An Application of Fuzzy C-Means, Fuzzy Cognitive Maps, and Fuzzy Rules to Forecasting First Arrival Date of Avian Spring Migrants

Dariusz Czerwinski
Department of Computer Science
Lublin University of Technology
Lublin, Poland
d.czerwinski@pollub.pl

Adam Kiersztyn
Department of Computer Science
Lublin University of Technology
Lublin, Poland
adam.kiersztyn.pl@gmail.com

Rafał Łopucki
Centre for Interdisciplinary Research
The John Paul II Catholic University of Lublin
Lublin, Poland
lopucki@kul.pl

Paweł Karczmarek
Department of Computer Science
Lublin University of Technology
Lublin, Poland
pawel.karczmarek@gmail.com

Ignacy Kitowski
The State School of Higher Education in Chełm
Chełm, Poland
ignacyk@autograf.pl

Adam Zbyryt
The Polish Society for Bird Protection (PTOP)
Białystok, Poland
adam.zbyryt@wp.pl

***Abstract*—In this study, we propose an approach based on the advanced fuzzy techniques such as Fuzzy C-Means and Fuzzy Cognitive Maps to cluster the birds species, based on the information of first arrival date, into more coherent and uniform groups. The birds are very suitable subject for modelling the climate changes. Very popular indicator to forecast bird migration dynamic is the first arrival date. In many reported studies, this indicator is shown as very useful. However, there is still a lack of precise methods grouping the birds into the classes in satisfying manner producing detailed information about species and the relations between them. As evidenced in the experimental series section, the proposed approach enables the researchers and practitioners working with that important area of ecology to observe subtle dependencies between various bird species. Moreover, this work sheds the light on the novel application of both Fuzzy C-Means and Fuzzy Cognitive Maps as the efficient tools to analyse the ecological data collected in changing climatic environment.**

***Index Terms*—fuzzy clustering, bird migration, arrival date, forecasting**

I. Introduction

Studies considering phenological changes in response to climate change are numerous and concern of many groups of biota [1]. One of the most-used scientific research models in the context of climate changes are birds. It is assumed that climate change will affect the shape and dynamics of periodic processes of bird populations, such as migration and breeding. One of the most popular indicators used to study bird migration is first arrival date (FAD) to the breeding areas [2]–[4]. Many previous comparative studies from various regions of the world, including Europe, indicate that FAD has changed significantly in recent decades, but the rate and significance of these changes is generally higher in short distance migrants than in long distance ones [2], [4]–[6]. As a result, phenological changes of FAD could cause mismatching in food chains, thus inducing important perturbations in ecosystem functioning, for example, by breaking the relationship between parasitic birds and their avian hosts, particularly in relation to short distance migrants [7], [8].

II. Description of the Periods of Bird Migration

Despite many studies showing the usefulness of FAD in recording the consequences of climate change, there are also critical papers showing that first appearance dates behave as a very inaccurate and biased estimator regarding any phenological data set [7]. For example, a frequently raised argument is that population size could impact on the detection of bird arrival time as there is a higher probability of observing earlier arrival when the population size is greater and the song activity of birds is increased, as occurs with a larger population [9]. Therefore, new statistical tools are constantly being sought to reduce the limitations of the methods used to collect data and the indicators calculated on their basis. The use of fuzzy numbers can be one such solution. Therefore, the aim of the work was the introduction of fuzzy classes dividing bird species into smaller, more uniform and coherent groups. Thanks to this division, it will be possible to use close relationships between species within individual classes and to develop models forecasting arrival dates of birds based on data from representatives of each class.

III. Fuzzy C-Means (FCMeans) Clustering

In this section, we recall a well-known classic method of data clustering as proposed by [10] settled in the nomenclature of a spatio-temporal time series. Let us assume that the data is a set of the form $x_1, x_2, ..., x_n$, where $x_i = x_{i1}, x_{i2}, \ldots, x_{im}, i = 1, 2, \ldots, n, m \geq 1$. Here, the elements of the vector x_i can be identified as the changing values in

978-1-7281-6932-3/20/$31.00 ©2020 IEEE

the discrete time intervals marked by m. This kind of data can obviously be an input to the FCM algorithm. Its goal is to find c clusters, also called information granules, namely $v_1, v_2, ..., v_c$ and a fuzzy partition matrix $U = [u_{pq}], p = 1, 2, \ldots, c, q = 1, 2, ..., n$, with values in-between 0 and 1. Moreover, $\sum_{i=1}^{c} u_{iq} = 1$ for each q and $0 < \sum_{i=1}^{n} u_{pi} < n$ for each p. The matrix U elements u_{pq} are obtained as a result of the following optimization task: to minimize the following function sum

$$\sum_{i=1}^{c}\sum_{q=1}^{n} u_{iq}^{M} d^2(v_i, x_q) \tag{1}$$

with an arbitrary $M > 1$ sought as a fuzzification coefficient. In most of the applications it is assumed $M = 2$, d is a distance measure, usually a Euclidean one because of the easiness of the next calculations. The result of the FCMeans is described as the following two formulas:

$$v_i = \sum_{q=1}^{n} u_{iq}^{M} x_q / \sum_{q=1}^{n} u_{iq}^{M} \tag{2}$$

for the center prototypes, and

$$u_{iq} = 1 / \sum_{i=1}^{c} \left(\frac{d(v_i, x_q)}{d(v_j, x_q)} \right)^{\frac{2}{M-1}} \tag{3}$$

for the participation matrix entities. These two formulas are used in a series of iteration to get the optimized result. The algorithm is stopped after a certain number of iterations or if the distance between two successive partition matrices is small enough.

In our case, x_is are the vectors containing the series of bird incoming times each of the m years. Therefore, the clustering is seen as the initial division onto the group of birds in relation to the time of migration.

IV. Fuzzy Cognitive Maps

Fuzzy cognitive maps (FCM) are a very effective tool for modelling complex structures and data. In general, a fuzzy cognitive map [11], [12] is a directed graph in which nodes are factors such as (for example): operating principles, events, and branches reflect causal relationships. In FCM there is interconnection between each two concepts C_i and C_j. Two concepts are connected with the directed edge w_{ij}, which indicates the strength of relationships between concepts. There are three possible cases for the weight value:

- $w_{ij} > 0$ indicates a positive causality between two concepts, and means that an increase or decrease in the value of concept C_i and leads to the increase or decrease of the value of concept C_j,
- $w_{ij} < 0$ indicates a negative causality between concept two concepts, meaning that an increase in the value of concept C_i leads to a decrease of the value of concept C_j or a decrease of one value causes an increase in the other,
- $w_{ij} = 0$ indicates no relationship between concepts.

Each concept in a fuzzy cognitive map has the value A_i which express a concept strength. Each subsequent value of the causal state is calculated by Kosko [11] using previous state and weight matrix. The relationship is given by the formula:

$$A_i(m+1) = f\left(\sum_{j=1, j\neq 1}^{N} (w_{ji} \times A_j(m))\right) \tag{4}$$

The value of the particular concept strength is calculated during each step of simulation taking into consideration the influence of other concepts in weight matrix.

In this formula f is the threshold function which can be: bivalent, trivalent, sigmoid and hyperbolic tangent [13]. In our case the threshold function was sigmoid and expressed below:

$$f(x) = \frac{1}{1 + e^{-\lambda \times x}} \tag{5}$$

Parameter λ in our case was set to 1.

V. Fuzzy Rule Classifier

Fuzzy classifiers are a very broad class of algorithms that share a common denominator: we call the fuzzy classifier ”Any classifier that uses fuzzy sets or fuzzy logic in the course of its training or operation” [14].

The simplest fuzzy rule-based classifier is a fuzzy if-then system, analogous to that used in fuzzy control. The fuzzy classifier can be specified using a sequence of rules. Consider the example in which the coordinates of points $(x, y) \in R^2$ in two-dimensional space will be classified. In this case, we can specify sample rules

IF *x is small* ***AND*** *y is medium* ***THEN*** *class number 1*
IF *x is small* ***AND*** *y is large* ***THEN*** *class number 2*
IF *x is medium* ***AND*** *y is small* ***THEN*** *class number 3*
IF *x is medium* ***AND*** *y is large* ***THEN*** *class number 4*

The x, y coordinates are numeric here, but the classification rules use linguistic descriptors. It should be noted that if we allow the use of N descriptors and the number of features (coordinates) of the point is k, then using a logical conjunction and N^k different results can be obtained.

If the fuzzy classifier includes all such rules, then it turns into a simple lookup table. However, unlike lookup tables, fuzzy classifiers can provide results for combinations of language values that are not included as one of the rules. Each language value is represented by a membership function Fig. 1.

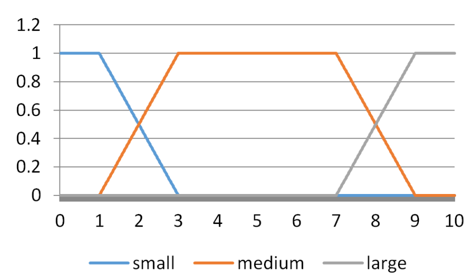

Fig. 1. Examples of membership functions for the descriptors considered.

Consequently, the classification rules described above can be expressed in the notation of membership function

$$\tau_1(x,y) = \mu^{(1)}_{small}(x)\, AND\, \mu^{(2)}_{medium}(y) \quad (6)$$

$$\tau_2(x,y) = \mu^{(1)}_{small}(x)\, AND\, \mu^{(2)}_{large}(y) \quad (7)$$

$$\tau_3(x,y) = \mu^{(1)}_{medium}(x)\, AND\, \mu^{(2)}_{small}(y) \quad (8)$$

$$\tau_4(x,y) = \mu^{(1)}_{medium}(x)\, AND\, \mu^{(2)}_{large}(y) \quad (9)$$

Superscripts are necessary because the range of individual descriptors for each coordinate may be different. The AND operation is typically implemented as minimum but any other t-norm may be used. The rule "votes" for the class of the consequent part. The weight of this vote is $\tau(x, y)$. To find the output of the classifier, the votes of all rules are aggregated. An algorithm [15] implemented in the KNIME environment was used in the experiments.

VI. Experimental results

A. Data description

The dates and first arrival date (FAD) were examined. The considerations are limited to birds of close migration, presented together with their Latin names in Table I. The data were recorded in the years 1996-2016 by the large group of ornithologists belonging to the Polish Society for Bird Protection (PTOP).

TABLE I
NUMBERS, ENGLISH NAMES AND LATIN NAMES OF THE BIRDS

Bird number	English name	Latin name
1	Eurasian bittern	*Botaurus stellaris*
2	Marsh harrier	*Circus aeruginosus*
3	Lapwing	*Vanellus vanellus*
4	Wood pigeon	*Columba palumbus*
5	Eurasian blackcap	*Sylvia atricapilla*
6	Black redstart	*Phoenicurus ochruros*
7	Spotted crake	*Porzana porzana*
8	Common redshank	*Tringa totanus*
9	Common snipe	*Gallinago gallinago*
10	European serin	*Serinus serinus*
11	Woodlark	*Lullula arborea*
12	Coot	*Fulica atra*
13	Great crested grebe	*Podiceps cristatus*
14	Red-necked grebe	*Podiceps grisegena*
15	Common chiffchaff	*Phylloscopus collybita*
16	White wagtail	*Motacilla alba*
17	Dunnock	*Prunella modularis*
18	Reed bunting	*Schoeniclus schoeniclus*
19	Eurasian penduline-tit	*Remiz pendulinus*
20	Eurasian skylark	*Alauda arvensis*
21	Eurasian woodcock	*Scolopax rusticola*
22	Black-headed gull	*Chroicocephalus ridibundus*
23	Song thrush	*Turdus philomelos*
24	Meadow pipit	*Anthus pratensis*
25	Western water rail	*Rallus aquaticus*
26	Common crane	*Grus grus*

B. Fuzzy C-Means

In this section we describe the results of clustering using the Fuzzy C-Means method to obtain three groups of birds divided according to arrival times. The cluster centres are presented in Fig. 2. We observe that the clusters are mostly relatively far to each other. However, in some years it may be relatively difficult to find three different groups of birds. However, these are sporadic cases. Moreover, it cannot be observed that one of the groups (1-3) of migrating birds tends more to other group. Therefore, the division into three groups is relatively stable and gives clear results. Table II presents the results of clustering of the above listed birds. One can observe that in general each bird species belongs to mainly one group of birds. Only one species, namely #16, is somewhere in-between of two groups (1 and 2). However, the trend clearly shows that it is more in the first group.

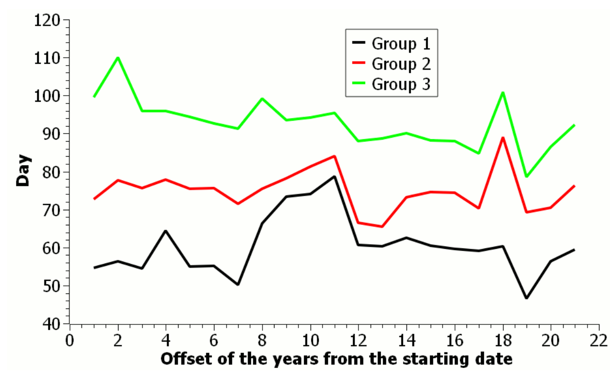

Fig. 2. Cluster centres of groups of migrating birds.

TABLE II
MEMBERSHIPS OF PARTICULAR BIRD SPECIES IN MIGRATION GROUPS

Bird number	Group 1	Group 2	Group 3
1	0.09	0.34	0.57
2	0.12	0.75	0.14
3	0.89	0.09	0.03
4	0.22	0.69	0.09
5	0.06	0.13	0.81
6	0.12	0.68	0.21
7	0.05	0.11	0.84
8	0.05	0.89	0.06
9	0.06	0.89	0.05
10	0.04	0.14	0.82
11	0.30	0.63	0.07
12	0.22	0.67	0.11
13	0.12	0.67	0.21
14	0.03	0.10	0.87
15	0.03	0.10	0.87
16	0.48	0.42	0.09
17	0.04	0.16	0.80
18	0.87	0.11	0.03
19	0.05	0.13	0.82
20	0.79	0.15	0.06
21	0.16	0.67	0.17
22	0.87	0.10	0.03
23	0.37	0.52	0.11
24	0.26	0.66	0.07
25	0.07	0.19	0.74
26	0.94	0.05	0.01

C. Fuzzy Cognitive Maps

An important input for a fuzzy cognitive map is the weight matrix, which represents the relation between the elements (concepts). In this research the weight matrix was calculated as a modified matrix of correlation coefficients between a set of 26 birds. This was done with an algorithm implemented in the *rcorr* function. function of the R language. The matrix of Pearson correlation coefficients and asymptotic P-values were obtained. Later, with the use of code written by the authors, the matrix values were modified in the following way: diagonal has been set to 0 to remove the self loops in the fuzzy cognitive maps, and only the values of correlation coefficients for which P-values smaller than 0.05 were taken into consideration.

Next we used the *fcm* package in the R environment [16] to calculate the strength of the elements. The activation vector for each bird was set to 1 and the number of iteration was 25. For better explanation the results of calculations were presented in the form of energy Fruchterman Reingold graphs shown in Fig. 3 to Fig. 5. In this kind of graph the energy of the whole system is minimised and the nodes are moving until the system reaches its equilibrium state. The diameter of the circle is proportional to the concept strength and the thickness of the line is proportional to the relation between elements. The colours of the lines show positive (black) or negative (red) relation between the graph elements.

Additionally we checked how the clustering with Fuzzy C-Means correlates with fuzzy cognitive maps. We performed the calculations for the 3, 4 and 5 clusters of birds emerging with the fuzzy C-Means. These clusters were shown with different colours i.e: yellow (cluster 1), green (cluster 2), blue (cluster 3), brown (cluster 4), teal blue (cluster 5). Membership of the bird in the cluster was assumed according to the maximum value obtained from the fuzzy C-Means calculations (for example for 3 clusters, bird No 6 belong to the second cluster as shown in Table II).

The results of analysis were shown in Figures 3 – 5. It can be noticed that, due to the fuzzy C-Means analysis, birds change the membership in the clusters. However the fuzzy cognitive map analysis shows that birds stay in the same graph energy region.

There are some birds which stay always in the same cluster and in the same energy region, e.g.: common crane, black-headed gull, Eurasian skylark, lapwing and reed bunting. This can be seen in Figures 3 – 5: 3 – marked in yellow, 4 and 5 – marked in green.

Analysing the fuzzy cognitive map results we can say that the most birds belong to the three cluster centres. Comparing FCM with Fuzzy C-Means gives us more specific information on membership of the birds in the same classes obtained from both analyses. The results are presented in Table III.

D. Fuzzy Rules

Fuzzy classifiers for the analysed data were used in two classification problems. On the one hand, the effectiveness of classification of bird species by time of arrival was examined.

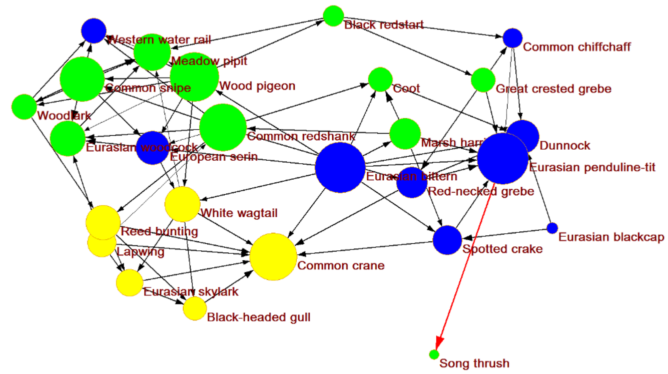

Fig. 3. Fuzzy cognitive map representation as a Fruchterman Reingold graph for 3 clusters emerging with the use of fuzzy C-Means.

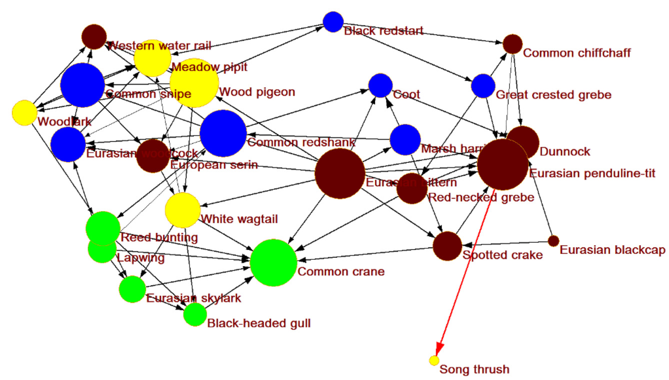

Fig. 4. Fuzzy cognitive map representation as a Fruchterman Reingold graph for 4 clusters emerging with the use of fuzzy C-Means.

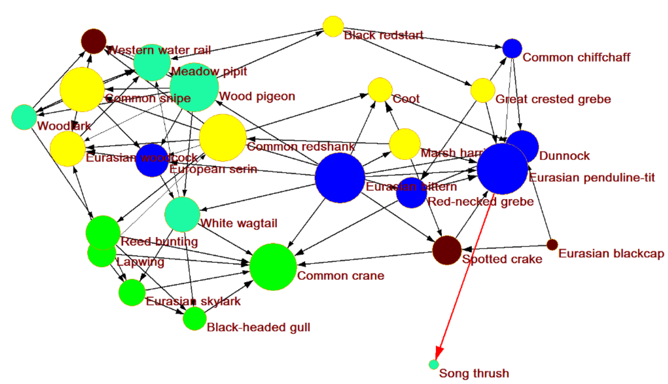

Fig. 5. Fuzzy cognitive map representation as a Fruchterman Reingold graph for 5 clusters emerging with the use of fuzzy C-Means.

TABLE III
MEMBERSHIP OF PARTICULAR BIRD SPECIES IN CLUSTER GROUPS BASED ON THE FUZZY COGNITIVE MAPS AND FUZZY C-MEANS

Cluster number	Cluster members
1	3, 18, 20, 22, 26 Lapwing, Reed bunting, Eurasian skylark, Black-headed gull , Common crane
2	4, 11, 24 Wood pigeon, Woodlark, Meadow pipit
3	1, 14, 17, 19 Eurasian bittern, Red-necked grebe, Dunnock, Eurasian penduline-tit

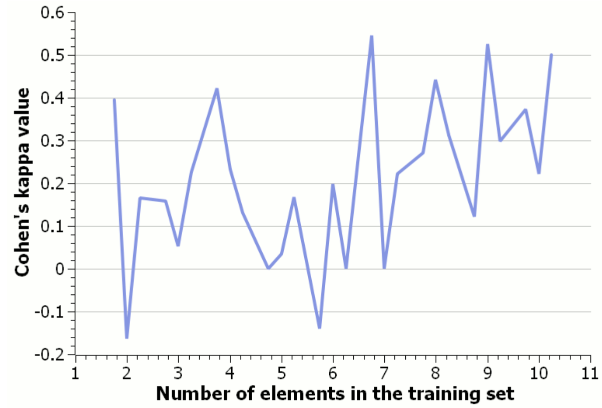

Fig. 7. Cohen's kappa coefficient values for subsequent random elements of the training set.

On the other hand, the effectiveness of classification of individual years was examined, due to the arrival time of the species studied, for warm years (birds arrive quickly) and for cool years (late arrivals of birds).

In the case of classifying years into warm and cold ones, based on the terms of bird arrivals, the results are not satisfactory. The percentage of proper classification of objects (individual years) for different numbers of elements in the training set is presented in Fig. 6. For each of the cases considered, 3 independent tests of random selection of elements for the learning set were carried out.

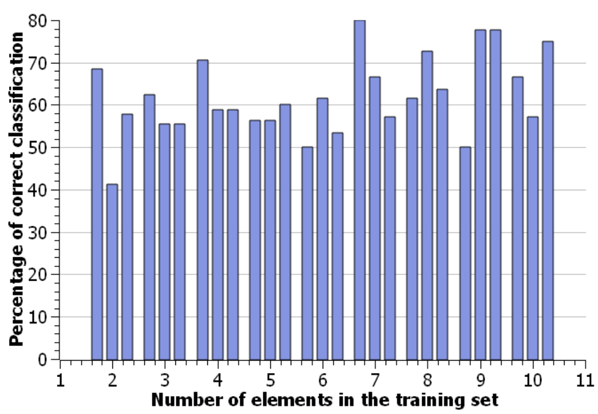

Fig. 6. Efficiency of year classification based on bird arrival times.

The results obtained are characterised by great diversity within the same number of learning sets, which suggests that the right selection of elements is of decisive importance here. The selection of elements for the learning set is still an open problem and will be implemented as part of future work. The values of the Cohen kappa coefficient for individual tests are presented in the Fig. 7.

As one can see, in some cases a negative coefficient suggests that random allocation to individual categories would be better. In the case of classification of birds into species characterised by early and late arrival, we receive much better results of classification (see Fig. 8). In this case, also the selection of elements for the learning set is important and will be tested as part of future work.

Taking into consideration classification efficiency for Fuzzy C-Means and Fuzzy Cognitive Maps shown in Fig. 9 and Fig 10 it is noticeable that the first method is much more efficient.For threshold 20% Fuzzy C-Means has about 80% classification efficiency while Fuzzy Cognitive Maps has about 60% efficiency. The full classification efficiency in the first method is obtained for a threshold equal of 60% , while for the second method for a threshold equal to 90%.

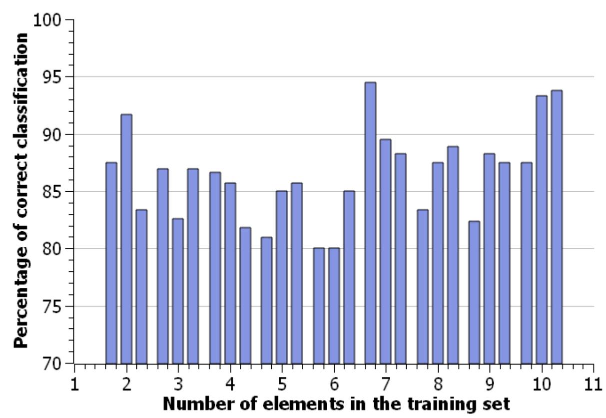

Fig. 8. Efficiency of bird classification.

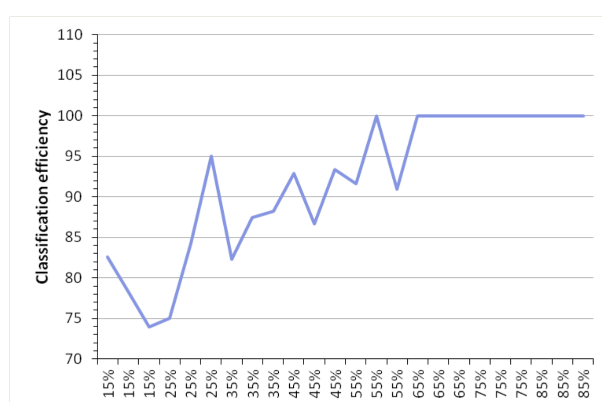

Fig. 9. Classification efficiency for clusters designated by using Fuzzy C-Means.

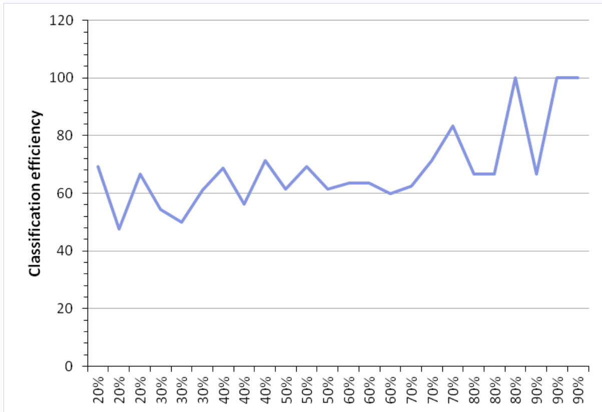

Fig. 10. Classification efficiency for clusters designated by using Fuzzy Cognitive Maps.

The results obtained suggest that the analysed method of classifying birds based on their arrival dates in individual years is very promising and opens new possibilities for avian ecologists. It is enough to observe the arrival dates of only a few bird species to most likely determine the approximate arrival date of other species belonging to the same bird category. Appropriate selection of predictors (bird species in the development of which predictive models will be developed) remains an open problem.

VII. Conclusions and Future Work

The obtained results showed that analysing data of birds' first arrival dates with the use of fuzzy classifiers gives us additional information. It was possible to divide the whole group of early coming birds into smaller clusters in which the arrivals date depends on the coming of a few birds. The use of fuzzy algorithms allows to form clusters (subgroups) of bird species having similar relationship to the birds shown in Table III, and to build forecasting models.

In further research the Fuzzy Cognitive Map classification efficiency can be improved upon by incorporating sub maps from other aspects, such as average winter temperature or short distance migration information. Later the complexity of introducing many concepts from fuzzy models can be verified with the use of machine learning. It is worth noting that this study is innovative and it is reported an early stage of research here. In addition, due to changing climatic conditions, it is difficult to forecast strongly stable future behaviour of bird groups. Therefore, it is hard to find a ground truth for obtained results. Hence, the task of determination of model parameters which could serve as comparative values instead of arbitrary numbers appearing in presented experiments will be an interesting research field to explore in the nearest future.

Acknowledgment

Funded by the National Science Centre, Poland under CHIST-ERA programme (Grant no. 2018/28/Z/ST6/00563).

References

[1] T. L. Root, et al., "Fingerprints of global warming on wild animals and plants," Nature, vol. 421, no. 6918, pp. 57–60, 2003.

[2] C. J. Butler "The disproportionate effect of global warming on the arrival dates of short-distance migratory birds in North America," Ibis, vol. 145, 484–495, 2003.

[3] M. Sepp, V. Palm, A. Leito, K. Päädam, and J. Truu, "Effect of atmospheric circulation types on spring arrival of migratory birds and long-term trends in the first arrival dates in Estonia," Est. J. Ecol., vol. 60, pp. 111–131, 2011.

[4] C. Kullberg, et al., " Change in spring arrival of migratory birds under an era of climate change: Swedish data from the last 140 years," Ambio vol. 44, pp. 69–77, 2015.

[5] W. Biaduń, I. Kitowski, and E. Filipiuk, "Trends in the arrival dates of spring migrants in Lublin (E Poland)," Acta Ornithol. vol. 44, pp. 89–94, 2009.

[6] S. E. Newson, et al., "Long-term changes in the migration phenology of UK breeding birds detected by large- scale citizen science recording schemes," Ibis, vol. 158, no.3, pp. 481–495, 2016.

[7] J. P. Moussus, R.., Julliard, and F. Jiguet, "Featuring 10 phenological estimators using simulated data," Methods Ecol. Evol., vol. 1, pp. 140–150, 2010.

[8] N. Saino, et al., "Climate change effects on migration phenology may mismatch brood parasitic cuckoos and their hosts," Biol. Lett., vol. 5, pp. 539–541, 2009.

[9] P. S. T. H. Tryjanowski and T. H. Sparks, "Is the detection of the first arrival date of migrating birds influenced by population size? A case study of the red-backed shrike Laniuscollurio," Int. J. Biometeorol., vol. 45, no. 4, pp. 217–219, 2001.

[10] J. C. Bezdek, R. Ehrlich and W. Full, "FCM: The fuzzy c-means clustering algorithm," Comput. Geosci., vol. 10, no. 2-3, pp. 191–203, 1984.

[11] B. Kosko, "Fuzzy cognitive maps," Int. J. Man Mach. Stud., vol. 24, pp. 65–75, 1986.

[12] N.K. Kasabov, "Foundations of Neural Networks, Fuzzy Systems and Knowledge Engineering," 2nd ed., A Bradford Book The MIT Press Cambridge, Masschusetts, London (England) 1998.

[13] G. Nápoles, L. Concepción, R. Falcon, R. Bello, and K. Vanhoof, "On the accuracy–convergence trade off in sigmoid fuzzy cognitive maps," IEEE Trans. Fuzzy Syst., vol. 26, no. 4, pp. 2479–2484, 2017.

[14] L. Kuncheva, Fuzzy classifier design. Springer Science & Business Media, 2000.

[15] T. R. Gabriel and M. R. Berthold, " Influence of fuzzy norms and other heuristics on "mixed fuzzy rule formation"". Int. J. Approximate Reasoning, vol. 35 no. 2, pp. 195–202. 2004.

[16] Z. Dikopoulou and E. Papageorgiou, "Inference of Fuzzy Cognitive Maps (FCMs)," https://cran.r-project.org/web/packages/fcm, July 2017.