Statistical Approach to Fuzzy Cognitive Maps

Niskanen, Vesa Antti

Springer International Publishing AG
2020-03-06

Niskanen , V A 2020 , Statistical Approach to Fuzzy Cognitive Maps . in S N Shahbazova , M Sugeno & J Kacprzyk (eds) , Recent Developments in Fuzzy Logic and Fuzzy Sets . Studies in fuzziness and soft computing , vol. 391 , Springer International Publishing AG , pp. 33-59 . https://doi.org/10.1007/978-3-030-38893-5_3

http://hdl.handle.net/10138/341319
https://doi.org/10.1007/978-3-030-38893-5_3

acceptedVersion

Statistical Approach to Fuzzy Cognitive Maps

Vesa A. Niskanen
Dept. of Economics & Management
University of Helsinki
vesa.a.niskanen@helsinki.fi

1 Introduction

Fuzzy systems have proven to be very useful in model construction over the past decades. This is by virtue of their linguistic and approximate reasoning approaches which will provide a basis for simple and user-friendly modelling. Thousands of practical applications have confirmed the goodness of these systems.

However, we still expect certain supplementary methods in fuzzy modelling in particular in the human sciences because the methods used in the natural sciences, especially in the engineering sciences, which seem to prevail in the fuzzy systems, are unable model all the specific features of the human behavior. Examples of these are the human intentions, decisions, motives and ethical aspects.

Within fuzzy systems we are also applying insufficiently such prevailing crisp methods which may enhance our fuzzy model construction or the interpretation of the model's outcomes. The conventional statistical methods, for their part, may provide resolutions to this problem area, and we will apply them below.

Statistical methods, and fuzzified statistical methods, have been applied to certain fuzzy models already [2,6,10,15,32], but still more studies are expected. Below we will apply statistical methods to fuzzy cognitive maps from the human scientific standpoint because there already seems to be several studies on these maps but they mainly stem from the methods of neural networks and engineering, and even subjective or ad hoc reasoning is performed in this context. In the quantitative human sciences, on the other hand, we aim at objectivity and also apply empirically justified methods to our data analysis. Hence, we will study how such statistical methods as regression analyses may enhance or supplement the fine-tuning and interpretation of these models. Our results may also be applied to other fuzzy systems in general.

Chapter 2 presents the basic principles of the fuzzy cognitive maps. Chapter 3 considers at theoretical level how certain statistical methods may be utilized within the fuzzy cognitive maps and also provides a concrete example. Chapter 4 concludes our examination.

2 Fuzzy Cognitive Maps

The fuzzy cognitive maps (FCM) mainly base on the ideas of Axelrod and Kosko [1,13], and they are used for simulating and forecasting such phenomena of the real world which consist of numerous variables and their interrelationships. The FCMs may also include feedback operations. The concept maps [19,20], in turn, provide an example in which case even more wide-ranging interrelationships may be used, and they are quite much applied in the human sciences.

In statistics the structural equation models, such as Mplus™, LISREL™ and AMOS™, are used for this purpose as well as time series analysis, but the FCMs are usually simpler and more robust in model construction [16,17].

The traditional Axelrod's cognitive maps base on classical bivalent or trivalent logic and mathematics and hence they can only model coarsely their interrelationships [1]. Kosko enhanced these maps by applying the Hebbian neural networks and by using the numeric concept variable values which usually range from 0 to 1, and these values denote the degrees of activation of the variables or concepts. The degrees of relationship between the variables, in turn, may range from -1 to 1 in which case the benchmarks -1, 1 and 0 denote full negative effect, full positive effect and no effect, respectively [13].

Due to the mathematical properties of the numeric cognitive maps, in iterations on the time axis the values of its concepts may oscillate (limit cycles), are chaotic or will finally become stable (fixed-point attractors). For example, in control applications our goal is usually to achieve certain fixed points or other type of stabilities.

If empiric data, history data, in a given period of time is unavailable, we only operate with *a priori* maps, and thus we apply human intuition or expertise in our constructions, otherwise we may construct *a posteriori* maps and then we apply such methods as statistics (e.g., regression and path analysis), neural networks or evolutionary computing [9,12,14,22-26]. Hence, appropriate data may yield usable FCMs in a more or less automatic manner.

On the other hand, the construction of a posteriori FCM models still await such usable methods which can yield stable interrelationships between the concepts, i.e., similar relationship outputs if the model construction with the data is repeated, because today we seem to lack this feature, but this problem is not considered here. Many FCM constructions also seem to include subjective or ad hoc interpretations and decisions, and this problem is discussed below.

The numeric FCMs can only establish monotonic causal interrelationships between the concepts, whereas fuzzy linguistic cognitive maps enable us to avoid this problem [4,27,28]. The latter approach is also more user-friendly due to its linguistic nature and, in this sense, they resemble more the concept maps [20]. We only focus on the numeric FCMs below because they provide the basis for their analyses.

If the prevailing methods are used in the numeric FCM simulations, the concept (node) values range from zero to unity and their weights or intensities of the interrelationships belong to the closed intervals -1 to 1 [9]. These weights are presented in the connection matrices. Hence, in the basic FCM computer simulations we may apply the matrix product, *,

$$V_{t+1}=f(V_t{*}M), \qquad (2.1)$$

in which the state vector, V_t, contains m concept values at time=t, M is an $m{\times}m$ connection matrix, f is the transformation function and vector V_{t+1} contains the new concept values at time=t+1 [9]. The function f is usually the logistic function,

$$f(x) = 1 / (1+exp(-\boldsymbol{\lambda}\cdot x)), \qquad (2.2)$$

or the hyperbolic tangent function

$$f(x) = (exp(\boldsymbol{\lambda}\cdot x)-exp(-\boldsymbol{\lambda}\cdot x)) / (exp(\boldsymbol{\lambda}\cdot x)+exp(-\boldsymbol{\lambda}\cdot x)) \qquad (2.3)$$

in which *exp* is the exponential function and the parameter lambda, $\boldsymbol{\lambda}$, is a positive value [9]. This function transforms the matrix product values at time=t+1 into the closed interval from 0 to 1 or -1 to 1, respectively. Below we also use (2.2), and then lambda quite often is having the values of 1 or 5 (Figg. 2.1 and 2.2). We will prefer the value of 1 because then the obtained concept values may have larger dispersions [11].

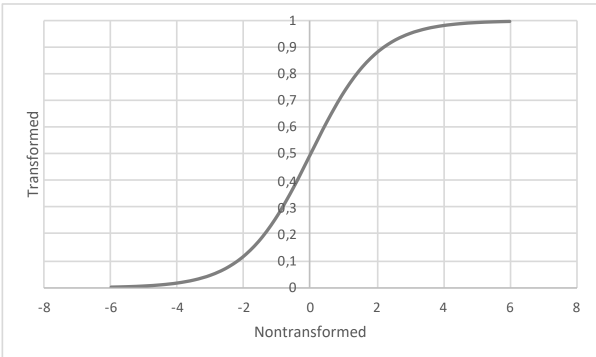

Fig. 2.1. Transformation function (2.2) when lambda=1.

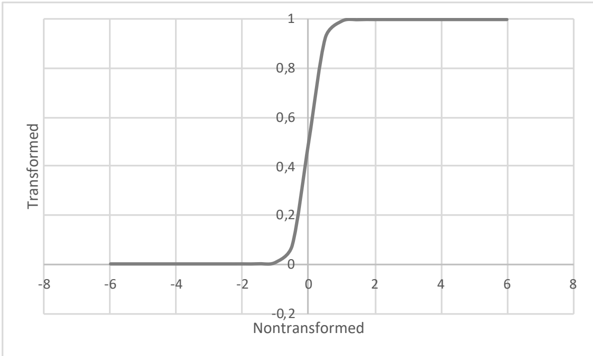

Fig 2.2. Transformation function (2.2) when lambda=5.

The mainstream studies on the FCMs seem to stem from the methods, and even the patterns of thought, of the neural networks, in particular from the Hebbian learning methods [9,13]. Hence, the interpretations on the weights in the connection matrices may be ambiguous and subjective even if empiric history data is used for their specifications. In addition, as within the neural networks, their weights may even be regarded as such black-box values which only play a secondary role when the FCMs are used in practice. In this context, the leading role of the transformation functions is only to serve as a rescaling tool for the initial concept outputs, and they are also formulated mainly on the intuitive grounds. The fine-tuning of the original FCMs also seems to rely quite often on subjective or ad hoc decisions.

In the quantitative human sciences, in turn, we apply the widely-accepted stochastic interpretations and decisions when we study our parameters, functions and model outcomes. We also aim at the stability of our models when the model construction is repeated. In this manner, we may understand well the interrelationships in our models. In practice this means that the use of the prevailing statistical methods may provide additional information for us when we study the FCMs. Below we will adopt this approach.

3 Statistical Approach to Fuzzy Cognitive Maps

The author has, also with others, studied the general features of FCMs from the statistical standpoint by using thousands of random initial concept vectors and connection matrices in computer simulations with Matlab™ and SPSS™, and the examples of these can be found in [11,18]. These studies more or less differ from the mainstream studies on the FCMs due to our human-scientific and statistical approach. In this context we only study one such random-valued FCM model, but our results may also be generalized to other FCM models.

We focus on the connection matrix in Table 3.1, whose number of concepts and weights were created randomly, because then it will represent well our general approach. The weights in the diagonal of Table 3.1 were assigned to be zeros which seems quite typical in the engineering applications even though in the real-world FCM applications of the human sciences the preceding values of concepts are often relevant in the succeeding iteration, i.e., the diagonal contains nonzero values.

Figure 3.1 depicts the graph based on Table 3.1, and the usual FCM methods mentioned in Chapter 2 yielded the concept values in Figures 3.2 and 3.3 when ten iterations and the lambda values one and five were used. This model yielded one fixed point for each of its concepts.

From the statistical standpoint, we may first consider all the possible values of the concepts at time=1 within a given cognitive map when the values at time=0 (the initial values) are given [11]. This procedure was carried out here by using 9 998 random initial concept vectors plus the vectors which only contained zeros and unities. The random vector sample was used because it was impossible to use all input vector combinations due to the numeric explosion.

Table 3.1. The original random weights of concepts.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
C1	0.00	0.54	0.97	-0.58	0.38	-0.51	-0.96	-0.06	0.73	-0.67
C2	0.49	0.00	-0.34	0.01	0.07	0.25	-0.53	-0.85	-0.05	0.58
C3	-0.54	0.94	0.00	0.72	0.39	-0.57	0.66	-0.33	-0.77	0.31
C4	-0.87	0.92	-0.94	0.00	-0.01	0.56	-0.66	-0.27	0.82	0.16
C5	-0.63	-0.71	-0.15	-0.99	0.00	-0.77	-0.74	0.45	-0.54	0.20
C6	0.06	-0.49	0.09	0.08	-0.38	0.00	-0.12	-0.74	-0.27	0.39
C7	0.65	-0.39	-0.03	0.66	-0.05	-0.84	0.00	0.24	-0.92	0.59
C8	-0.73	-0.38	-0.17	-0.14	0.65	-0.32	0.16	0.00	-0.84	0.75
C9	0.17	-0.58	0.78	0.17	0.35	0.54	0.80	0.03	0.00	0.59
C10	-0.99	-0.43	-0.83	-0.22	0.14	-0.62	0.18	0.56	0.27	0.00

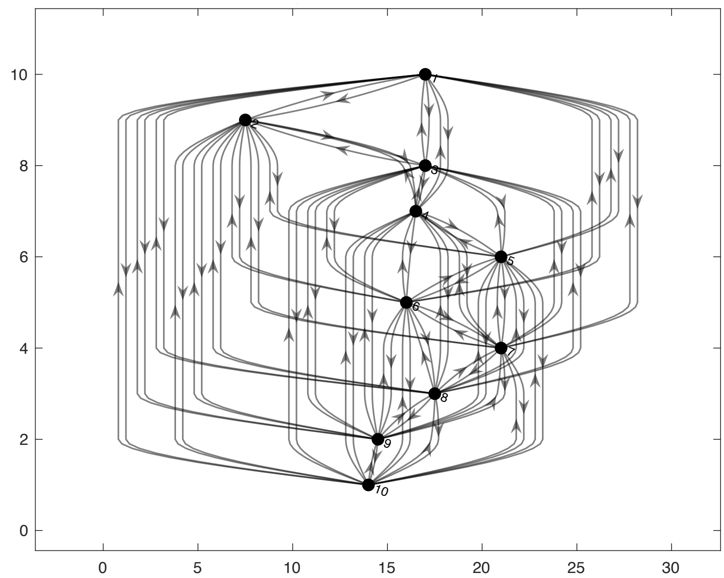

Fig. 3.1. Graph of the connection matrix in Table 3.1.

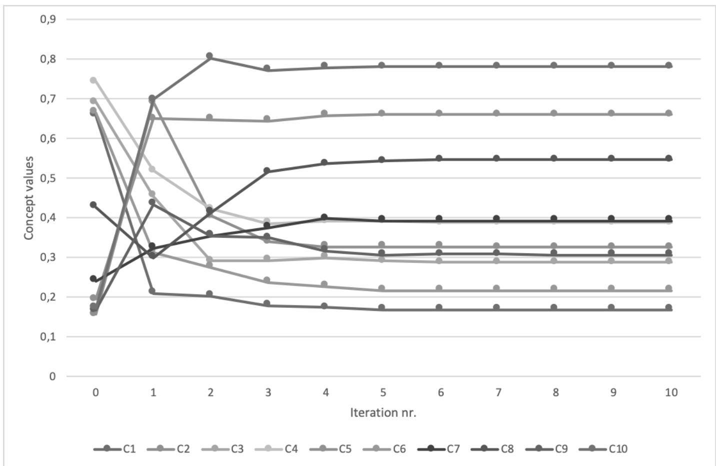

Fig. 3.2. Concept values in ten iterations when lambda=1 (iteration 0 denotes the original random values).

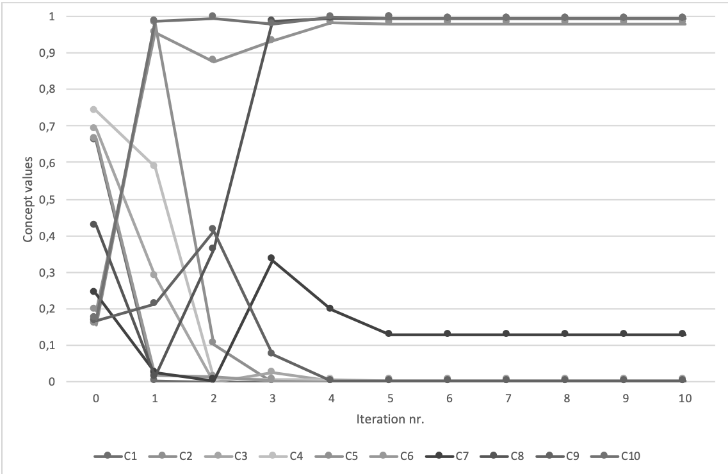

Fig. 3.3. Concept values in ten iterations when lambda=5 (iteration 0 denotes the original random values).

The weights in matrix 3.1 may have various interpretations in engineering applications, such as the degrees of intensity of the relationship, but still their real meanings in practice may nevertheless remain unclear or ambiguous. We may also perform subjective reasoning when we consider the relevance of the driver concepts for a given target concept because, for the sake of simplicity, we often aim at removing the irrelevant drivers. In other words, the simplicity means low density (many zero weights) in the connection matrix [9].

In the quantitative human sciences, in turn, we aim at objectivity and the foregoing type of subjectivity may be reduced by applying such statistical methods as regression analysis. Given a target concept, it will be the response variable (dependent variable) in a regression model, whereas the its drivers will be the predictors (independent variables). Hence, the weights in the corresponding column in the connection matrix are directly the linear regression coefficients for the given nontransformed target concept. Thanks for this approach, various useful statistics and other outcomes will be obtained in our studies.

We will consider our FCM at two stages. The first stage examines its nontransformed output values and this is more interesting analysis because these values are always, in a sense, the fixed values when the connection matrix is fixed. The second stage considers their transformed values, and then various mathematical transformation functions and output vectors are possible thus increasing the degrees of freedom in our models.

3.1 Stage 1, the Linear Nontransformed FCM Model

Consider now stage one. We will focus on the first concept in Table 3.1, C1, below because its examination already elucidates sufficiently well the application of our suggested methods.

Table 3.1.1 presents the descriptive statistics of these concepts at time=t+1 prior to their transformations, and these concept values will take below the priority over the transformed ones in our analysis because then we may exclude the effect of the parameter lambda in our studies. We notice that these possible concept values, which based on 10 000 random initial vectors given at time=t, have dissimilar distributions and generally these distributions are also dependent upon the number of their driver concepts. The latter fact is quite seldom taken into account even though it clearly affects the transformed values of the concepts. We will nevertheless preclude this discussion here.

Table 3.1.2 presents the corresponding transformed concept values when lambda=1. We notice that the dispersions are quite small, and thus these outcomes may often be undesirable in the concrete applications.

Figures 3.1.1 and 3.1.2 depict the possible values of C1 at time=t+1 both prior to and after the transformation with lambda=1 according to our random initial vectors at time=t. Since its nontransformed values are mainly negative, their transformed counterparts are centered below the value of 0.5. Figure 3.1.3 depicts how the nontransformed values of C1 are transformed when lambda=1.

Table 3.1.1. The descriptive statistics of the possible nontransformed concept values when 10 000 initial random vectors were used.

Concept	Range	Minimum	Maximum	Mean	Std. Deviation
C1	3.98	-3.07	.91	-1.1963	.54831
C2	4.07	-2.41	1.66	-.2896	.55202
C3	3.30	-2.07	1.22	-.3128	.52685
C4	2.73	-1.54	1.19	-.1377	.44012
C5	1.87	-.13	1.74	.7675	.29210
C6	3.49	-2.99	.49	-1.1466	.50610
C7	3.54	-2.37	1.17	-.5944	.52323
C8	2.67	-1.81	.87	-.4775	.41512
C9	4.08	-2.94	1.14	-.7806	.56001
C10	3.25	-.23	3.03	1.4479	.44432

Table 3.1.2. The descriptive statistics of the possible transformed concept values when 10 000 initial random vectors were used.

Concept	Range	Minimum	Maximum	Mean	Std. Deviation
C1	.67	.04	.71	.2454	.09780
C2	.76	.08	.84	.4329	.12732
C3	.66	.11	.77	.4270	.12210
C4	.59	.18	.77	.4672	.10521
C5	.38	.47	.85	.6797	.06272
C6	.57	.05	.62	.2524	.09240
C7	.68	.09	.76	.3638	.11511
C8	.56	.14	.70	.3873	.09534
C9	.71	.05	.76	.3254	.11670
C10	.51	.44	.95	.8006	.06995

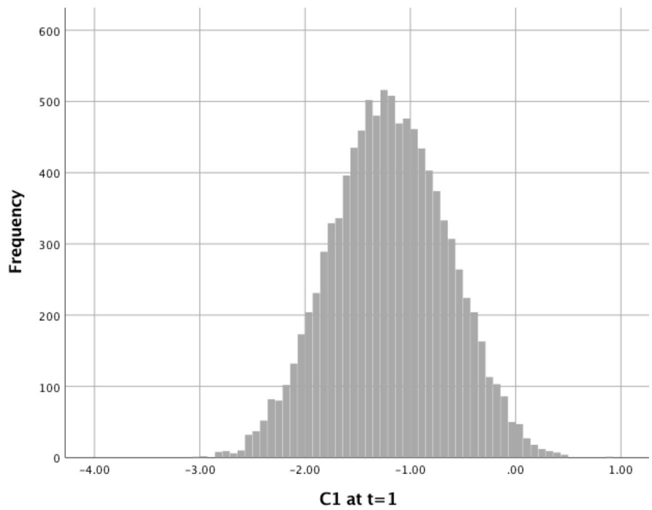

Fig. 3.1.1. The distribution of the possible concept values of C1 prior to the transformations.

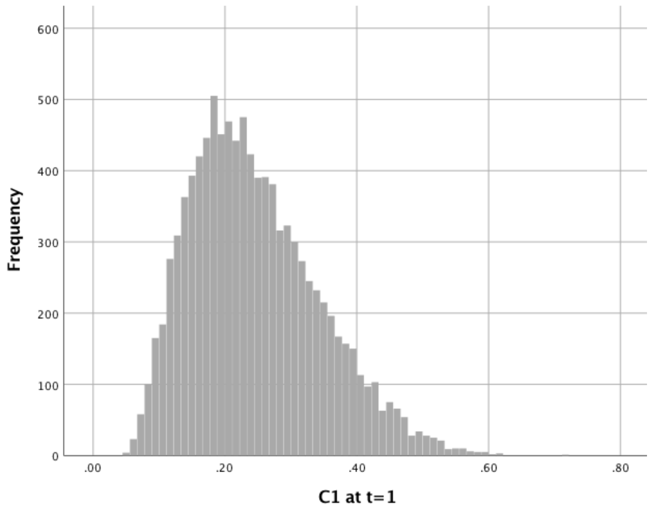

Fig. 3.1.2. The distribution of the possible concept values of C1 after the transformations.

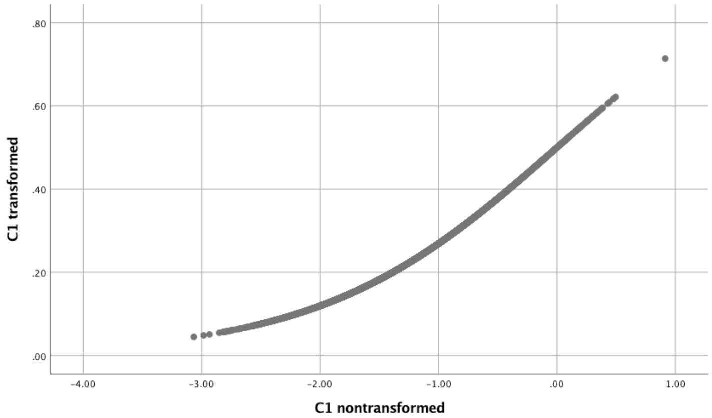

Fig. 3.1.3. The transformed vs. nontransformed values of concept C1 at time=1.

Since the possible nontransformed concept values are actually the outcomes of linear functions, we may apply linear regression analyses thru their origins in this context [7,16,17]. If we use, for example, the concept C1 as our response concept and the concepts C2 to C10 are its predictors, we may construct a linear regression model with our 10 000 data vectors. Then, the initial random vectors of the concepts

C2 to C10 at time=t constitute the predictor data and the concept values of C1 at time=t+1 represent the response values.

When using all predictors in this regression model, the estimates of the regression coefficients are naturally similar to the weights in Table 3.1. If we aim at removing the irrelevant or insignificant predictors for achieving a simpler model, we may apply stepwise regression models. These models were first constructed with SPSS for finding the significant predictors. In this manner, we may have stochastic justifications for removing the insignificant predictors, and thus the predictors C6 and C9 were removed from our final model. We notice that these predictors also have small weights in column C1 in Table 3.1, but now these removals were not performed on subjective grounds.

Our final regression model had the R-square value of 0.998 and its coefficient estimates are presented in Table 3.1.3. Hence, this model should yield output values which are virtually similar to the original values of C1 at time=t+1. We notice that the regression coefficients (B values) are not identical to the corresponding weights in matrix 3.1 because two predictors were removed. The t-tests indicate that the remaining predictors seem significant (their levels of significance are below 0.05). The beta values indicate that the concepts C10 and C4 seem to be the most relevant predictors (the highest absolute values), whereas C2 and C3 seem to play minor roles. The tolerances are quite low, and thus the undesirable multicollinearity exists to some extent. Figure 3.1.4 depicts the residuals, and they seem normally distributed around zero. Hence, now we have a plausible statistical model which yields nontransformed concept values to C1 by only using the significant predictors.

Table 3.1.3. The estimates of linear regression coefficients for concept C1.

Concept	Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
C2	.524	.002	.229	283.811	.000
C3	-.511	.002	-.224	-275.610	.000
C4	-.840	.002	-.368	-455.429	.000
C5	-.593	.002	-.260	-319.250	.000
C7	.678	.002	.300	366.492	.000
C8	-.698	.002	-.306	-380.801	.000
C10	-.959	.002	-.424	-520.389	.000

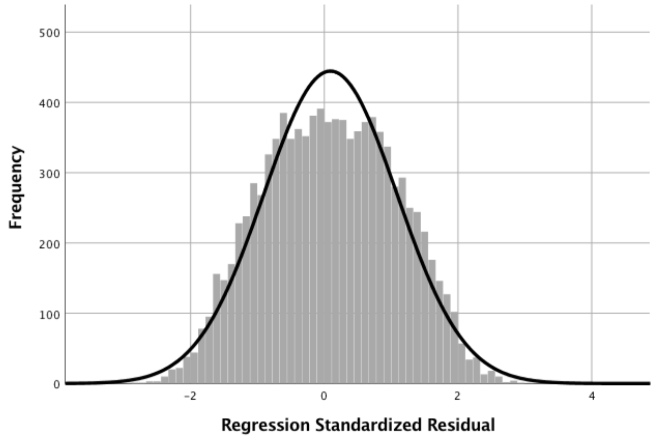

Fig. 3.1.4. The distribution of residuals within the linear regression analysis for concept C1.

With our regression coefficient estimates (which are the slopes of the predictors) we may now perform such what-if reasoning as

If the value of concept C2 at time=t increases 0.5 units, the nontransformed value of C1 at time=t+1 will increase B·0.5 = 0.524·0.5 = 0.262 units.

However, if collinearity exists, we must be more or less cautious with this type of reasoning.

In this manner, we may construct an appropriate regression model to each such target concept which has at least one driver. In Table 3.1.4 the weights of Table 3.1 are replaced with the new weights when the insignificant weights are removed according to our stepwise regression models. Hence, the small weights are removed on the statistical grounds. Table 3.1.4 also presents the R-square values of our corresponding regression models. Figure 3.1.5 depicts the corresponding simplified version of the original graph.

Figure 3.1.6 depicts our concept values when 10 iterations are calculated with our simplified FCM and lambda=1. The average deviation from the original FCM values is 0.013. In the human sciences such low level of errors will not arouse any critical problems.

Table 3.1.4. The simplified connection matrix and the R-squares of the regression models when the column concepts were used as the response variables.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
C1	0.00	0.50	0.97	-0.56	0.40	-0.52	-0.93	0.00	0.72	-0.57
C2	0.52	0.00	-0.33	0.00	0.00	0.00	-0.50	-0.81	0.00	0.69
C3	-0.51	0.90	0.00	0.73	0.42	-0.58	0.69	-0.30	-0.78	0.00
C4	-0.84	0.88	-0.93	0.01	0.00	0.55	-0.62	-0.24	0.81	0.00
C5	-0.59	-0.76	-0.14	-0.98	0.00	-0.78	-0.70	0.48	-0.54	0.00
C6	0.00	-0.53	0.00	0.00	-0.35	0.00	0.00	-0.71	0.00	0.49
C7	0.68	-0.44	0.00	0.67	0.00	-0.85	0.00	0.00	-0.92	0.70
C8	-0.70	0.00	-0.16	-0.13	0.68	0.00	0.00	0.00	-0.84	0.85
C9	0.00	-0.62	0.79	0.19	0.38	0.53	0.84	0.00	0.00	0.70
C10	-0.96	-0.48	-0.83	-0.21	0.00	-0.63	0.00	0.60	0.00	0.00
R-square	0.998	0.965	0.998	0.997	0.996	0.991	0.989	0.986	0.986	0.992

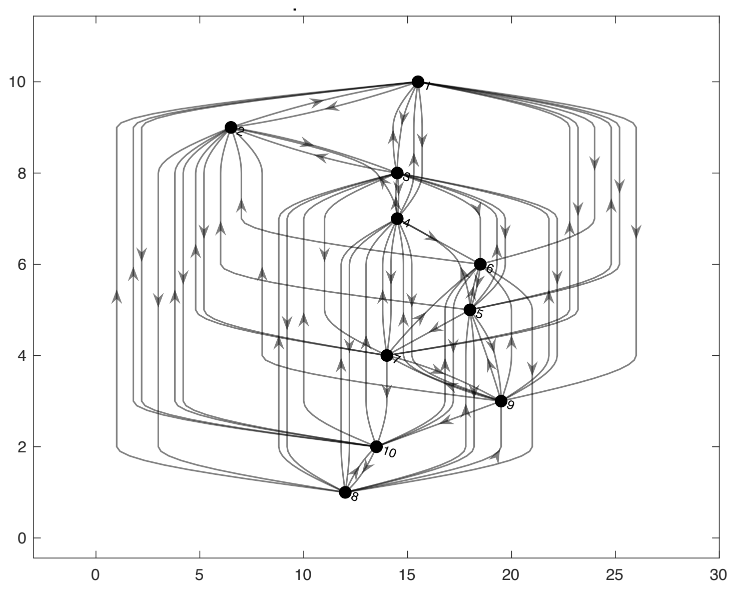

Fig. 3.1.5. The simplified graph based on the regression models.

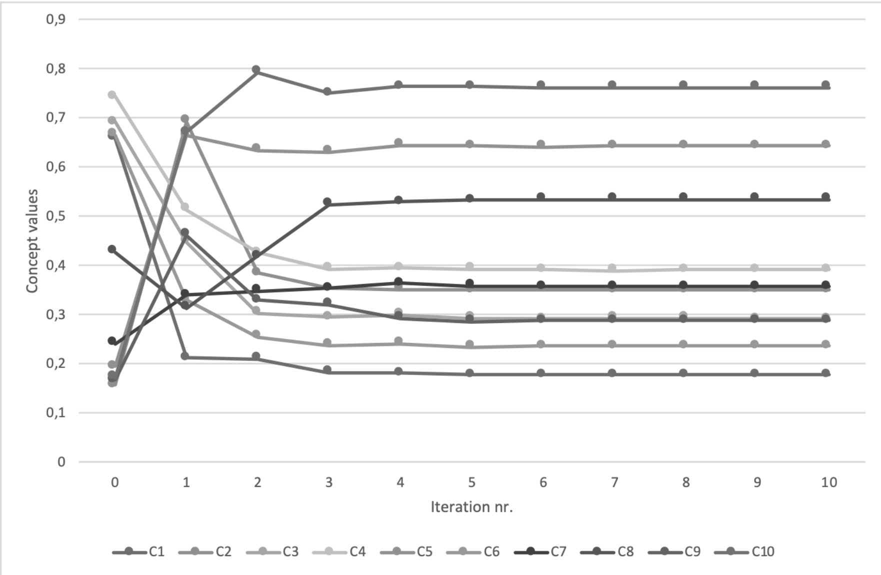

Fig. 3.1.6. Concept values in ten iterations in the simplified FCM when lambda=1 (iteration 0 denotes the original random values).

3.2 Stage 2, the Transformed FCM Model

At the first stage above, we always obtain unique FCM outputs when the weights in the connection matrix are fixed. At the second stage, in turn, our outputs also depend upon the transformation function, and thus various models may be constructed but the functions (2.2) and (2.3) seem usual. We only focus on transformation function (2.2), but our results may also be applied to many other functions.

At the outset, we are also interested in the descriptive statistics of the possible concept values when random initial concept values are used. Examples of these are provided in Tables 3.1.1 and 3.1.2. In the ordinary FCM studies we may also attempt to find the optimal parameter values for functions (2.2) or (2.3) according to the given criteria, even the individual parameters for each concept [3,11]. Hence, we first assign the coefficients to the linear functions for calculating the untransformed values and then we transform these values into a certain interval.

Within statistics, we may also examine directly with the regression models how to predict the transformed target values with the given drivers if the appropriate data is available. In this context, we may apply the nonlinear models, but this task is usually quite challenging, in particular when several predictors are involved, because the appropriate regression functions may often be unknown to us. The fuzzy reasoning systems may resolve this problem fluently, but they are not in the focus in this study because numerous such studies are available already [29-32].

One usual statistical method is logistic regression analysis in which case we may obtain the probabilities of acquiring certain concept values [16,17], and, in a sense, this method is analogous to FCM modelling. Consider thus our foregoing 10 000 transformed observations for concept C1 at time=t and lambda=1 in the function (2.2). If we, for example, will examine when these transformed values of C1 will be above its median at time=t+1, viz. above 0.23, we will first create a new dichotomous response variable, C1d, and C1d = 0 when C1 ≤ median, otherwise C1d = 1. After this, we may construct our model and we aim to estimate the probabilities of obtaining C1d=1.

Our final logistic regression analysis model with SPSS, which based on various stepwise analyses, yielded the similar significant predictors as above with Nagelkerke's pseudo R-square value of 0.953 (when 1 is the maximum value), and the regression coefficients are presented in Table 3.2.1.

We notice in Table 3.2.1 that the Wald tests assume our predictors to be significant (small levels of significance). The regression coefficients, B, indicate that if the values of the predictors C2 and C7 at time=t will increase, we will have higher probability of obtaining the value above the median for C1 at time=t+1 because these predictors have significant positive coefficients. The other predictors will decrease this probability if their values will increase at time=t due to their negative signs (the odds ratio values are useless here because the predictors are continuous variables).

Table 3.2.1. The estimates of logistic regression coefficients for concept C1.

Concept	B	S.E.	Wald	df	Sig.
C2	17.297	.725	568.587	1	.000
C3	-19.000	.782	590.253	1	.000
C4	-30.308	1.219	617.789	1	.000
C5	-22.539	.930	587.733	1	.000
C7	22.503	.912	609.072	1	.000
C8	-25.848	1.047	609.832	1	.000
C10	-34.959	1.393	629.706	1	.000
Constant	46.655	1.873	620.371	1	.000

Hence, the estimates on the individual probabilities of obtaining the response values of C1 above the median at time=t+1 are based on the logistic regression equation [16,17]

$$\text{Probability} = 1 / (1+exp(-1 \cdot Z)), \text{ when} \quad (3.2.1)$$
$$Z = 17.297 \cdot C2 - 19.0 \cdot C3 - 30.308 \cdot C4 - 22.539 \cdot C5 + 22.503 \cdot C7 - 25.848 \cdot C8 - 34.959 \cdot C10 + 46.655$$

and *exp* is the exponential function. Hence, if Z is approximately above -1.25, there is high probability of obtaining the value above the median for C1 (Fig. 3.2.1). We notice that this calculation is analogous to the calculation of the FCM values when lambda=1 with function (2.2) even though now the linear function is specified by applying the maximum likelihood method and thus the regression coefficients are different [16,17].

For example, given the predictor values at time=t,

C2=0.16, C3=0.69, C4=0.74, C5=0.20, C7=0.24, C8=0.43, C10=0.17

the estimated probability that C1 > 0.23 is 0.092 at time=t+1.

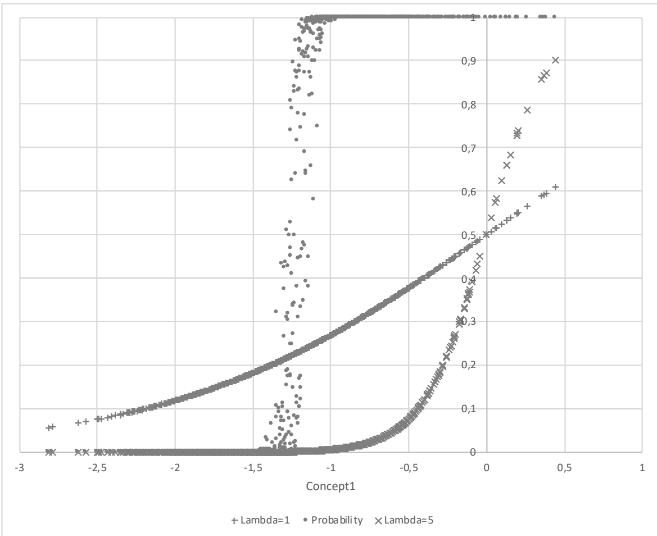

Fig. 3.2.1. Transformed values of concept C1 when lambda is 1 and 5 vs. their untransformed values. Also, the probabilities of concept values of C1 being above 0.23 according to the nontransformed values of C1.

Cluster analysis is also useful when we examine how the possible target values are obtained according to its drivers. In this context, we may apply both ordinary and fuzzy clustering techniques [16,17]. Table 3.2.2 provides an example on the cluster centers with Matlab's fuzzy subtractive clustering tool when four clusters are created with our data [5].

Table 3.2.2. Examples of cluster centers of the drivers and the corresponding target values.

Concepts at time=t	Cluster center 1	Cluster center 2	Cluster center 3	Cluster center 4
C2	0.500	0.460	0.613	0.710
C3	0.633	0.459	0.287	0.050
C4	0.527	0.456	0.280	0.405
C5	0.511	0.499	0.148	0.768
C7	0.472	0.564	0.586	0.293
C8	0.376	0.679	0.210	0.726
C10	0.671	0.283	0.314	0.702
C1 at time=t+1	0.208	0.263	0.454	0.184

Hence, for example, according to the first cluster center, we may reason that

If initially C2≈0.500 and C3≈0.633 and C4≈0.527 and C5≈0.511 and C7≈0.472 and C8≈0.376 and C10≈0.671, then we will obtain C1≈0.208.

In fact, these clusters and their centers may be used in fuzzy rule-based reasoning models if we aim at predicting the target values according to their drivers [2,8,29]. This widely-adopted approach is thus a fluent nonlinear method for predicting directly the transformed values from the initial concept vectors. Discriminant analysis, in turn, is an example on the corresponding statistical method [16,17].

We may also apply the multinomial regression models in this context in which case the categorical response variable may have more than two values [16-18]. An example of this is provided below.

The foregoing statistical methods provide both novel and supplementary information on FMCs which base on widely-used stochastic estimates and reasoning. In this manner, we may thus avoid certain subjective and ad hoc conclusions when we interpret or fine-tune our FCM models. We will apply these ideas to an empiric example below.

3.3 The Liquid Tank Model

Our suggested methods seem useful within the human sciences which is in the focus in this study. But we may also apply these methods to other fields and thus our example considers the well-known control application because fuzzy control models still play a central role in the fuzzy community.

Our empiric example considers the control application presented in [21,23]. In this model two valves, valve1 and valve 2, supply different liquids into the tank.

These liquids are mixed for a certain chemical reaction, and our goal is to maintain the desired liquid level (amount of liquid) and specific liquid gravity in the tank. The third valve, valve3, is used to drain liquid from the tank.

This FCM model applies the connection matrix presented in Table 3.3.1 which is given in [23], and now the preceding values of the target concepts are also used in simulations (self-loops, thus the diagonal values in the matrix are 1). The transformed concept values will use formula (2.2) with lambda=1 as in the original model. Table 3.3.2. presents the possible concept values prior to and after the transformations when 10 000 random initial concept vectors were used as above. Figures 3.3.1 and 3.3.2 depict the corresponding graph (without self-loops) and the concept values in ten iterations, respectively. The model will yield fixed-points to the concepts.

Table 3.3.1. The original weights of concepts.

	Liquid level	Valve1	Valve2	Valve3	Gravity
Liquid level	1	-0.207	-0.112	0.064	0.264
Valve1	0.298	1	0.061	0.069	0.067
Valve2	0.356	0.062	1	0.063	0.061
Valve3	-0.516	0.07	0.063	1	0.068
Gravity	0.064	0.468	0.06	0.268	1

Table 3.3.2. The descriptive statistics of the possible nontransformed and transformed concept values when 10 000 initial random vectors were used.

Concepts	Range	Minimum	Maximum	Mean	Std. Deviation
Nontransformed					
Liquid level	1.92	-.36	1.56	.5976	.34447
Valve1	1.63	-.11	1.52	.6952	.32617
Valve2	1.20	-.07	1.13	.5324	.29359
Valve3	1.46	.00	1.46	.7336	.30088
Gravity	1.46	.00	1.46	.7261	.30022
Transformed					
Liquid level	.42	.41	.83	.6413	.07757
Valve1	.35	.47	.82	.6633	.07170
Valve2	.27	.48	.76	.6275	.06770
Valve3	.31	.50	.81	.6722	.06545
Gravity	.31	.50	.81	.6706	.06541

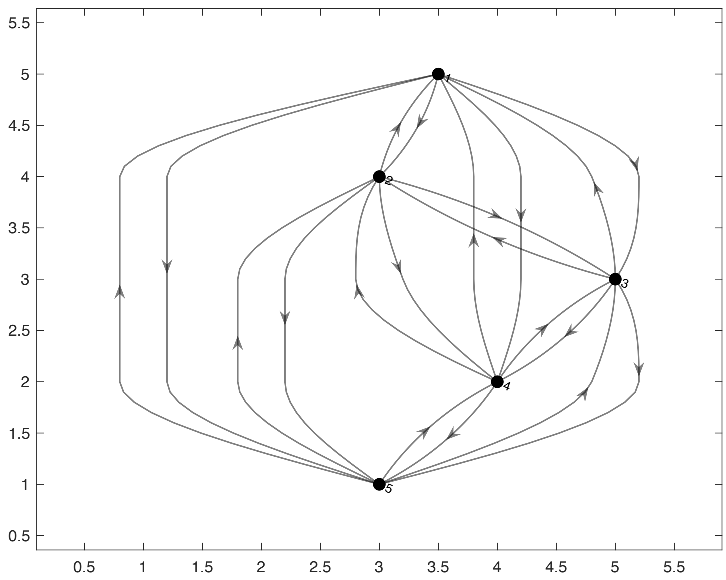

Fig. 3.3.1. Graph of the connection matrix in Table 3.3.1. (1=liquid level, 2=valve1, 3=valve2, 4=valve3, 5=gravity).

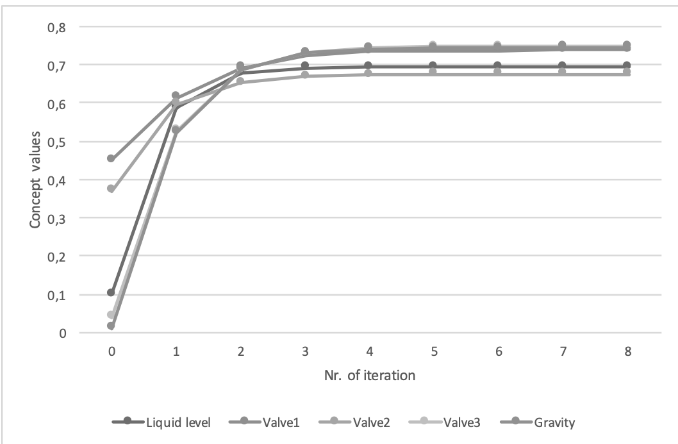

Fig. 3.3.2. Concept values in ten iterations in the original FCM when lambda=1 (iteration 0 denotes the original random values).

When stepwise linear regression analyses were applied to such models in which each concept at a time was the response variable and the other concepts acted as the possible predictors, we obtained the simplified connection matrix in Table 3.3.3, and this Table only contained the statistically significant drivers to each target concept. For example, Gravity was insignificant driver to Liquid level. We notice that our final regression models yielded very high R-square values. The average error in the concept values in ten iterations was 0.008 when our new FCM values were compared to the original ones. Figure 3.3.3 depicts our simplified FCM when the self-loops are not presented.

Table 3.3.3. The simplified connection matrix and the R-squares of the regression models.

	Liquid level	Valve1	Valve2	Valve3	Gravity
Liquid level	1	-0.166	-0.074	0	0.349
Valve1	0.313	1	0.098	0.106	0
Valve2	0.37	0	1	0	0
Valve3	-0.501	0	0	1	0
Gravity	0	0.507	0	0.304	1
R-square	0.999	0.998	0.997	0.998	0.996

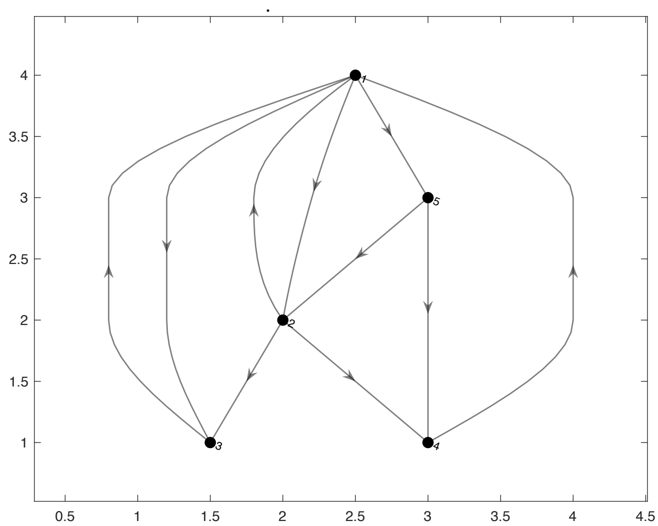

Fig. 3.3.3. Graph of the connection matrix in Table 3.3.3. (1=liquid level, 2=valve1, 3=valve2, 4=valve3, 5=gravity).

We will only focus on predicting the liquid levels in the tank. Hence, in [21,23] these rules were given,

- if Liquid_level < 0.68, the level is low.
- If 0.68 ≤ Liquid_level ≤ 0.70, the level is appropriate (the goal level).
- If Liquid_level > 0.70, the level is high.

If we, for example, will construct such logistic regression model with SPSS which may provide the probabilities of obtaining low liquid level according to our driver concept values, we should first create a new dichotomous target concept, Liquid_d,

- Liquid_d = 1, when the liquid level < 0.68,
- Liquid_d = 0, otherwise.

The statistics of this logistic regression model, whose Nagelkerke R-square value is 0.974, is presented in Table 3.3.4 (the insignificant predictor Gravity was removed). Hence, we notice that our final predictors seem significant according to the Wald tests and the signs of the linear regression coefficients, B, indicate (quite self-evidently) that, among others,

- If the initial liquid level increases, the risk to low liquid level is lower (its B value is negative).
- If the initial flow restriction increases in valve1, the risk to low liquid level is lower (its B value is negative).
- If the initial flow restriction increases in valve2, the risk to low liquid level is lower (its B value is negative).
- If the initial flow restriction increases in valve3, the risk to low liquid level is higher (its B value is positive).

Table 3.3.4. The Estimates of the logistic regression coefficients for the new liquid level.

Concepts	B	S.E.	Wald	df	Sig.
Initial liquid level	-94.964	5.195	334.192	1	.000
Valve1	-28.492	1.585	323.082	1	.000
Valve2	-34.163	1.892	325.998	1	.000
Valve3	48.440	2.672	328.687	1	.000
Constant	69.063	3.780	333.862	1	.000

As above, the probabilities of obtaining the low liquid level are calculated with the function,

Probability = 1 / (1+*exp*(-1·Z)), when (3.3.1)
Z = -94.964·Liquid_level – 28.492·Valve1 – 34.163·Valve2 + 48.440·Valve3 + 69.063

and *exp* is the exponential function (Fig. 3.3.4). For example, if the initial predictor values are

Liquid level=0.96, Valve1=.032, Valve2=0.53, Valve3=0.96 at time=t,

the probability of obtaining low liquid level is 0.029 at time=t+1.

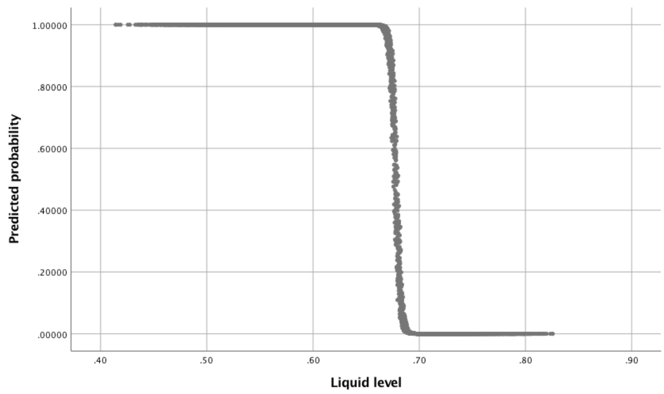

Fig. 3.3.4. The probabilities of obtaining the liquid level below the median vs. the transformed liquid levels.

In [18] the corresponding multinomial regression model with SPSS was also constructed in which case the Nagelkerke R-square was 0.979, and these results are in Table 3.3.5. In this case the goal level in [21,23],

0.68 ≤ liquid level ≤ 0.70,

was our reference class and the low and high levels were below and above this level, respectively. This model will actually yield two logistic regression models, and their probabilities base on the analyses of,

- low level compared to the goal level
- high level compared to the goal level

Table 3.3.5. The estimates of the multinomial logistic regression coefficients for the liquid tank model.

Liquid level		B	St. Error	Wald	df	Sig.
Goal vs. low	Intercept	79.136	14.739	28.827	1	0.000
	Liquid_level	-109.707	20.377	28.985	1	0.000
	Valve1	-33.785	6.341	28.392	1	0.000
	Valve2	-40.126	7.551	28.238	1	0.000
	Valve3	54.93	10.236	28.797	1	0.000
Goal vs. high	Intercept	-126.517	30.898	16.766	1	0.000
	Liquid_level	151.329	37.194	16.554	1	0.000
	Valve1	45.568	11.036	17.049	1	0.000
	Valve2	52.309	12.773	16.772	1	0.000
	Valve3	-79.126	19.38	16.67	1	0.000

Hence, the interpretation on our regression coefficients, B, in Table 3.3.5 is similar to that of logistic regression and thus we may reason on the stochastic grounds, among others,

- Goal vs. low: the increase in the initial liquid level will cause lower risk to achieve low liquid level (B value is negative).
- Goal vs. low: the increased flows in the valves 1 and 2 will cause lower risk to achieve low liquid level from the goal level (their B values are neg
- Goal vs. low: The increased flow in valve 3 will cause higher risk to achieve low liquid level from the goal level (B value is positive).
- Goal vs. high: the increase in the initial liquid level will cause higher risk to achieve high liquid level (B value is positive).
- Goal vs. high: The increased flows in valves 1 and 2 will cause higher risk to achieve high liquid level from the goal level (their B values are positive).
- Goal vs. high: The increased flow in valve 3 will cause lower risk to achieve high liquid level from the goal level (B value is negative).

These examples thus correspond well to the basic principles for controlling this system with its FCM.

The specific probabilities of the low and high liquid levels may be calculated with the linear regression coefficients in Table 3.3.5, and then with the logistic function, 1/(1+*exp*(-Z)), as above. Hence,

$$Z_{\text{low level}} = -109.707 \cdot \text{Liquid\_level} - 33.785 \cdot \text{Valve1} - 40.126 \cdot \text{Valve2} + 54.930 \cdot \text{Valve3} + 79.136 \quad (3.3.2)$$

$Z_{high\ level}$ = 151.329·Liquid_level + 45.568·Valve1 + 52.309·Valve2 -79.126·Valve3 - 126.517 (3.3.3)

For example, given the initial concept values at time=t,

Liquid_level=0.45, Valve1=0.03, Valve2=0.89, Valve3=0.06

the probabilities of obtaining the low, goal and high liquid levels at time=t+1 are 0.03, 0.97 and 0.00, respectively.

If we focus on those initial vector values which will lead to the goal values of the liquid level, Table 3.3.6 presents the descriptives of these values, and Table 3.3.7 presents examples of such typical initial vectors when the subtractive clustering was applied. We notice in Table 3.3.6 that the liquid level should be at least 0.19 for achieving its goal level in the subsequent FCM iterations, whereas the other concepts may vary more freely.

Table 3.3.6. The descriptives of the initial concept values that lead to the goal values of the liquid level.

Concept	Minimum	Maximum	Mean	Std. Deviation
Liquid level	.19	1.00	.6748	.18255
Valve 1	.00	1.00	.5202	.28840
Valve 2	.00	1.00	.5290	.28294
Valve 3	.00	1.00	.4854	.28565
Gravity	.01	1.00	.4988	.29010

Table 3.3.7. Examples of cluster centers of the initial concept values that lead to the goal values of liquid level.

Concepts	Center 1	Center 2	Center 3	Center 4	Center 5	Center 6	Center 7	Center 8	Center 9
Liquid level	0.723	0.447	0.789	0.556	0.873	0.669	0.699	0.765	0.378
Valve 1	0.504	0.550	0.183	0.812	0.867	0.931	0.169	0.060	0.871
Valve 2	0.427	0.766	0.706	0.266	0.280	0.755	0.142	0.787	0.624
Valve 3	0.553	0.233	0.756	0.191	0.908	0.820	0.136	0.487	0.231
Gravity	0.454	0.463	0.791	0.216	0.630	0.227	0.758	0.101	0.963

Thanks for our stochastic approach, we may again avoid better subjective and ad hoc decisions when we interpret our connection matrix weights, simplify our original models and forecast our FCM concept values in the simulations.

4 Conclusions

Fuzzy systems have proven to be applicable in various computer model constructions, and numerous scientific articles are already available about these models. The fuzzy cognitive maps were considered above, and this field of fuzziness has also been studied quite much already. However, many of the studies on these maps still seem to pivot on the methods of the engineering sciences and neural networks, and thus they do not necessarily reveal sufficiently their model performance. Hence, we expect certain methods which may supplement or clarify our research outcomes and even rely less on subjective or ad hoc reasoning.

Our approach above considered the fuzzy cognitive maps from the standpoint of the quantitative human sciences, and in this context, we applied more wide-ranging methods. In particular, statistical methods were also applied because they play a central role in this field. Thanks for these methods, we may avoid subjectivity better and simultaneously acquire further information on our models. We only focused on the numerical fuzzy cognitive maps, and especially on their a priori models, because these models will provide the basis for their other studies. The statistical analysis of a posteriori models with the history data and their linguistic versions will be an interesting objective of the future examinations.

The possible values of the fuzzy cognitive maps were studied by using the random initial concept vectors and with a random connection matrix. In this manner, we were able to estimate the performance of these maps, and due to our randomized approach, the generalization capability of our outcomes seems better. Various regression models were also applied for providing a stochastic basis for this performance and also reducing subjectivity. Our results indicated that they corresponded well to those of obtained by the prevailing cognitive map methods. Furthermore, our methods provided supplementary and objective information on these models as well as enhanced their fine-tuning. Naturally, we may also apply other statistical methods, and these are possible objectives of the future studies.

Our approach still awaits further justifications which will base on the performance of the concrete applications, and thus future studies are expected in this problem area.

Acknowledgements

I express my thanks to the distinguished Editors for having this opportunity to be one of the contributors of this book. This article is dedicated to the memory of my mentor and friend, the great Professor Lotfi Zadeh.

References

1. R. Axelrod, Structure of Decision, The Cognitive Maps of Political Elites. Princeton University Press, Princeton, 1976.
2. H. Bandemer and W. Näther, Fuzzy Data Analysis. Kluwer, Dordrecht, 1992.
3. M. Buruzs, M. Hatwágner and L. T. Kóczy, “Expert-Based Method of Integrated Waste Management Systems for Developing Fuzzy Cognitive Map”, in Q. Zhu and A. Azar (Eds.), Complex System Modelling and Control Through Intelligent Soft Computations. Studies in Fuzziness and Soft Computing, Vol. 319, pp. 111-137, 2015.
4. J. P. Carvalho and J. Tome, “Rule Based Fuzzy Cognitive Maps in Socio-Economic Systems”, Proceedings of the IFSA Congress, pp. 1821-1826, Lisbon, 2009.
5. S. Chiu, “Fuzzy Model Identification Based on Cluster Estimation”, Journal of Intelligent and Fuzzy Systems 2, 267-278, 1994.
6. V. Dimitrov and B. Hodge, Social Fuzziology – Study of Fuzziness of Social Complexity. Physica Verlag, Heidelberg, 2002.
7. D. Freedman, “Statistical models: Theory and practice”, Cambridge University Press, Cambridge, 2005.
8. Fuzzy Logic User’s Guide 2018a, Mathworks, 2018, www.mathworks.com/help/pdf_doc/fuzzy/fuzzy.pdf.
9. M. Glykas, (Ed.), Fuzzy Cognitive Maps, Springer, Heidelberg, 2010.
10. P. Grzegorzewski, O. Hryniewicz and M. Gil, Soft Methods in Probability, Statistics and Data Analysis. Physica Verlag, Heidelberg, 2002.
11. M. Hatwagner, V. Niskanen and L. Koczy, “Behavioral Analysis of Fuzzy Cognitive Map Models by Simulation”, Proceedings of the IFSA ’17 Congress, Otsu, Japan, https://ieeexplore.ieee.org/abstract/document/8023345/
12. S. Kim, C. Lee, “Fuzzy Implications of Fuzzy Cognitive Map with Emphasis on Fuzzy Causal Relationship and Fuzzy Partially Causal Relationship”, Fuzzy Sets and Systems 97 (3) pp. 303-3013, 1998.
13. B. Kosko, Fuzzy Engineering, Prentice Hall, Upper Saddle River, New Jersey, 1997.
14. K. C. Lee, W. J. Lee, O. B. Kwon, J. H. Han and P. I. Yu, “Strategic planning simulation based on fuzzy cognitive map knowledge and differential game”, Simulation, Vol. 71 (5), pp. 316–327, 1998.
15. R. Kruse and K. Meyer, Statistics with Vague Data. Reidel, Dordrecht, 1987.
16. J. Metsämuuronen, Essentials in Research Methods in Human Sciences, Multivariate Analysis, Sage, London, 2017.
17. J. Metsämuuronen, Essentials in Research Methods in Human Sciences, Advanced Analysis, Sage, London, 2017.
18. V. A. Niskanen, “Application of Logistic Regression Analysis to Fuzzy Cognitive Maps”, in L. Zadeh and R. Aliev (Eds.): Fuzzy Logic Theory and Applications, World Scientific, 2018, forthcoming.
19. V. A. Niskanen, “Concept map approach to approximate reasoning with fuzzy extended logic”, in M. Fedrizzi, M. Collan and J. Kacprzyk (Eds): Fuzzy Technology: Present Applications and Future Technology, Studies in Fuzziness and Soft Computing, Vol. 335, pp. 47-70, Springer, Heidelberg, 2016.
20. J. Novak, Learning, Creating, and Using Knowledge: Concept Maps as Facilitative Tools in Schools and Corporations, Lawrence Erlbaum Associates, Inc., New Jersey, 1998.

21. E. Papageorgiou, E. Stylios and P. Groumpos, “Fuzzy cognitive map learning based on nonlinear Hebbian rule”, in T. Gedeon and L. Fung (Eds.): AI 2003. LNCS (LNAI), Vol. 2903, pp. 256-268, Springer, 2003.
22. W. Pedrycz, A. Jastrzebska and W. Homenda, “Design of Fuzzy Cognitive Maps for Modeling Time Series”, IEEE Transactions of Fuzzy Systems, Vol. 24 (1), pp. 120-130, 2016.
23. W. Stach, L. Kurgan and W. Pedrycz, “Expert-based and computational methods for developing fuzzy cognitive maps”, in M. Glykas (Ed.): Fuzzy cognitive maps, pp. 24-41, Springer, 2010.
24. W. Stach, L. A. Kurgan and W. Pedrycz, "Numerical and Linguistic Prediction of Time Series With the Use of Fuzzy Cognitive Maps," IEEE Transactions on Fuzzy Systems, vol. 16, 2008.
25. W. Stach, L. Kurgan, W. Pedrycz and M. Reformat, "Genetic learning of fuzzy cognitive maps," Fuzzy Sets and Systems, vol. 153, pp. 371-401, 2005.
26. W. Stach, L. Kurgan and W. Pedrycz, "A Survey of Fuzzy Cognitive Map Learning Methods," Issues in Soft Computing: Theory and Applications, pp. 71-84, 2005.
27. C. Stylios and P. Groumpos, “Modeling Complex Systems Using Fuzzy Cognitive Maps”, IEEE Transactions on Systems, Man and Cybernetics Part A 34 (1), pp. 155-162, 2004.
28. F. Wenstøp, “Quantitative Analysis with Linguistic Values”, Fuzzy Sets and Systems 4, pp. 99-115, 1980.
29. L. Zadeh, “Fuzzy Logic = Computing with Words”, IEEE Transactions on Fuzzy Systems, 2, pp. 103-111, 1996.
30. L. Zadeh, “Toward a Theory of Fuzzy Information Granulation and Its Centrality in Human Reasoning and Fuzzy Logic”, Fuzzy Sets and Systems 90/2, pp. 111-127, 1997.
31. L. Zadeh, “From Computing with Numbers to Computing with Words - From Manipulation of Measurements to Manipulation of Perceptions”, IEEE Transactions on Circuits and Systems 45, pp.105-119, 1999.
32. L. Zadeh, “Toward a Perception-Based Theory of Probabilistic Reasoning with Imprecise Probabilities”, Journal of Statistical Planning and Inference 105/2, pp.233-264, 2002.