A Fuzzy Data-Driven Paradigmatic Predictor

Farzad Amirjavid * **Hamidreza Nemati** ** **Sasan Barak** ***

* *Edward S. Rogers, Department of Electrical and Computer Engineering, University of Toronto, Toronto, Canada (e-mail: farzad.amirjavid@utoronto.ca).*
** *Engineering Department, Lancaster University, Lancaster, UK (e-mail: h.nemati@lancaster.ac.uk).*
*** *Department of Management Science, Lancaster University Management School, Lancaster, UK (e-mail: s.barak@lancaster.ac.uk).*

Abstract: Data-driven prediction of future events is to provide decision-makers Predictive Information (PI) to decrease human-error. They usually desire possession of a predictor which works independently from the humanized configurations and works efficiently and accurately. The accurate data-driven prediction of the systems' behavior is the primary focus of this paper. We define the future state of a system is a set of uncertain values, which can be modeled by fuzzy numbers. The future machine state is not very dissimilar to the current status, and the next event is a sort of behavior repetition. The PI also justifies the system being in a trend to achieve a goal, and it counts the unplanned contextual reactions of the system. In this paper, we come up with a fuzzy data-driven predictor application to foretell the system behavior.

Keywords: Fuzzy logic, temporal data analytics, adaptive learning, systems theory.

1. INTRODUCTION

The Predictive Information (PI) is a sort of actionable insight, which allows decision-makers to act upon, and which gives them enough insight into the future that the actions that should be taken become clear. Technically, to provide initial data of system behaviors experts perform the frequent observations on the systems and form the Systematic Temporal Datasets (STDs). Then to machinery mining of the systems, we attribute four characteristics to their behavior: stability, radicalism, goal achievement, and behavior repetitions [Djaferis and Schick (2000); Rapoport (1986); Lobry (1973)]. Per each paradigm, there is a predictor, and the aggregator concludes the individual predictions into a totalitarian hypothesis, which is the final prediction of the Fuzzy Data-Driven Predictor (FDDP) as shown in Fig.1. We also consider the role of an Anomaly Recognizer, which compares the realized events with predicted ones to find the anomalies. We argue that in Data-Driven Machine Learning (DDML) and prediction, by improving the clustering rule, we can get more accurate results but in faster time. The reason is that it helps to understand the previous and current states more quickly and more precisely to provide biases for prediction. The other criterion to improve the DDML predictions is attention to the new events and trends that happen after training, and we address this issue by the term "adaptivity." We concentrate on the prediction of the future states of the real-world systems as the initial problem. This paper is organized as the following. In section (2) we discuss the preliminaries of the current paper and we review some important and relevant works. In section (3), we propose our FDDP model. In the next section, we experiment the

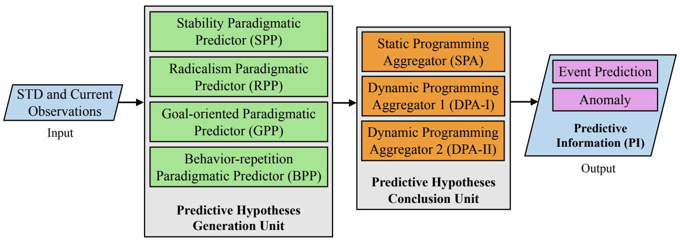

Fig. 1. The model of Fuzzy Data-Driven Predictor (FDDP) and Anomaly Recognizer (FDDAR)

proposed model (4) and compare the results with others', and the functionality of the model with other works. In 5, we perform conslusion and propose suggestions to the interested researchers. Finally, we present the applied references.

2. PRELIMINARY AND THE RELATED WORKS

Data modeling provides biases for machinery perception of the previous and current states of the systems. In this regard, data classification and clustering are the two popular methodologies used for DDML and prediction, and Each instance of these algorithms operate based on a particular logic and philosophy. An important parameter to list the related works is that they are either Knowledge-Driven (KD) or Data-Driven (DD) [Amirjavid (2014)]. In KD, the experts have prior knowledge about the congestion specs of the observation data points [Lu et al. (2017)]. In these problems, expert experimentally knows the boundaries of the clusters and the ranges of the target data. But, in most of the issues, the user does not initially know where the data points are congested around and how much the size and proportion of the congestions are. A first solution is to choose a clustering method such as k-means or Fuzzy C-Means (FCM). Due to the segmentation operation with k-means, it finds the centroids rapidly. But, knowing the number of clusters is still a requirement. The fuzzy subtractive clustering is a known popular algorithms for the problems in which the experts intend to get tailored cluster sizes and positions from STDs [Kruger et al. (2017)]. This algorithm reviews the data, finds the congested areas and defines the soft cluster boundaries in the neighborhood of the data points that are the farthest to the corresponding centroids. There is also a group of works that focus on classification and data modeling rule to improve the prediction rate. This rule is done in a KD manner. We identified the cited papers in Table 1 out of the recent related works that focused on *"Clustering"* and *"Classification"* rules in temporal data mining. These works extended/customized derivations of more known algorithms such as "k-means" and "SVM" to their concerning problems.

According to the presented information in Table 1, and considering that a full data-mining cycle includes both clustering and classification then an ideal approach would predict as accurate as 96.5% $\times$ 96% $=$ 92.64% with KD strategy and 95.75% $\times$ 94% = 90.005% with DD strategy. These rates are usually taken by more complexity order algorithms and the total complexity will be of $O(n^2) : O(n^4)$. However, a full cycle of DDML, which works independent of expert configurations with light complexity order ($O(x)$, where $x < n^2$) is not proposed. To our best knowledge, if such cycle be proposed then the average prediction accuracy rate with the recent and relevant works will span around the rate of $90 - 93\%$. The proposed predictors are not generally extensible to predict the STDs of other domains and a full human-independent DDML cycle regarding the model knowledge (unsupervised) and model-training process (DD) is required.

3. FDDP MODELLING

For data-driven machine learning of the system, the m features ($f_i | 1 \leq i \leq m$) are observed. Each feature is assumed to have totally minimum and maximum values, which means $\{x | x \in f_i, x \in [0\ 1], 1 \leq i \leq m\}$, and the FDDP analyzes m-dimensional space. The inputs to the FDDP are the observations ($o_{i,j}$ in equation 1), which are frequently taken off the system behaviors. The observations matrix is STD, which has m columns and n rows, whereas $m, n > 0$. Therefore, the current observation will sit on the last row, and it will be the $n + 1^{th}$ row of the STD matrix.

$$STD = \begin{array}{c} \\ S_{t_1} \\ S_{t_2} \\ \vdots \\ S_{t_n} \end{array} \begin{array}{c} \begin{array}{cccc} f_1 & f_2 & \dots & f_m \end{array} \\ \begin{pmatrix} O_{1,1} & O_{1,2} & \dots & O_{1,m} \\ O_{2,1} & O_{2,2} & \dots & O_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ O_{n,1} & O_{n,2} & \dots & O_{n,m} \end{pmatrix} \end{array} \quad (1)$$

In Equation (1), f_i is the i^{th} world feature, S_{t_j} is the j^{th} observation or the observation in time j.

In [Amirjavid et al. (2018)], a derivation of subtractive clustering is proposed, which inputs the STD dataset and the desired similarity degrees between the data points of the clusters (ε). The first output of this algorithm is the list of centroids (matrix C) and the sizes of the clusters (matrix Sz).

The first sort of specific clusters that the FDDP forms is the "machine state". This a set of continuous observations or a set of sequential rows ($s_{t_{i..j}}$) in STD that are similar to each other with similarity degree of ε. A machine state is a fuzzy concept such as:

$$\begin{aligned} \tilde{s}_{t_{i..j}} &= \{S_{t_x} | S_{t_x} \in STD, x \in [i..j] | 0 \leq i \leq j \leq n, \\ &\exists C(k) | \forall S_{t_x} \rightarrow |S_{t_x} - C(k)| < 1 - \varepsilon\}, 0 \leq k \leq w \end{aligned} \quad (2)$$

Based on this concept, we determine the present/current state and the previous state. By present/current state, we mean a subset of machine state that are recently observed. This fuzzy set is defined as the following:

$$\tilde{S}_{present} = \{x | x \in \tilde{s}_{t_{i..j}},\ |x.time - current\, time| \leq 1 - \varepsilon\} \quad (3)$$

Alike the "current state", the "previous state" is a sort of machine state that includes all the previous machine states but excludes the recent observations:

$$\tilde{S}_{past} = \{x | x \in \tilde{s}_{t_{i..j}},\ |x.time - current\, time| < \varepsilon\} \quad (4)$$

We use the basic concepts mentioned above in design of the paradigmatic predictors.

3.1 Stability Characteristic

The systems are stable, and their behavior is stable over the time. By this statement, we infer that the two continuous observations from a system are not very dissimilar to each other, and the system changes the machine states smoothly/slowly. If $\boldsymbol{S}_{t_i}$ and $\boldsymbol{S}_{t_{i+1}}$ are two consecutive observations, then a threshold measure symbolized by ε defines for the maximum difference between two observation vectors:

$$\left|\boldsymbol{S}_{t_i} - \boldsymbol{S}_{t_{i+1}}\right| < \varepsilon, \forall i | 0 > i, i+1 \leq n, \varepsilon > 0 \quad (5)$$

The equation (5) indicates each two consecutive observations are similar to each other. Therefore, if $\boldsymbol{S}_{t_i}$ is the latest observed status from the world, then the future upcoming state $\boldsymbol{S}_{t_{i+1}}$ will be a measure in following constraints:

$$|\boldsymbol{S}_{t_n}| - \varepsilon < \left|\boldsymbol{S}_{t_{n+1}}\right| < |\boldsymbol{S}_{t_n}| + \varepsilon \quad (6)$$

According to equation (6), the future event will be a vector similar to the present vector, which means $|\boldsymbol{S}_{t_n}| \cong \left|\boldsymbol{S}_{t_{n+1}}\right|$. See Figure. 2. In Figure. 2.A, we are presenting that the feature i of the future state of the system will be of Gaussian fuzzy number, in which the current value of feature i has the highest fuzzy membership value to occur in future. In Figure. 2.B, we are presenting a two dimensional presentation of the future state of the features i and j. This space is the product of $s_{t_n}.f_i \times s_{t_n}.f_j$. This prediction space is a fuzzy set. It gives the highest possibility degree of occurrence to the center of the circular space, which refers to the current value of feature i and the current value of feature j. The ε of equation (6) indicates a soft

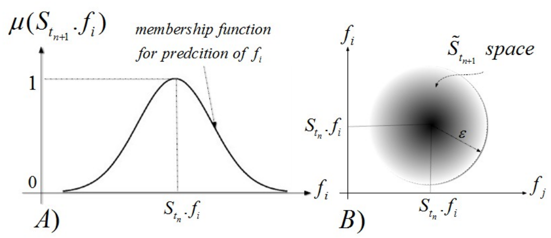

Fig. 2. Fuzzy predictive hypothesis generation by paradigmatic predictors. A: Gaussian function, B: 2-D representation.

Table 1. Review of related works that focused on improvement of clustering rule for prediction

KD/DD	Approach	Accuracy	Derived algorithm	Complexity	Focus
KD	Kim et al. (2016)	85.40%	KNN	$O(n)$	Classification
	Arif et al. (2013)	90.02%	ID3 and C4.5	$O(n)$	Classification
	Vluymans et al. (2016)	91%	FIS	O(n.log(n))	Classification
	Lopez-Garcia et al. (2016)	83.10%	FRBS	N/A	Classification
	Lippi et al. (2013)	95%	Regression and SVM	$O(n)$	Classification
	Baggenstoss and Harrison (2016)	93%	SVM	N/A	Classification
	Abercrombie and Friedl (2016)	96.50%	HMM	$O(n^2)$	Classification
KD	Huang et al. (2013)	75%	Regression	$O(n)$	Clustering
	Debonnaire et al. (2016)	87%	K-means	$O(n)$	Clustering
	Wang et al. (2017)	96%	Hierarchical	O(n^2.log(n))	Clustering
DD	Yin et al. (2017)	95.75%	SVM	$O(n^2)$	Classification
	Mori et al. (2016)	71%	K-means	$O(n)$	Clustering
DD	Kruger et al. (2017)	90%	FCM	$O(n)$	Clustering

radius of a circle that represents the possible future values of the feature i. The ε does not represent a hard boundary indicating the impossibility measures of the world features, but it shows the limitation implied in current calculations. A possibilistic/fuzzy 2D space as similar as the Figure. 2.B can be drawn per each couple of the features to predict totally the $\boldsymbol{S}_{t_{n+1}}$ vector. The possibility degrees at the radius borders and beyond there are not zero, but not noticeable. As a result by this we indicate that by exclusive respect to the stability feature of the systems, the vector $\boldsymbol{S}_{t_{n+1}}$ is predicted in the form of a multidimensional fuzzy number, which is $\tilde{S}_{t_{n+1}}$.

Definition 1: Current state. The *Current observations* cluster is the cluster, which includes the current observation and the last recent observations from STD. The algorithm for determination of current state cluster is in Fig. 3. The SPP performs this process. We show fuzzy space of $\tilde{S}_{t_{n+1}}$, which is produced by SPP, by the notion $\tilde{P}_{SPP}$. By P_{SPP} we point to the defuzzified of the $\tilde{P}_{SPP}$.

3.2 Radicalism Characteristic

By radicalism, we address the tendency of a dynamic system to leave the current state regardless of whatever reason makes the system leaves the current status. A fuzzy operator called Absolute Fuzzy Symmetrizer Operator (AFSO) finds the most dissimilar status to the current status [Amirjavid (2014); Amirjavid et al. (2018)]:

$$AFSO(\boldsymbol{S}_{t_n}) = \boldsymbol{S}'_{t_n} = \{1-x|x \in S_{t_n}.fi, \forall i|i \in [1..m]\} \quad (7)$$

By this paradigm, the future state of the system is:

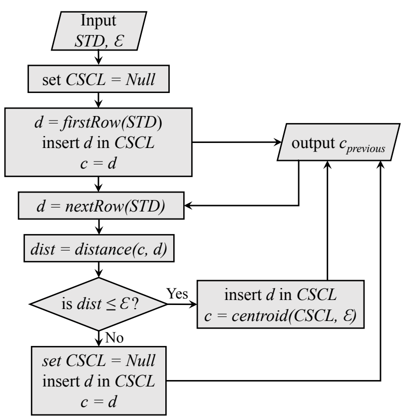

Fig. 3. Algorithm regarding Dynamic formation of current state cluster (current fuzzy state)

$$\boldsymbol{S}_{t_{n+1}} = (AFSO(\boldsymbol{S}_{t_n}) + \boldsymbol{S}_{t_n}) \times \lambda \ , \ 0 \leq \lambda \leq 1 \quad (8)$$

In (8), the λ is a factor that indicates the tendency of the system to the radicalism behavior. We show fuzzy space of $\tilde{S}_{t_{n+1}}$, which is produced by RPP, by the notion $\tilde{P}_{RPP}$.

3.3 Goal Orientation Characteristic

Systems are to achieve a particular goal. The goal is a world state, and to make it the system transits a series of temporary states. The goal is a fuzzy vector $\tilde{G}$, which justifies the next state of the system in transition to meet it. Therefore, the future machine state will be in a possibilistic space between the $\boldsymbol{S}_{t_n}$ and the $\tilde{G}$:

$$\boldsymbol{S}_{t_{n+1}} = \{\tilde{x}|\tilde{x} \in [\boldsymbol{S}_{t_n}, \tilde{G}]\} \quad (9)$$

In (9), the $\tilde{G}$ is not known, but the next machine state is in trend of the two precedent states to achieve the goal:

$$\boldsymbol{S}_{t_{n+1}} = (\boldsymbol{S}_{t_n} - \boldsymbol{S}_{t_{n-1}}) + \boldsymbol{S}_{t_n} = 2\boldsymbol{S}_{t_n} - \boldsymbol{S}_{t_{n-1}} \quad (10)$$

The present state in (10) comes from the algorithm in Fig. 3, and the previous state comes from the algorithm in Fig. 4. The GPP performs the prediction based on goal orientation feature of the systems. We show fuzzy space of $\tilde{S}_{t_{n+1}}$, which is produced by GPP, by the notion $\tilde{P}_{GPP}$. By P_{GPP} we point to the defuzzified of the $\tilde{P}_{GPP}$.

3.4 Behavior Repetition Paradigm

The systems usually repeat their behaviors. We expect the system repeats a behavior whenever the context (a set of preliminary conditions) for this issue comes up. We define the context as the similar state in STD and the "behavior repetition" as the following state that the system has already experienced in that situation:

$$\tilde{S}_{t_{n+1}} = \tilde{S}_{t_{x+1}} | \tilde{S}_{t_n} \cong \tilde{S}_{t_x}, x \in [1...n-1] \quad (11)$$

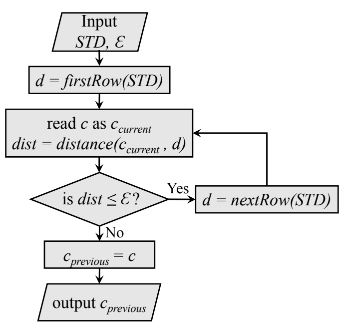

Fig. 4. The method to compute the previous state

In (11), we are indicating that the future event will be an action that the system has already accomplished whenever it was in similar context.

Theorem 1: According to the stability paradigm in (5, 6) since two consecutive observations are similar, then the BPP output (P_{BPP}) as the future state of the machine is the most similar centroid in history of the system behavior:

$$\tilde{S}_{t_{n+1}} = \tilde{S}_{t_x} | \tilde{S}_{t_n} \cong \tilde{S}_{t_x}, x \in [1...n-1] \tag{12}$$

Therefore, according to the (12), the BPP calculates the centroid, which is the most similar centroid to the current observation as the future event. We show fuzzy space of $\tilde{S}_{t_{n+1}}$, which is produced by BPP, by the notion $\tilde{P}_{BPP}$. By P_{BPP} we point to the defuzzified of the $\tilde{P}_{BPP}$.

3.5 Aggregation of The Paradigmatic Predictions

We apply an aggregator to achieve a single number as of the final prediction. See Fig. 1. Presuming the "XPP" and "YPP" are two typical paradigmatic predictors, and we define the $Aggregate(P_{XPP}, P_{YPP})$. The predictive values of XPP and YPP are either intersected or distinct/disjoint. In Fig. 5 and Fig. 6 we are providing a visual representation of the intersection and distinction between two fuzzy predictions.

If the centroids of $\tilde{P}_{XPP}$ and $\tilde{P}_{YPP}$ are close to each other then they make intersection area like in Fig. 5.A and Fig. 6.A. The radius or effective range of centroids are in L matrix. The L_values correspond the radius measures of the clusters. The predictive centroids make intersection area if:

$$ABS(\tilde{P}_{YPP} - \tilde{P}_{XPP}) \leq L\_values_{XPP} + L\_values_{YPP} \tag{13}$$

In the case that equation (13) is valid, then the two fuzzy numbers coming from paradigmatic predictors produce intersection areas.

Proposition 1: If the predictors make an intersection area (see Fig. 7.A, Fig. 8.A), then the possible area of the occurrence of the predictive event is a member of the intersection area of two paradigmatic predictors, and the plausible punctual event is at the average (in half distance) of the corresponding centroids.

$$\begin{aligned} \tilde{P}_{FDDP} &= \tilde{P}_{XPP} \cap \tilde{P}_{YPP}, \\ \tilde{P}_{FDDP} &= Average(P_{XPP}, P_{YPP}) = \frac{P_{XPP} + P_{YPP}}{2} \end{aligned} \tag{14}$$

If Eq. (13) is not valid, Eq. (14) will not be valid too. The reason is that although, by (13) the plausible event

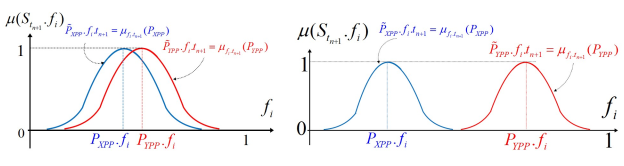

Fig. 5. 1-D intersected (A) and disjoint (B) prediction spaces

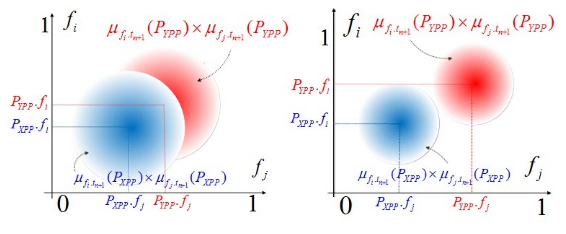

Fig. 6. 2-D intersected (A) and disjoint (B) prediction spaces

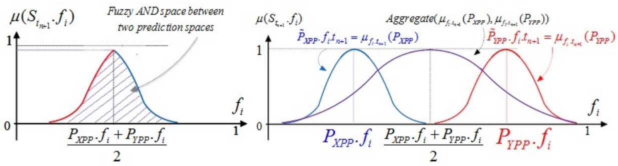

Fig. 7. 1-D aggregation of the intersected (A) and disjoint (B) XPP and YPP

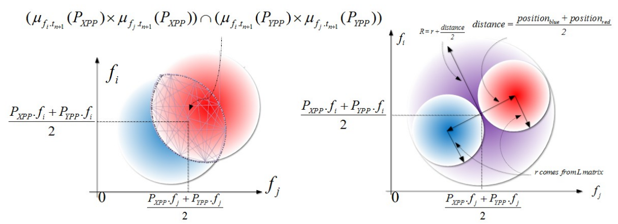

Fig. 8. 2-D aggregation of the intersected (A) and disjoint (B) XPP and YPP

is in space between the predictors, the equation (14) recommends contradictorily low impossibility degree for this estimation. To resolve this problem, the FDDP creates a third larger cluster, which includes, both of the clusters as the predictive estimation. See Fig. 7.B, Fig. 8.B.

Proposition 2: The aggregated prediction of disjoint paradigmatic predictions are computed by (14). This is the new centroid of the new auxiliary cluster. The radius of the new cluster is (R):

$$\begin{aligned} R &= L\_values + \frac{distance(\tilde{P}_{YPP}, \tilde{P}_{XPP})}{2} \\ &= L\_values + \frac{ABS(\tilde{P}_{YPP} - \tilde{P}_{XPP})}{2} \end{aligned} \tag{15}$$

In Fig. 8.B, we show the auxiliary cluster formed by two disjointed fuzzy numbers. The position of the new centroid is calculated by (14), but the new radius of the new cluster (R) is:

$$\begin{aligned} R &= L\_values + \frac{distance(\tilde{P}_{YPP}, \tilde{P}_{XPP})}{2} \\ &= L\_values + \frac{ABS(\tilde{P}_{YPP} - \tilde{P}_{XPP})}{2} \end{aligned} \tag{16}$$

The R of (16) is demonstrated in Figure. 8.B. The final prediction of the FDDP is a fuzzy space represented by the symbol $\tilde{P}_{FDDP}$, and the plausible measure is P_{FDDP}, which is the defuzzified of $\tilde{P}_{FDDP}$.

Proposition 3: We define the "*Problem Solution Space (PSS)*" is the auxiliary cluster, which includes all the individual clusters formed by paradigmatic predictors. "R" from equation (16,15) is the radius of this cluster, and the position of the center of this cluster is:

$$Pos(PSS) = Average(Pos(SPP), Pos(RPP), Pos(GPP), Pos(BPP)) \tag{17}$$

The PSS indicates that the FDDP expects that the predictive values are in PSS and we consider it to be *impossible* that the event occurs out of the PSS specs.

Theorem 2: The "*Static Programming Aggregation (SPA)*" function inputs the predictive values in type of the fuzzy numbers from individual paradigmatic predictors and returns the aggregated predictive value by:

$$\begin{aligned} [PSS, \tilde{P}_{FDDP}] &= SPA(\tilde{P}_{SPP}, \tilde{P}_{RPP}, \tilde{P}_{GPP}, \tilde{P}_{BPP}) \\ &= \pi.R^2, Average(P_{SPP}, P_{RPP}, P_{GPP}, P_{BPP}) \end{aligned} \tag{18}$$

The R values in equation (18) comes from equations (15,16). The smaller values of "R" reveal more cer-

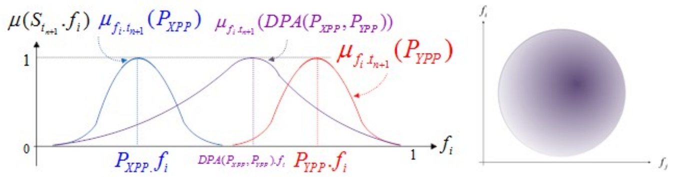

Fig. 9. By dynamic programming the plausible hypothetical prediction moves from average/center toward the more precise predictor (YPP), while the PSS does not change.

tain/reliable predictions regarding the UOS behavior in comparison to bigger values of "R".

3.6 Dynamic programming Aggregation (DPA)

In equation (18), we presume all the paradigmatic predictors have equal weights in the SPA aggregation process. However, in reality, one or more paradigmatic predictors usually work more accurately than the others *temporarily* or *permanently*. To adjust the weight of the paradigmatic predictors in aggregation process FDDP evaluates the functionality of the individual predictors:

$$Prediction\,Accuracy\,of\,XPP = |P_{XPP,n} - S_{t_n}| \quad (19)$$

In (19), the XPP is a typical paradigmatic predictor: $XPP = \{x | x \in \{SPP, RPP, GPP, BPP\}\}$. S_{t_n} is the current observation, $P_{XPP,n}$ is the prediction of XPP; the concerning process is done in time $n-1$ (previous time) and supposed to be observed in time n (current time).

Dynamic Programming Aggregation type 1 (DPA-I): Presuming XPP_i, $i = 1..4$ predicts more accurately than others in time n (current time), then it will have additional weight in average making rule 14 to predict the $S_{t_{n+1}}$:

$$P_{FDDP} = Average(P_{XPP_i}, P_{XPP_i}, P_{XPP_k}) \mid k = 1..4, k \neq i \quad (20)$$

In DPA-I method, the *temporarily* effect of the "more accurate" predictor is taken into account and at the next observation FDDP redoes the processes in equations (19, 20).

Dynamic Programming Aggregation type 2 (DPA-II): This reflects the permanent effect of the predictors in prediction accuracy. DPA-II counts the number of times each of the XPPs are the most accurate in precision since the beginning to the current time:

$$P_{FDDP} = \frac{(i_{SPP}.P_{SPP}) + (i_{RPP}.P_{RPP}) + (i_{GPP}.P_{GPP}) + (i_{BPP}.P_{BPP})}{i_{SPP} + i_{RPP} + i_{GPP} + i_{BPP}},$$
$$i_{XPP} \geq 0, \quad i_{XPP} \in N \quad (21)$$

The visual presentation of the effect of dynamic programming is in Fig. 9. The objective of dynamic programming is to drive the predictions to the more accurate predictors while the PSS does not increase. In this way, the certainty of the predictions ("R" values) does not decrease but the accuracy of the predictions might increase.

4. EXPERIMENTING THE FDDP

We experimented the functionality of the FDDP on the monthly Fama-French 3-Factor Model Dataset.This data is an indicator of the systematic risk of an investment arising from exposure to general market movements. The FF 3-factor model dataset includes three parameters, which are $(r_m - r_f)$, SMB and HML. We predicted the data of FF three parameters model since July 1926 to November 2016. The first step to deal with this dataset is to normalize the dataset so that the data of f_i in STD. We examined the FDDP with multiple values of ε (an indicator of cluster size) for values between zero and one. In overal, the SPA predicts most accurately at $\varepsilon = 0.15$. In Fig. 10, we are representing the inaccuracy rates of the paradigmatic predictors as well as the aggregators at $\varepsilon = 0.15$. According to our test results, in average the static programming aggregation gives the best outcome; however, the dynamic method improves the maximum amount of prediction inaccuracy. See Table 2.

The main objective to accomplish this experiment is to

Table 2. Dynamic programming aggregation of the FF data in $\varepsilon = 0.15$

Inaccuracy	DPA-Il	DPA-II
Average	0.02591598	0.026053
Minimum	0.000894579	0.001572
Maximum	0.215938889	0.218264

show the functionality of the clustering rule within the paradigmatic predictors. We also reflected the application of fuzzy logic in aggregation of predictions that led to more reliable accuracy rates. Moreover, the adoptive learning technique improved the results. Statistically, the BPP predicted 25.6931608% of the observations more accurate than the others, then SPP by 25.23105366%, RPP by 24.7689464% and GPP by 24.1219963% predicted the more accurate the FF data, which explains the characteristic of the FF dataset. According to the related works in section (2), the other proposed methods are customized to and dependent on the problem context. In average they do not propose an accuracy rate any better than 96%. Especially, if we desire a data-driven approach that works fast (being of $O(n)$), then the expected accuracy rate drops to less than $71-90\%$ [Yang et al. (2015); Mori et al. (2016)]. Since, regression lets us model and predict the datasets with no expert prejudgment, then we selected the linear regression method to compare our works with it.

To compare the FDDP results on FF dataset with the regression results, we set a linear regression model to learn each data column by the other columns. For example, column 1 learns the columns 2 and 3, and the column 2 learns the column 3. The average RSQ rate of the linear regression on FF data is 0.229701927, which indicates the

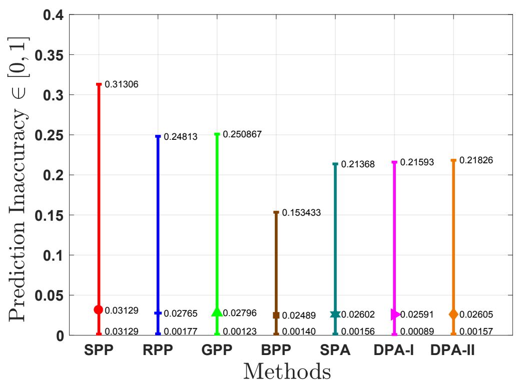

Fig. 10. Inaccuracy rates in prediction of FF dataset at $\varepsilon = 0.15$

inaccuracy rate at the learning step. Therefore, at the prediction time the average inaccuracy rate for prediction is expected to be worse. The FDDP with DPA-II method achieved the inaccuracy rate of 0.026053 at the prediction time. This is though the FDDP works with no preconfiguration and training data, but the regression model is trained with full FF data. [Arif et al. (2013)] for classification of the same dataset. The RSQ rate (inaccuracy rate at the learning step) for the Kagaru Airborne Dataset is 3.064886%, while the FDDP achieved the rate of 1.3112% for prediction inaccuracy of that dataset.

5. CONCLUSION

Four paradigms in the interpretation of systematic temporal data series are considered to predict future events. The FDDP explains the behavior of a system regarding its predictability proportion with the paradigmatic predictors. For instance, since GPP and RPP predicted the FF data the most accurate then we resemble the FF dataset as a radical and goal oriented dataset. We experimented the FDDP on FF dataset to survey the accuracy of prediction, and we verified the role of dynamic programming to propose inaccurate average predictions. We recommend the future researchers to expand the paradigmatic predictors. In particular, we suggest adding the probabilistic predictors in the box of paradigmatic predictors and study their role if it could improve the prediction specs such as accuracy rate and the reliability factor. We propose also to survey the proposed model to recognize the anomalies in economical systems so that the systematic economical crises could be inferred before their occurrence.

ACKNOWLEDGEMENTS

The first author is profoundly thankful and highly influenced by the works of professor Kostas N. Plataniotis (kostas@ece.utoronto.ca), and Dr. Arash Mohammadi (arash.mohammadi@concordia.ca). The second author's work is supported by the Engineering and Physical Sciences Research Council (EPSRC), grant number EP/R02572X/1, and the National Centre for Nuclear Robotics (NCNR). The third author is supported by the Czech Science Foundation (GACR Project GA 17-22662S) and Operational Program Education for Competitiveness - Project No. CZ.1.07/2.3.00/20.0296.

REFERENCES

Abercrombie, S. and Friedl, M. (2016). Improving the consistency of multitemporal land cover maps using a hidden markov model. *IEEE Trans. Geosci. Rem. Sens.*, 54, 703–713.

Amirjavid, F. (2014). A fuzzy temporal data-mining model for activity recognition in smart homes. *PhD Dissertation*, University of Quebec in Chicoutimi.

Amirjavid, F., Spachos, P., and Plataniotis, K. (2018). 3-d object localization in smart homes: a distributed sensor and video mining approach. *IEEE Systems Journal*, 12, 1307–1316.

Arif, F., Suryana, N., and Hussin, B. (2013). A data mining approach for developing quality prediction model in multi-stage manufacturing. *Int. J. Computer Appli.*, 69, 35–40.

Baggenstoss, P. and Harrison, B. (2016). Class-specific model mixtures for the classification of acoustic time series. *IEEE Trans. Aero. Elect. Sys.*, 52, 1937–1952.

Debonnaire, N., Stumpf, A., and Puissant, A. (2016). Spatio-temporal clustering and active learning for change classification in satellite image time series. *IEEE J. Selected Topics in Appl. Earth Observ. Rem. Sens.*, 9, 3642–3650.

Djaferis, T. and Schick, I. (2000). *System Theory: Modeling, Analysis and Control.* Springer, New York.

Huang, B.W., Wang, K.W., Wei, L.Y., and Peng, W.C. (2013). A hybrid prediction algorithm for traffic speed prediction. *Adv. Intell. Syst. Appl.*, 1, 281–296.

Kim, J., Han, Y., and Lee, J. (2016). Data imbalance problem solving for smote based oversampling: Study on fault detection prediction model in semiconductor manufacturing process. *ITCS 2016*, 79–84.

Kruger, B., Vogele, A., Willig, T., Yao, A., Klein, R., and Weber, A. (2017). Efficient unsupervised temporal segmentation of motion data. *IEEE Trans. Multimedia*, 19, 797–812.

Lippi, M., Bertini, M., and Frasconi, P. (2013). Short-term traffic flow forecasting: an experimental comparison of time-series analysis and supervised learning. *IEEE Trans. Intell. Transport. Syst.*, 14, 871–882.

Lobry, C. (1973). *Dynamical Polysystems and Control Theory.* Springer, Dordrecht.

Lopez-Garcia, P., Osaba, E., Onieva, E., Masegosa, A., and Perallos, A. (2016). Short-term traffic congestion forecasting using hybrid metaheuristics and rule-based methods: A comparative study. *17th CAEPIA 2016*, Salamanca, Spain, 290–299.

Lu, M., Lai, C., Ye, T., Liang, J., and Yuan, X. (2017). Visual analysis of multiple route choices based on general gps trajectories. *IEEE Trans. Big Data*, 3, 234–237.

Mori, U., Mendiburu, A., and Lozano, J. (2016). Similarity measure selection for clustering time series databases. *IEEE Trans. Knowl. Data Engin.*, 28, 181–195.

Rapoport, A. (1986). *General System Theory: Essential Concepts & Applications.* Routledge, Cambridge.

Vluymans, S., Tarrago, D., Saeys, Y., Cornelis, C., and Herrera, F. (2016). Fuzzy multi-instance classifiers. *IEEE Trans. Fuzzy Syst.*, 24, 1395–1409.

Wang, L., Zhao, X., Si, Y., Cao, L., and Liu, Y. (2017). Context-associative hierarchical memory model for human activity recognition and prediction. *IEEE Trans. Multim.*, 19, 646–659.

Yang, L., Wang, J., and Khuhro, M. (2015). Pippy search: Anomaly traffic clustering. *IEEE Int. Conf. Smart City/SocialCom/SustainCom (SmartCity)*, Chengdu, 378–383.

Yin, S., Jiang, Y., Tian, Y., and Kaynak, O. (2017). A data-driven fuzzy information granulation approach for freight volume forecasting. *IEEE Trans. Indust. Elect.*, 64, 1447–1456.