Evolving Systems (2010) 1:73–82
DOI 10.1007/s12530-010-9002-1

ORIGINAL PAPER

Evolvable fuzzy systems: some insights and challenges

Witold Pedrycz

Received: 28 December 2009 / Accepted: 10 June 2010 / Published online: 1 July 2010

Abstract The emerging category of evolvable fuzzy systems has opened a new uncharted territory of system modeling by enhancing the capabilities of existing fuzzy models and formulating new methodological and algorithmic challenges and opportunities. In this study, we revisit the underlying concept and identify a number of essential optimization problems arising therein. It is shown that the behavior and characteristics of evolvable systems can be classified under the rubric of perception-based evolvability (being inherently associated with the human-centric systems and the development of efficient mechanisms of relevance feedback) and a distribution of knowledge representation resources of evolvable systems. We elaborate on the essence of these problems and define the corresponding optimization criteria. A selected detailed design scenario is presented as well in which the dynamics of information granules is exploited as a vehicle to cope with the evolving modeling environment.

Keywords Evolvable systems · Dynamic information granulation · Fuzzy clustering · Perception-based evolvability · Adjustable knowledge representation resources

1 Introduction

Evolvability of systems is one of the properties that makes fuzzy models or models, in general, being more in rapport with reality. The concept itself exhibits many facets and is quite often linked with an idea of adaptive systems- the architectures, which have been around for a long time. The systems (models) have to evolve because the real world is often nonstationary. The perception of the world changes depending upon a certain observer (user) and his preferences and in this way the model evolves to adhere to the needs of the users.

From the algorithmic perspective, being treated in a narrow sense, the evolvable system should possess an ability to realize some parametric adaptation, which means that it has to be endowed with a suite of parameters (so that the required flexibility is present in the system) along with mechanisms of adaptation so that the required adjustments can be realized in an effective fashion. Neural networks are typical representative examples of adaptive systems.

In general terms, the evolvable systems are characterized by abilities to adjust their structure as well as parameters to the varying characteristics of the environment (with the term of environment embracing processes/ phenomena with which the system has to interact or deal with the users using the system). Human-centric systems, including recommending systems (recommenders), are usually evolvable systems as they have to cope with continuously changing preferences of the users and address the need to deal with adjustable requirements. The relevance mechanisms have to be carefully accommodated through building the evolvability capabilities of the system.

The notion of evolvable systems along with the development strategies have undergone a number of refinements

W. Pedrycz (✉)
Department of Electrical and Computer Engineering,
University of Alberta, Edmonton, AB T6R 2V4, Canada
e-mail: pedrycz@ee.ualberta.ca

W. Pedrycz
Warsaw School of Information Technology, Warsaw, Poland

and with this regard one can refer to a series of representative studies carried out so far [see (Angelov 2004a, b); (Angelov et al. 2008); (Huang et al. 2007); (Jain et al. 1999)]. The concept of evolvability is perceived in different ways and in some cases it looks at the underlying optimization processes (which exhibit a great deal of adaptive mechanisms). The category of genetic-based optimization of fuzzy models has assumed here a visible position [the reader may refer here to (Akbarzadeh-T et al. 2008); (Al-Razgan and Domeniconi 2006); (Aliev et al. 2009); (Antonelli et al. 2009); (Molina et al. 2006, 2007); (Park et al. 2007); (Pham and Castellani 2006)].

The objective of this study is to elaborate on the facets of conceptual and algorithmic diversity of evolvable systems. We discuss on the two essential and general directions present in the realm of evolvable systems, that is a concept of adjustable knowledge representation resources and an idea of evolvability of perception perspectives. Those are the two dominant dimensions of the concept of evolvability. While this taxonomy is of general nature, for illustrative purposes, our investigations are focused on fuzzy rule-based systems as this architecture of fuzzy models can help exemplify the main ideas in a clear and convincing manner. While the detailed algorithmic studies are presented as well, some aspects of general development character are only highlighted.

The material is structured into six sections. We start with a brief introduction to the functional components of rule-based models and discuss a main ways in which they are constructed. The perspective of adjustable knowledge representation is presented in Sect. 3, which is followed by a discussion on evolvability of perception perspectives where a certain descriptive fuzzy model is cast in the suitable setting by adjusting the nonlinear mapping of the input spaces (Sect. 4). Section 5 is devoted to the detailed algorithmic developments of evolvable systems where the underlying design is discussed in detail. Conclusions are covered in Sect. 6.

As already mentioned, to focus our investigations, we consider a broad category of fuzzy systems (models) structured as a collection of rules of the form

$$\text{if } \mathbf{x} \text{ is } R_{\mathrm{i}} \quad \text{then } y = f_i(\mathbf{x}, \mathbf{a}_i) \tag{1}$$

In a more direct forms of fuzzy models (in which all input variables are articulated explicitly), we encounter the rules of the following form:

$$\text{if } x_1 \text{ is } A_i \text{ and } x_2 \text{ is } B_i \text{ and } x_3 \text{ is } C_i \ldots \quad \text{then } y = f_i(\mathbf{x}, \mathbf{a}_i) \tag{2}$$

where the inputs (**x**) come from $\mathbf{R}^{\mathrm{n}}$ and R_i is a fuzzy set defined in the Cartesian product of the input variables. The vector of parameters of the local functions f_i standing in the conclusion part of the rules is denoted by $\mathbf{a}_i$. Fuzzy sets formed for the individual variables are denoted by A_i, B_i, C_i, … etc. The number of rules is equal to “c”.

2 Fuzzy models: a generic functionality and design

In spite of the genuine diversity of differences and ensuing design algorithms, there are a few common generic components of rule-based fuzzy models and the ways, in which they are put together to construct the fuzzy model.

2.1 Information granules

The information granules are usually realized through clustering (Bodenhofer and Bauer 2003) or fuzzy clustering, say Fuzzy C-Means (FCM) (Bezdek 1981); (Pedrycz 2005); (Pedrycz and Hirota 2007). FCM operates on numeric data and produces a family of “c” clusters (information granules) R_1, R_2,…, R_{c} with membership functions conveyed by the partition matrix. The prototypes are the other component of the granular structure of data. The geometry of the clusters—fuzzy sets is implied by the form of the distance function used in the generation of the clusters (typically a Euclidean distance is considered) and the value of the fuzzification coefficient (m). The higher the number of information granules, the larger the number of rules and potentially the higher accuracy of the resulting model albeit its interpretability (transparency) becomes lower.

2.2 Local models

Those are descriptors of input–output relations whose relevance is limited to the region of the input space specified by the information granules standing in the condition part of the rule. There is a high diversity possible here. Local models could be linear models, polynomials, neural networks, differential equations; just listing a few examples. Local models set up a tradeoff between approximation capabilities and interpretability. For instance, local models coming in the form of neural networks exhibit high accuracy but low interpretability. Obviously, along with the nature of the model, comes a certain level of computational effort and optimization procedures. They could be either gradient-based learning or simple pseudo-inverse matrix computing yielding a global minimum of the performance index guiding the estimation of parameters of the local linear models.

2.3 Aggregation

The rules are aggregated giving rise to the overall output of the overall model. The common aggregation mechanism uses a weighted sum of the local models where the weights are the activation levels of the condition parts of the corresponding rules. The typical expression reads as follows

$$\hat{y} = \sum_{i=1}^{c} \phi(R_1, R_2, \ldots, R_c)\mathrm{L}_i(\mathbf{x}) \tag{3}$$

with ϕ being a function of the degrees of activation of the conditions (R_1, R_2,…, R_c) of the consecutive rules present in the system. Here L_i stands for the local model present in the conclusion part of the i-th rule.

3 Evolvable systems: a perspective of adjustable knowledge representation resources

The observation about the architecture of fuzzy rule-based models, especially when cast in the context of human-centric and interpretable systems (Pedrycz 2005), presented in the previous section, entails that from a more abstract perspective such evolvable models can be viewed in terms of its knowledge representation perspective or knowledge representation resources. More specifically, we can talk here about several important components such as e.g.,

(a) the number of rules, which is implied by the number of information granules used to granulate input variables or collections of input variables,
(b) the form of information granules, and
(c) the dimensionality of the input space

These components are usually considered and adjusted when optimizing multidimensional systems [see (Molina et al. 2006)]. Let us assume that these knowledge representation resources (capabilities) can be quantified in the form of a single index Γ. The index itself can be treated as a function of the number of clusters, dimensionality of the input space, and other factors in addition to those listed above. Of course, an assumption of having this single index is quite strong and one might be cognizant that in some cases forming this index might not be feasible and may call for additional effort in which the domain knowledge has to be engaged.

In a simple version, the index Γ can be formed by taking a weighted sum of the individual knowledge representation resources, say

$$\Gamma = w_1\Gamma_1 + \mathrm{w}_2\Gamma_2 + \cdots + w_p\Gamma_p \tag{4}$$

with w_1, w_2, …, w_p being the weights of the individual components of the index while Γ_1, Γ_2,…, Γ_p representing the indexes quantifying the nature of the individual resources available to the realization of the system. For instance, the quantification of some of them could be quite straightforward by being linked directly to the corresponding resources, say $\Gamma_1 = c$ (the number of rules), $\Gamma_2 = p$ (the dimensionality of the input space). Both of them can vary as the granularity of the model (the number of rules) changes and the adjustable dimensionality of the input space (which could be modified by invoking the mechanisms of feature reduction) can be modified.

In the context, is worth recalling a wealth of studies reported in the existing literature [see, for instance, (Bodenhofer and Bauer 2000, 2003, 2005)]. With regard to the readability of rule bases, one can allude to the components such as the number of rules, total rule length, and an average rule length. The intuitively appealing idea could be quantified through the multiplicative form of the expression, Comp * Part * Cov, where Comp represent the complexity of the classifier (as the rule-based model is focused on classification problems), Part is the average normalized partition index (computed as the inverse of the number of information granules decreased by 1), and Cov is the normalized coverage level of the fuzzy partition. The study reported in Alonso et al. (2009) takes into account a more elaborate index in which a larger number of factors are involved. More specifically, there are in total six variables such as: total number of rules, total number of conditions, number of rules which use only a single input, number of rules which use two variables, number of rules using three or more variables, the total number of labels—information granules defined per input.

Some other indexes may require more attention when proceeding with the quantification of the underlying effect, say, the form of information granules in which case we may quantify the aspect of transparency of information granules (in terms of interaction occurring between them and an explicit quantification of individual variables versus the use of multidimensional fuzzy sets defined over all input variables). The values of the weights are estimated on a basis of input provided by the user/expert.

The well-known indexes of complexity of system modeling including minimum description length (MDL) principle, the Bayes information criterion (BIC) or the Akaike information criterion (AIC) might be worth considering albeit one should be cognizant that they apply to numeric models not distinguishing between different semantics of the individual parameters of the granular (rule-based) models.

The multifaceted nature of knowledge representation becomes more apparent considering that in addition to the interpretability of the produced source of knowledge, an issue of "stability" of knowledge is of interest as well. In a nutshell, a collection of information granules being an abstract view at the experimental evidence can be constructed in different ways resulting in similar albeit not identical information granules. The variability of the information granules contributes to the stability of the acquired knowledge. More specifically, let us denote by

$\boldsymbol{G} = \{G_i\}$ and $\boldsymbol{F} = \{F_j\}$, $i = 1, 2,\ldots, c_1$, $j = 1, 2,\ldots, c_2$ information granules built on the basis of the same source of experimental evidence $\boldsymbol{X}$. These information granules are expressed through the corresponding rows of the partition matrices being the result of running some algorithm of fuzzy clustering, say FCM. Note that in general c_1 is not equal to c_2. The stability of the obtained information granules can be expressed in the form of the following index stability index S expressing the distance $||\cdot||^2$ of the corresponding proximity matrices Prox(G_i) and Prox(F_j),

$$S = ||\Pr\text{ox}(\boldsymbol{G}) - \Pr\text{ox}(\boldsymbol{F})||^2 \tag{5}$$

where the entries of the matrix Prox($\boldsymbol{G}$) are determined on a basis of the partition matrix U obtained as a result of granulation of $\boldsymbol{X}$, that is

$$\text{Prox}(\boldsymbol{G}) = U^{\text{T}} \circ U \tag{6}$$

with its (k_1, k_2) entry being computed by summing the minima of the corresponding columns of the partition matrix, that is

$$\sum_{i=1}^{c} \min(u_{ik_1}, u_{ik_2}) \tag{7}$$

The facet of stability of information granules (forming the condition part of the rules) can be taken into consideration in the realization of the index Γ.

The evolvability of the model is guided by the values of Γ, which are used as a high-end objective function governing all modeling activities. For instance, the task can be formulated as follows:

- design a fuzzy model such that the associated index of knowledge representation is Γ_0 and the accuracy of the models becomes maximized.

The evolvability aspect comes into play because even though the value of Γ_0 is given, the distribution of the resources can evolve (change)—clearly we might have a large number of modeling scenarios characterized by different allocations of knowledge representation resources from which those offering the highest level of accuracy are selected.

One may envision scenarios in which the evolvability of the model is governed by the increasing values of Γ:

- over the course of modeling (in successive releases of the model), keep increasing the level of knowledge representation so that in consecutive releases of the model, they exhibit increasing values of Γ while increasing the accuracy of modeling.

This strategy is concerned with the creation of a series of models, which become more specific. The opposite strategy may promote evolvability towards the use of more compact systems by assigning less knowledge representation resources:

- over the course of modeling (in successive releases of the model), keep decreasing the level of knowledge representation so that in consecutive releases of the model, they exhibit decreasing values of Γ while potentially retaining the accuracy of modeling.

Given the complexity of the underlying problem, the optimization of the fuzzy models characterized in this manner calls for evolutionary optimization as one of the feasible development alternatives. If we confine ourselves to Genetic Algorithms, the main task is to capture the structure in the suitable format of the chromosome.

4 Evolvability of perception perspectives

A fuzzy model regarded as a collection of rules interacting among themselves can be treated to a significant extent as a construct of prescriptive nature where knowledge about a system of phenomenon is encapsulated in a form of rules. The knowledge could be the one of commonsense nature, which is quite general, but not necessarily capturing the specificity and minute details of the system but instead focusing the very essence of the phenomenon. The same system may be perceived from different points of view depending upon the objectives of the observers and their objectives. The same model (which captured the prescriptive nature of the phenomenon) is still relevant but needs calibration. Fuzzy sets offer this possibility as in contrast to symbols (which semantics cannot be easily adjusted) they come with numeric membership functions, which could be made evolvable. This important aspect becomes a crux of the evolvability expressed in terms of the associated perception perspective. Figure 1 highlights the essence of this concept.

The fuzzy model positioned at the center of this figure is of prescriptive nature and usually specified in advance through capturing the underlying domain knowledge. For instance, it can be composed through a collection of fairly commonsense—driven rules whose relevance is self-explanatory but for which we might anticipate some need of further calibration or adjustment. With this regard, information granules and information granulation play an important role [cf. (Zadeh 1997)]

The realization of adjustments of perception can be conveniently realized by a nonlinear transformation of the input space in which fuzzy sets are formed. Consider a family of fuzzy sets A_1, A_2,…, A_c defined in **R**, A_i: **R** → [0,1]. The nonlinear transformation of the universe of discourse described as F: $x \to z$ results in the collection of transformed information granules, that is $A_{\text{i}}(F(x))$, $i = 1$,

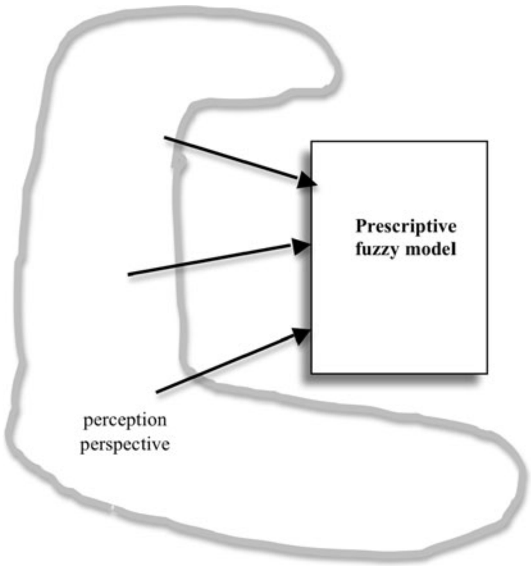

Fig. 1 The concept of evolvability of perception perspectives impacting the development of the model

2,…,c. They produce an individual perception to which the general model adheres. The flexibility of the nonlinear transformation is provided by expressing it by a collection of piecewise linear functions, see Fig. 2, or a sum of translated sigmoid functions. In both cases we note a nonlinear transformation of the original input space. For instance, as shown in Fig. 2, the original segment [a_1, a_2] is "stretched" while some other, say [a_i, a_{i+1}] is very much "suppressed" to a very limited region in the transformed space. The realization of focus (attenuation) of attention (where the region is stretched) or reduction of attention is accomplished through the linear segments of the transformation function of different slopes. A similar effect is realized by running a series of sigmoid functions having different slopes.

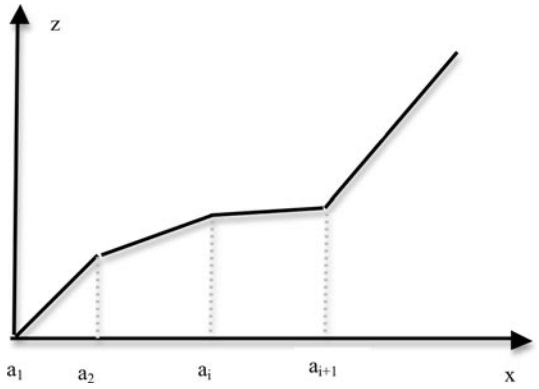

Fig. 2 Example of nonlinear (piecewise linear) transformation F of the input space x

The realization of the transformation is completed as an optimization problem of the best representation of the prescriptive model. Denote this prescriptive model by FM(A_1, A_2,…, A_c). The incorporation of the perception through the nonlinear transformation results in the fuzzy model expressed in the form FM($\mathbf{x}$; $A_1, A_2, \ldots, A_c, f_1, f_2, .., f_n$) where the individual transformation functions play a pivotal role. The experimental data associated with the specific perception of the system is given as input–output pairs ($\mathbf{x}(k)$, target(k)), $k = 1, 2, \ldots, N$. The performance index Q to be minimized is defined as a sum of errors

$$Q = \sum_{k=1}^{N} [\mathrm{FM}(\mathbf{x}_k; A_1, A_2, \ldots, A_c, f_1, f_2, \ldots, f_c) - \mathrm{target}_k]^2 \tag{8}$$

Its minimization is realized with respect to the transformation functions f_1, f_2,.., f_n (and their parameters such as e.g., endpoints of the segments of piecewise approximation or translations and slopes of the sigmoid functions)

$$\mathrm{Min}_{f_i \in \boldsymbol{F}} Q \tag{9}$$

where $\boldsymbol{F}$ is a family of all admissible transformation functions under consideration. So far we have discussed only two classes of such functions, which are intuitively appealing. In general, we can think of the families such that they satisfy the condition of continuity and monotonicity. The nonlinear functions could be either monotonically increasing (those were the functions already discussed) or monotonically decreasing. In this case, we may think of reversal (antonyms) of semantics of the original linguistic terms used in the prescriptive model. The overall schematic view of the realization of the perception-based evolvability concept via nonlinear transformation of the input spaces is illustrated in Fig. 3.

Given the nature of the optimization task, instead of considering gradient-based algorithms (which could be of limited relevance here), one can envision the use of techniques of evolutionary optimization or population-based methods.

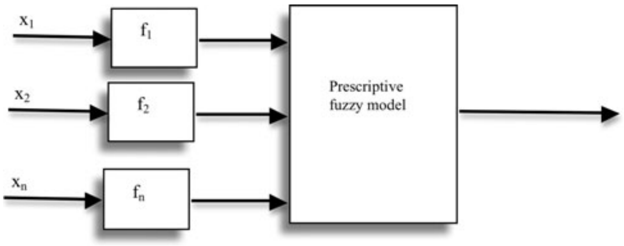

Fig. 3 Nonlinear transformations of input spaces feeding original inputs to the prescriptive fuzzy model

5 Development of evolvable systems

In this section, we get into a more detailed scheme of the design of evolvable systems. Before proceeding with this, let us show two representative practical scenarios.

(a) In processing temporal slices of multivariable data such as e.g., in sensor networks, stock market, currency exchange rates, etc. readings are analyzed over some time slices and for each time slice a fuzzy model is constructed.
Apparently the phenomenon varies over time and its complexity and variability could fluctuate by moving up and down in consecutive time segments (say, months, quarters, etc.). To accommodate this component of inherent variability, the models constructed in successive steps can change their granularity as schematically displayed in Fig. 4.

(b) In forming models for successive slices of volume, 3D type of data (such as e.g., those present in 3D biomedical data, atmospheric data and alike), the models constructed for successive slices have to exhibit some continuity but at the same time have to accommodate the variability of the phenomenon observed at successive volumes (see Fig. 5).

In both cases, there is an obvious necessity to build an evolvable architecture of the fuzzy model. The integral component of the structure that evolves is concerned with information granules and their variability over neighboring models (with these models either being in the relationship of spatial or temporal neighborhood). Once they have been established, building local models becomes predominantly a matter of parametric optimization.

In presence of dynamically collected data, information granules move around, become more refined or generalized—all of these effects are driven by the nature of data being collected and their continuous changes. Similarly, it is intuitive to anticipate that there is some structural continuity that is the clusters (information granules) evolve in a “smooth”, continuous manner when moving from one model to the next one. In other words, in the next time slice, the clustering is run again starting from the initial configuration, which has been “inherited” from the previous slice. As stressed, the number of clusters is not fixed in advance but could fluctuate from one data snapshot to another depending upon the structure of data themselves and the complexity of the underlying dynamics of the data (that is the underlying phenomenon generating the data). With reference to the evolvability of information granules and updates of the level of granularity, it is worth recalling some related studies pertaining to clustering data streams, see e.g., (Beringer and Hullermeier 2006); (Crespo and Weber 2005); (Guha et al. 2003); (Huang et al. 2007); (Park and Lee 2007); (Strehl et al. 2002); (Bandyopadhyay et al. 2006); (Tasoulis and Vrahatis 2005).

Referring to Fig. 6, let us elaborate in more detail on the process of clustering. We start with the first data snapshot, run the FCM on the data present there by producing a collection of the prototypes $\mathbf{v}_1[1]$, $\mathbf{v}_2[1]$, …, $\mathbf{v}_c[1]$ (the number in the square bracket denotes that those prototypes are built on the basis of the data coming the first data segment $\mathbf{X}[1]$). The number of clusters is determined based on the acceptable level of the predetermined optimization criterion. We elaborate on the computational details later on. When running the FCM on the second data snapshot $\mathbf{X}[2]$, we proceed with the prototypes that were formed for $\mathbf{X}[1]$ and use them as the starting point when running the

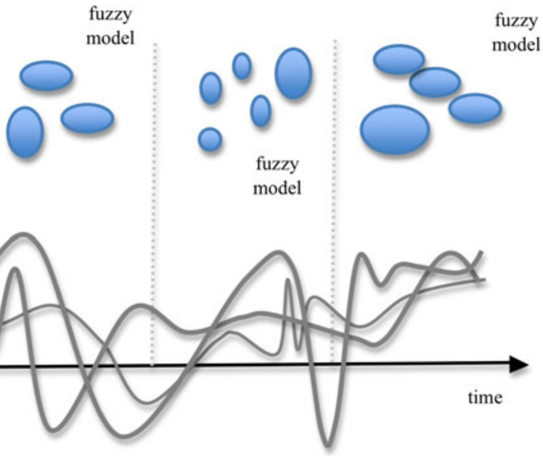

Fig. 4 Fuzzy models built for successive time slices of multivariable temporal data

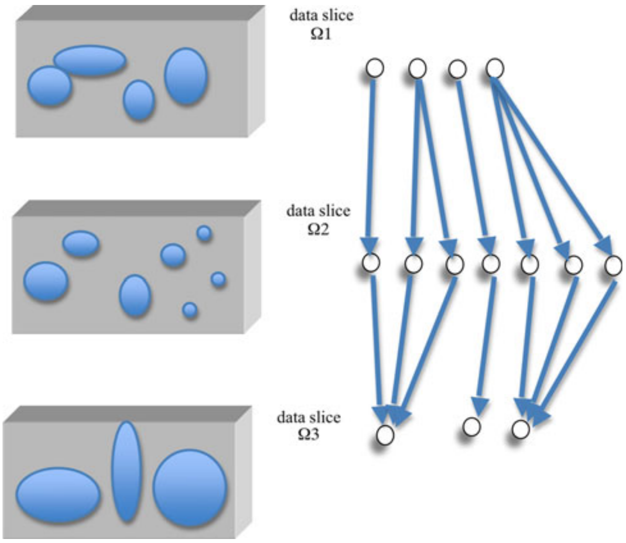

Fig. 5 Development of models for successive slices of volume data; shown is an evolution of underlying information granules (position in the input space, splitting and merging)

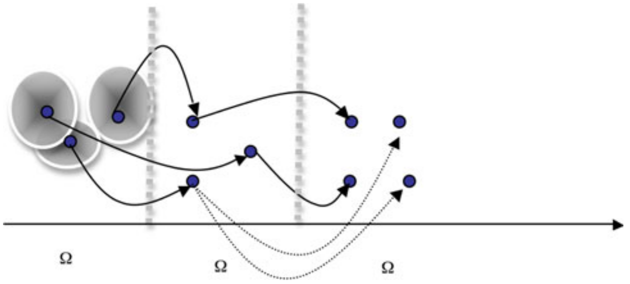

Fig. 6 The concept of dynamic clustering- progression of the FCM constructs (prototypes and partition matrices) in successive data snapshots and dynamic adjustments of the number of the clusters (prototypes) through the mechanisms of cluster merging and cluster splitting

FCM for **X**[2]. Inevitably, when iterating FCM, the prototypes change their location as a result of revealing the topology of the actually processed data. Once done, it could well be that a certain clustering quality indicator (performance index) produced there could exceed some required threshold or could assume values that are far lower than this threshold. In both cases, the number of clusters needs to be dynamically adjusted. In the first case, it must be increased—we could do that by splitting the most heterogeneous (diversified) cluster. In the second case, the number of clusters could be too high (capturing too many unnecessary details) and some merging process could be of interest.

In what follows, we are concerned with (a) splitting and merging criteria, (b) assessment of quality of clusters which could be directly used when controlling the dynamics of the clusters and deciding upon the change of the number of clusters themselves, and (c) optimization capabilities of fuzzy clustering (FCM) which could be exploited upfront when running the algorithm.

5.1 Reconstruction criterion in the assessment of quality of clusters

The number of clusters is another important design aspects of any clustering algorithm. While there has been a number of so-called cluster validity indexes (Bezdek 1981); (Jain et al. 1999), which are geared to assess the most "suitable" number of clusters that is supposed to correspond with the "true" structure present in the data. The inherent problem with them is that quite often one receives with several conflicting suggestions depending which cluster validity index has been selected. In this study, we follow a different path which could be referred to as a granulation-degranulation strategy [cf. (Pedrycz and Hirota 2007); (Pedrycz and Gomide 2007)]. In essence, once the information granules have been formed by the FCM, we use the prototypes to represent each pattern in the form of information granules by computing the corresponding membership degrees. At the next step, we complete a de-granulation, which is estimate the pattern on a basis of the prototypes and the membership degrees. Ideally, we would expect that the result of such de-granulation should return the original pattern. While this is quite unlikely, we may consider the distance between the original pattern and its de-granulated version to constitute a viable measure of quality of obtained reconstruction. Let us move to the technical details which provide some required formal description of the strategy.

Denote by $\mathbf{X}_i$ the data whose membership to the i-th cluster is the highest, that is

$$\mathbf{X}_i = \{\mathbf{x}_k \in \mathbf{R} | i = \arg_j \max \mathrm{u}_j(\mathbf{x})\} \tag{10}$$

The reconstruction error for the i-th cluster is computed as the following sum of distances between $\mathbf{x}_k$ and its degranulated version $\hat{\mathbf{x}}_k$

$$V_i = \sum_{\mathbf{x}_k \in \mathbf{X}_i} \|\mathbf{X}_k - \hat{\mathbf{x}}_k\|^2 \tag{11}$$

where

$$\hat{\mathbf{x}}_k = \frac{\sum_{i=1}^{c} u_{ik}^m \mathbf{v}_i}{\sum_{i=1}^{c} u_{ik}^m} \tag{12}$$

The maximal value of V_i, $V_{\max} = \max_i V_i$, can serve as a suitable measure describing the reconstructibility aspects provided by the information granules. By varying the number of the clusters (c) we can affect the values of V_i and in this way focus on selecting the most suitable structure in the data set.

5.2 Dynamic formation of fuzzy clusters—splitting and merging mechanisms

In what follows, we discuss the mechanisms of splitting and merging clusters. They become a crux of dynamic information granulation moving into details, we describe that the objective function minimized by the FCM for the 1st data snapshot exploits the standard sum of distances

$$Q[1] = \sum_{i=1}^{c[1]} \sum_{\mathbf{x}_k \in \mathbf{X}[1]} u_{ik}^m ||\mathbf{x}_k - \mathbf{v}_i[1]||^2 \tag{13}$$

(here $Q[1]$ concerns the first data snapshot). Let us also note that the number of clusters, say $c[1]$ may be selected on a basis of the assumed threshold level, $V_{\max}$. For the first data snapshot, we choose such $c[1]$ for which the reconstruction error does not exceed some predefined threshold level, that is $V_{\max} < e$; with the value of e, $e > 0$, given in advance. Here we may adjust the number of the clusters so that the required inequality is satisfied. The choice of the threshold level can be realized by observing how much the values of the reconstructability criterion are affected by the number of clusters; it might well be that V exhibits a visible knee-like relationship when varying the number of clusters. We select the minimal value at "c" at which observed is a radical reduction in the values of the performance index under consideration. Subsequently this particular value of the threshold "e" becomes associated with the corresponding value of V. With respect to the number of clusters, one may consider the use of any of so-called cluster validity indices (partition coefficient, partition entropy, etc.), which can navigate our search for the number of clusters, viz. the granularity of the findings. A certain drawback associated with these validity indexes is that each of them could eventually point at different number of clusters (which might be quite confusing) or some indexes could exhibit a fairly monotonic character which prevents us from using them as a reliable indicator of the preferred level of granularity of the structure. Some cross-validation techniques could be more suitable. Alternatively, one can envision running several clustering methods and then forming consensus as being encountered in consensus clustering [cf. (Pedrycz and Hirota 2008); (Strehl et al. 2002)].

5.2.1 *Cluster splitting*

If the reconstruction criterion of the clustering completed for some data snapshot does not meet the reconstruction criterion by exceeding the assumed threshold e, then some clusters need to be split. Intuitively, this split could involve the cluster characterized by the highest level of variability, that is the one for which the index V_i attains a maximum among all the clusters. Let this cluster be denoted by i_0. The split is realized by conditional clustering (Pedrycz 1998) of the data falling under the realm of the elements belonging to the i_0-th cluster that is $\mathbf{X}_{i_0}$. Denote the set of indexes of these patterns by $\mathbf{I}_0$. The formulation of the clustering problem is given in the form

$$\min \sum_{i=1}^{2} \sum_{k\in \mathbf{I}_0} f_{ik}^{m} ||\mathbf{x}_k - \mathbf{z}_i||^2 \tag{14}$$

subject to the constraint

$$\sum_{j=1}^{2} f_{jk} = u_{i_0k} \tag{15}$$

In essence, by minimizing (11) we invoke a computing process in which we split the i_0-th cluster into two clusters maintaining a balance of the membership grades of the original cluster. Here $\mathbf{z}_i$ are the prototypes and $F = [f_{ik}]$ denotes a partition matrix that satisfies constraint (15) meaning that the split is driven by the data and membership grades of the most diversified cluster. The term conditional clustering comes from the fact that we require that the resulting membership degrees f_{ik} are distributed according to the membership values available for the cluster to be split. The detailed calculations of the partition matrix F and the two prototypes $\mathbf{z}_1$ and $\mathbf{z}_2$ are carried out iteratively according to the two expressions

$$f_{ik} = \frac{u_{i_0k}}{\sum_{j=1}^{2} \left(\frac{||\mathbf{x}_k - \mathbf{z}_i||}{||\mathbf{x}_k - \mathbf{z}_j||} \right)^{1/(m-1)}} \tag{16}$$

and

$$\mathbf{z}_i = \frac{\sum_{k\in i_0} f_{ik}^{m} \mathbf{x}_k}{\sum_{k\in i_0} f_{ik}^{m}} \quad i = 1,2;\ \mathbf{x}_k \in \mathbf{X}_{i_0}. \tag{17}$$

5.2.2 *Cluster merging*

If the objective function for the number of the clusters is lower than the predefined level e then the number of the clusters could be reduced meaning that some original clusters could be combined together (merged). The way of doing that would be to merge two prototypes that are the closest to each other. Consider that these two are indexed by "s" and "l", that is $\mathbf{v}_\mathrm{s}$ and $\mathbf{v}_\mathrm{l}$. The new prototype $\mathbf{v}^{\sim}$ results from the minimization of the following expression with respect to it, the new prototype $\mathbf{v}^{\sim}$ results from the minimization of the following performance index

$$Q = \sum_{k=1}^{N} (u_{sk} + u_{lk})^m ||\mathbf{x}_k - \mathbf{v}^{\sim}||^2 \rightarrow \text{Min } Q \text{ with respect to } \mathbf{v}^{\sim} \tag{18}$$

If for these reduced number of the clusters the resulting reconstruction criterion is still lower than the critical value, we can move on with another reduction of the number of the prototypes. The process could be repeated by reducing the number of the clusters by one.

Let us rewrite the performance index (18) in a coordinate-wise fashion

$$\begin{aligned} Q &= \sum_{k=1}^{N} (u_{sk} + u_{lk})^m ||\mathbf{x}_k - \mathbf{v}^{\sim}||^2 \\ &= \sum_{k=1}^{N} \sum_{j=1}^{n} (u_{sk} + u_{lk})^m \left(x_{kj} - v_j^{\sim} \right)^2 / \sigma_j^2 \end{aligned} \tag{19}$$

We make the derivative of Q equal to zero, $\nabla_{\mathrm{v}^{\sim}} Q = 0$ and this leads to the new prototype

$$\mathbf{v}^{\sim} = \frac{\sum_{k=1}^{N} (u_{sk} + u_{lk})^m \mathbf{x}_k}{\sum_{k=1}^{N} (u_{sk} + u_{lk})^m} \tag{20}$$

5.3 Characterization of structural dynamics of the families of clusters

The structure of clusters changes from one data slice to another. It is of interest to characterize these changes at some general level so we could learn about the dynamics of the structural changes of the evolvable model. There are several possible alternatives we can consider in describing the evolvable structure.

5.3.1 Granularity of structure

The number of clusters $c[1]$, $c[2]$,…. reported in the consecutive data slices can be viewed as the simplest and the most concise structural characteristics one could think of. The increasing sequence of these numbers is indicative of the growing structural complexity (which manifests in the increasing number of the clusters one has to construct to deal with it). This does not tell us however at all about the relationships between the prototypes produced in two neighboring data slices.

5.3.2 Relationships between information granules

To capture more detailed structural relationships, we take a different detailed path whose crux is the following. Given the prototype $\mathbf{v}_i[\mathrm{ii}+1]$ we express it in terms of the prototypes $\mathbf{v}_j[\mathrm{ii}]$ available at the ii-th data snapshot. More specifically, we obtain degrees of membership of the i-th prototype expressed by $\mathbf{v}_j[\mathrm{ii}]$ which are determined in a "standard" manner

$$w_{ik} = \frac{1}{\sum_{j=1}^{c[\mathrm{ii}]} \left(\frac{||\mathbf{v}_i[\mathrm{ii}+1]-\mathbf{v}_k[\mathrm{ii}]||}{|\mathbf{v}_i[\mathbf{ii}+1]-\mathbf{v}_j[\mathbf{ii}]|}\right)^{2/(m-1)}} \tag{21}$$

where $i = 1, 2,\ldots, c[\mathrm{ii}+1]$ and $k = 1, 2,\ldots, c[\mathrm{ii}]$. This representation is schematically captured in Fig. 7.

If all these values are either close to 0.0 or close to 1.0, this becomes indicative of a sound representation of the prototypes in terms of the structure we have revealed so far. On the other hand, if the values of w_i tend to be close to $1/c[\mathrm{ii}]$, we may talk about a high level of uncertainty when expressing the prototypes in the language of the already available structure. More formally, we can quantify the structural diversity by introducing a functional φ defined on the set of membership degrees such that $\varphi\ (0) = \varphi\ (1) = 0$ which is increasing over $[0, 1/c[\mathrm{ii}])$ attains the maximal value at $1/c[\mathrm{ii}]$ and then monotonically decreases to zero. The sum of the following form

$$T = \sum_{i=1}^{\mathrm{c[ii+1]}} \varphi(w_i) \tag{22}$$

serves as a concise indicator of the uncertainty of structural differences observed from one data slice to another.

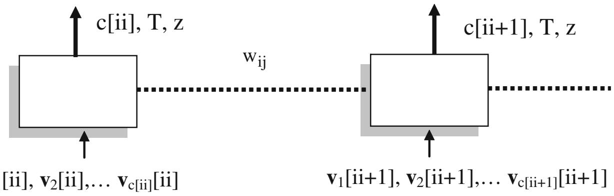
c[ii], T, z
w_ij
c[ii+1], T, z
v1[ii], v2[ii],… vc[ii][ii]
v1[ii+1], v2[ii+1],… vc[ii+1][ii+1]

Fig. 7 Expressing relationships between information granules through describing information granules in terms of the granules obtained for the previous data slice

The collection of connections $[w_{ij}]$ formed by the elements computed by (21) can be used to construct a relational model of structural relationships. Consider the i-th element (prototype) observed in the $(\mathrm{ii}+1)$-th data slice. It is described by the prototypes present in the ii-th data slice whose activation levels are w_{i1}, w_{i2}, …, $w_{ic[\mathrm{i}]}$. The same occurs for any other prototype available in the $(\mathrm{ii}+1)$-th data slice. This gives rise to the following collection of input–output pairs of data

1st prototype $\mathbf{y}(1) = \left[w_{11}w_{12}\ldots w_{1\mathrm{c[ii]}}\right]$
$\mathbf{z}(1) = [1 \quad 0 \quad 0 \quad 0 \quad \ldots \quad 0]$

2nd prototype $\mathbf{y}(2) = \left[w_{21}w_{22}\ldots w_{2c[\mathrm{ii}]}\right]$
$\mathbf{z}(1) = [0 \quad 1 \quad 0 \quad 0 \quad \ldots \quad 0]$

$c[\mathrm{ii}+1]$prototype
$\mathbf{y}(c[\mathrm{ii}+1]) = \left[w_{c[\mathrm{ii}+1]1}w_{c[\mathrm{ii}+1]2}\ldots w_{c[\mathrm{ii}+1]c[\mathrm{ii}]}\right]$
$\mathbf{z}(c[\mathrm{ii}+1]) = [0 \quad 0 \quad 0 \quad 0 \quad \ldots \quad 1]$

Given this structure of data we arrive at the estimation problem in which the output is expressed by means of the activation levels w_{ij}. In general we can formalize the relationship as a certain set of fuzzy relational equations

$$\mathbf{z}(k) = \mathbf{y}(k) \cdot R \tag{23}$$

$k = 1, 2,\ldots, c[\mathrm{ii}+1]$ where · stands for a certain relational operator, such as e.g., max–min, min–max or their generalized versions of max-t and min-s with "t" and "s" being the corresponding t-norm or t-conorm. The fuzzy relation R is to be determined and one could consider using one of the estimation methods (which might lead to exact or approximate solutions to the problem). One can also describe the i-th prototype in terms of the prototypes obtained for the ii-th data slice by considering the following aggregation of weighted prototypes available at the previous data slice

$$\hat{\mathbf{v}}_i[\mathrm{ii}+1] = \sum_{j=1}^{c[\mathrm{ii}]} w_{ij}^m \mathbf{v}_j[\mathrm{ii}]. \tag{24}$$

6 Conclusions

Evolvable systems enrich the landscape of fuzzy systems and make them well suited for human-centric systems in which the mechanisms of adaptability play a crucial role. We have stressed a need for a broad definition of such systems in which a general nature of evolvability has to be embraced. While the parametric adjustments are a part of the overall design agenda, the structural aspects of modifications are of far higher relevance. We introduced two general categories of evolvability, which emphasize the need for managing and securing a suitable allocation of knowledge representation resources (including the variable granularity of rules, order of polynomial local models and alike). Interestingly, the allocation process is predominantly concerned about the structure of the model and one has to proceed prudently with the optimization of the structure so that the accuracy of the model is also retained. Here the design framework supporting global optimization (including evolutionary and population-based optimization) becomes indispensible. The variability of perception perspectives (associated with human-centricity of systems embracing a realization of effective relevance feedback mechanisms) gives rise to another category of evolvable systems. We showed that one of the alternative approaches exercised here can focus on the adjustment of contexts via a nonlinear transformation of the individual input variables in which way we can emphasize the details of some ranges of the variables and generalize some others.

While in this study, we elaborated on the formulation of the pertinent problems and discussed some selected scenario at the algorithmic level, the detailed optimization cases have not been tackled and are left for further comprehensive studies.

Acknowledgments Support from the Natural Sciences and Engineering Research Council of Canada (NSERC) and Canada Research Chair (CRC) is gratefully acknowledged.

References

Akbarzadeh-T M-R, Davarynejad M, Pariz N (2008) Adaptive fuzzy fitness granulation for evolutionary optimization. Int J Approx Reason 49(3):523–538

Aliev RA, Guirimov BG, Fazlollahi B, Aliev RR (2009) Evolutionary algorithm-based learning of fuzzy neural networks. Part 2: recurrent fuzzy neural networks. Fuzzy Sets Syst 160(17):2553–2566

Alonso JM, Magdalena L, Gonzalez-Rodriguez G (2009) Looking for a good fuzzy system interpretability index: an experimental approach. Int J Approx Reason 51:115–134

Al-Razgan M, Domeniconi C (2006) Weighted clustering ensembles. In: Proceedings of 6th SIAM international conference on data mining, pp 258–269

Angelov P (2004a) An approach for fuzzy rule-base adaptation using on-line clustering. Int J Approx Reason 35(3):275–289

Angelov P (2004) A fuzzy controller with evolving structure. Inf Sci 161(1–2):21–35

Angelov P, Lughofer E, Zhou X (2008) Evolving fuzzy classifiers using different model architectures. Fuzzy Sets Syst 159(23):3160–3182

Antonelli M, Ducange P, Lazzerini B, Marcelloni F (2009) Learning concurrently partition granularities and rule bases of Mamdani fuzzy systems in a multi-objective evolutionary framework. Int J Approx Reason 50(7):1066–1080

Bandyopadhyay S, Giannella C, Maulik U, Kargupta H, Liu K, Datta S (2006) Clustering distributed data streams in peer-to-peer environments. Inf Sci 176:1952–1985

Beringer J, Hullermeier E (2006) Online clustering of parallel data streams. Data Knowl Eng 58:180–204

Bezdek JC (1981) Pattern recognition with fuzzy objective function algorithms. Plenum Press, New York

Bodenhofer U, Bauer P (2000) Toward an axiomatic treatment of interpretability. In: Proceedings of the 6th international conference on soft computing IIZUKA2000, pp 334–339

Bodenhofer U, Bauer P (2003) A formal model of interpretability of linguistic variables. In: Casillas J, Cordon O, Herrera F, Magdalena L (eds) Interpretability issues in fuzzy modeling, Springer, Berlin, pp 524–545

Bodenhofer U, Bauer P (2005) Interpretability of linguistic variables: a formal account. Kybernetika 41:227–248

Crespo F, Weber R (2005) A methodology for dynamic data mining based on fuzzy clustering. Fuzzy Sets Syst 150:267–284

Guha S, Meyerson A, Mishra N, Motwani R (2003) Clustering data streams: theory and practice. IEEE Trans Knowl Data Eng 15:515–528

Huang YP, Hsu CC, Wang SH (2007) Pattern recognition in time series database: a case study on financial database. Expert Syst Appl 33:199–205

Jain A, Murt M, Flynn P (1999) Data clustering: a review. ACM Comput Surv 31:264–323

Molina C, Rodríguez-Ariza L, Sánchez D, Vila MA (2006) A new fuzzy multidimensional model. IEEE Trans Fuzzy Syst 14(6):897–912

Montiel O, Castillo O, Melin P, Díaz AR, Sepúlveda R (2007) Human evolutionary model: a new approach to optimization. Inf Sci 177(10):2075–2098

Park NH, Lee WS (2007) Cell trees: an adaptive synopsis structure for clustering multi-dimensional on-line data streams. Data Knowl Eng 63:528–549

Park HS, Pedrycz W, Oh SK (2007) Evolutionary design of hybrid self-organizing fuzzy polynomial neural networks with the aid of information granulation. Expert Syst Appl 33(4):830–846

Pedrycz W (1998) Conditional fuzzy clustering in the design of radial basis function neural networks. IEEE Trans Neural Netw 9:601–612

Pedrycz W (2005) Knowledge-based clustering: from data to information granules. Wiley, Hoboken, NJ

Pedrycz W, Gomide F (2007) Fuzzy systems engineering. Wiley, Hoboken, NJ

Pedrycz W, Hirota K (2007) Fuzzy vector quantization with the particle swarm optimization: a study in fuzzy granulation–degranulation information processing. Signal Processing 87:2061–2074

Pedrycz W, Hirota K (2008) A consensus-driven clustering. Pattern Recognit Lett 29:1333–1343

Pham DT, Castellani M (2006) Evolutionary learning of fuzzy models. Eng Appl Artif Intell 19(6):583–592

Strehl A, Ghosh J, Cardie C (2002) Cluster ensembles: a knowledge reuse framework for combining multiple partitions. J Mach Learn Res 3:583–617

Tasoulis D, Vrahatis M (2005) Unsupervised clustering on dynamic databases. Pattern Recognit Lett 26(13):2116–2127

Zadeh LA (1997) Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets Syst 90:111–117