE. 2010.
81. Chen, Y.-W., et al., Application of Interactive Genetic Algorithms to Boid Model Based Artificial Fish Schools. KES 12th International Conference on Knowledge-Based Intelligent Information and Engineering Systems. Lecture Notes in Computer Science, Springer. Vol. 5178. 2008.
82. Eberhart, R. and Y. Shi, Comparing Inertia Weights and Constriction Factors in Particle Swarm Optimization. IEEE Congress on Evolutionary Computation. Vol. 1. pp. 84–88. 2000.
83. Clerc, M. and J. Kennedy, The particle swarm - explosion, stability, and convergence in a multidimensional complex space, in IEEE Transactions on Evolutionary Computation. pp. 58-73. 2002.
84. Fan, S.-K.S. and J.-M. Chang, A modified particle swarm optimizer using an adaptive dynamic weight scheme. 1st International Conference on Digital Human Modeling. Springer-Verlag Berlin Heidelberg. pp. 56–65. 2007.
85. Bratton, D. and J. Kennedy, Defining a Standard for Particle Swarm Optimization. SIS IEEE Swarm Intelligence Symposium. pp. 120–127. 2007.
86. Parsopoulos, K.E., et al., A First Study of Fuzzy Cognitive Maps Learning Using Particle Swarm Optimization, in IEEE Congress on Evolutionary Computation, IEEE Press. pp. 1440–1447. 2003.
87. Huelva, A.V., A Balanced Differential Learning algorithm in Fuzzy Cognitive Maps, in 16th International Workshop on Qualitative Reasoning. 2002.
88. Parenthöen, M., C. Buche, and J. Tisseau, Action Learning for Autonomous Virtual Actors, in ISRA, Toluca, Mexico. 2002.
89. Papageorgiou, E.I. and P.P. Groumpos, A weight adaptation method for fuzzy cognitive map learning. Springer-Verlag. 2005.
90. Eberhart, R. and J. Kennedy, A New Optimizer using Particle Swarm Theory. Sixth International Symposium on Micro Machine and Human Science. pp. 39–43. 1995.
91. Razali, N.M. and Y.B. Wah, Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. Journal of Statistical Modeling and Analytics. Vol. 2. No.1. pp. 21–33. 2011.
92. Mrówka, E. and P. Grzegorzewski, Friedman's test with missing observations. EUSFLAT 4th Conference of the European Society for Fuzzy Logic and Technology. pp. 621–626. 2005.
93. Rosner, B., R.J. Glynn, and M.-L.T. Lee, Incorporation of Clustering Effects for the Wilcoxon Rank Sum Test: A Large-Sample Approach. Biometrics. Vol. 59. pp. 1089–1098. 2003.

94. Demsar, J., Statistical Comparisons of Classifiers over Multiple Data Sets. *Journal of Machine Learning Research*. Vol. 7. pp. 1–30. ISSN: 1532-4435. 2006.
95. Xu, R. and D. Wunsch, Survey of clustering algorithms. *IEEE Transactions on Neural Networks*. IEEE Computational Intelligence Society. Vol. 16, No. 3. pp. 645–678. 2005.
96. Langfield-Smith, K. and A. Wirth, Measuring differences between cognitive maps. *The Journal of the Operational Research Society*. Vol. 42, No. 12. pp. 1135–1150. 1992.
97. Markíczy, L. and J. Goldberg, A Method for Eliciting and Comparing Causal Maps. *Journal of Management*. Vol. 21, No. 2. pp. 305–333. 1995.
98. Ortolani, L., et al., Analysis of Farmers' Concepts of Environmental Management Measures: An Application of Cognitive Maps and Cluster Analysis in Pursuit of Modelling Agents' Behaviour. Springer Berlin-Heidelberg. Vol. 247. pp. 363–381. 2010.
99. Alizadeh, S. and M. Ghazanfari, Using data mining for learning and clustering FCM. *International journal of computational intelligence*. Vol. 4. No. 2. 2007.
100. Eden, C., Analyzing cognitive maps to help structure issues or problems. *European Journal of Operational Research*. Elsevier. Vol. 159, No. 3. pp. 673–686. 2004.
101. Aronovich, L. and I. Spiegler, Bulk construction of dynamic clustered metric trees. *Knowledge and Information Systems*. Vol. 22. pp. 211–244. 2010.
102. Kianmehr, K., M. Alshalalfa, and R. Alhaji, Fuzzy clustering-based discretization for gene expression classification. *Knowledge and Information Systems*. Springer London. Vol. 24. pp. 441–465. 2010.
103. Czarnowski, I., Cluster-based instance selection for machine classification. *Knowledge and Information Systems*. Springer, Vol. 30. pp. 113–133. 2011.
104. Domeniconi, C., J. Peng, and B. Yan, Composite kernels for semi-supervised clustering. *Knowledge and Information Systems*. Springer London. Vol. 28. pp. 99–116. 2011.
105. Saha, S. and S. Bandyopadhyay, A new multiobjective clustering technique based on the concepts of stability and symmetry. *Knowledge and Information Systems*. Springer London. Vol. 23. pp. 1–27. 2010.
106. Bouguila, N. and D. Ziou, A countably infinite mixture model for clustering and feature selection. *Knowledge and Information Systems*. pp. 1–20. 2011.
107. Davies, D.L. and D.W. Bouldin, A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. 1, No. 2. pp. 224–227. 1979.
108. Bolshakova, N. and F. Azuaje, Cluster Validation Techniques for Genome Expression Data. *Signal Processing*. Vol. 83. pp. 825–833. 2002.
109. Glykas, M., *Fuzzy Cognitive Maps: Advances in Theory, Methodologies, Tools and Applications*. Studies in Fuzziness and Soft Computing. Springer. Vol. 247. 2010.
110. Banini, G.A. and R.A. Bearman, Application of Fuzzy Cognitive Maps to Factors Affecting Slurry Rheology. *International Journal of Mineral Processing*. Vol. 52, No. 4. pp. 223–244. 1998.
111. Yaman, D. and S. Polat, A Fuzzy Cognitive Map Approach for Effect-based Operations: An Illustrative Case. *Information Sciences*. Vol. 179, No. 4. pp. 382–403. 2009.
112. Stylios, C.D. and P.P. Groumpos, Fuzzy Cognitive Maps in Modeling Supervisory Control Systems. *Journal of Intelligent and Fuzzy Systems*. Vol. 8, No. 2. pp. 83–98. 2000.
113. Papageorgiou, E.I., et al., Brain Tumor Characterization using the Soft Computing Technique of Fuzzy Cognitive Maps. *Applied Soft Computing Journal*. Vol. 8, No. 1. pp. 820–828. 2008.
114. Lv, Z. and L. Zhou, Advanced Fuzzy Cognitive Maps Based on OWA Aggregation. *International Journal of Computational Cognition*. Vol. 5. 2007.
115. Noori, S., R.H. Amiri, and A. Bourouni, An FCM Approach to Better Understanding of Conflicts: a Case of New Technology Development. *International Journal of Business and Management*. Vol. 4, No. 3. pp. 106–115. 2009.
116. Salmeron, J.L., Augmented Fuzzy Cognitive Maps for Modelling LMS Critical Success Factors. *Knowledge-Based Systems*. Vol. 22, No. 4. pp. 275–278. 2009.
117. León, M., et al., Clustering Fuzzy Cognitive Maps for Travel Behavior Analysis. CF-WML-KD Cuba-Flanders Workshop on Machine Learning and Knowledge Discovery. Cuba. 2011.