ATLANTIS PRESS
Atlantis Studies in Uncertainty Modelling, volume 3
Joint Proceedings of the 19th World Congress of the International Fuzzy Systems Association (IFSA), the 12th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT), and the 11th International Summer School on Aggregation Operators (AGOP)

A Methodology for More sustainable Agriculture Through Early Crop Frost Forecasting

Jose M. Cadenas[a,*] and **M. Carmen Garrido**[a] and **Raquel Martínez-España**[b]

[a] Department of Information Engineering and Communication. University of Murcia. Murcia, Spain
jcadenas@um.es and carmengarrido@um.es

[b] Department of Computer Engineering. Catholic University of Murcia. Murcia, Spain
rmartinez@ucam.edu

Abstract

Climate change is causing abrupt changes in temperature, which adds to earlier flowering of crops and possible crop failure due to negative temperatures at night. Anti-frost techniques exist to save crops but they require natural and human resources which increase costs. Therefore the application of these techniques has to be as precise as possible. In order to facilitate the decision making process for the activate anti-frost techniques, e.g. sprinkler irrigation, the farmer needs an application that is freely available, simple to use and provides a quick and reliable prediction of the expected temperature in an immediate future. The application that provides this information must be available on readily available devices (e.g., a mobile phone). For the application to run on these devices it should have low requirements. For the design of this application, we propose a methodology that obtains small models, and provides it a fast and accuracy. The methodology includes the creation and use of colorreda decision model designed with an instance selection technique and including imprecise information and lagged attributes to respect the time series data and the nature of the information. A real case study is carried out to implement the proposed methodology using Nearest Neighbour technique both instance selection and classification. The results show reduced models with an high accuracy which allows us to create a lightweight model that can be easily updated in a mobile application.

Keywords: Sustainable agriculture, Condensed Nearest Neighbour, Lagged attributes, Imprecise information.

1 Introduction

Precision agriculture provides solutions to important problems for more sustainable agriculture. This discipline began to be implemented in the early 1990s with the initial objective of increasing profitability and reducing environmental impact by using integrated systems within information technologies, [2]. More and more farmers are beginning to have and use technology to address and improve processes in agriculture by trying to develop sustainable agriculture that will reduce their costs and increase their profits [15]. Sustainable agriculture is one of the most important areas of sustainable development. It is also a promising approach for the economic stability of agriculture, especially when it comes to the lives of people in developing countries [3]. In order to adapt and contribute to sustainability, farmers need information from reliable sources and new knowledge that leads to more sustainable decision-making. New technologies and new computational paradigms can help to provide farmers with more information to optimise agricultural processes [1].

Sustainable agriculture deals with problems associated with water resources, loss of crops due to climatic conditions, problems with soil contamination by pesticides, etc. Specifically, there are many areas where climate change is causing major catastrophes in agriculture leading to crop failure and degradation of farmers' economies. Every day, agricultural insurance is becoming more reluctant to insure imposing more restrictions and imposing higher prices on farmers in certain areas and crop varieties. Therefore, farmers must address solutions to mitigate crop damage, even if it means investing both natural and human resources to save crops.

In particular, in south-eastern Spain, climate change produces a situation of temperature imbalance between January and March, with temperatures exceeding 20 degrees Celsius at midday and dropping to negative values at night. The consequences of this temperature variability is the anticipation of crop growth. This anticipation leads to crop failure when temperatures reach negative values. In fruit trees, the anticipation consists of an early blooming of the fruit trees, resulting in the loss of the blossom and thus of the harvest when temperatures fall below zero. In the fight against frost, farmers have several techniques at their disposal that can prevent or mitigate the total loss of the crop. These techniques include covering crops with plastic sheeting, creating a smoke screen by burning fodder, applying chemicals that create a protective layer on flowers and sprinkler irrigation that uses thermal inversion to protect flowers from the cold, [19]. Because of these techniques consume a large amount of both human and natural resources, it is particularly important to use them when it is known as accurately as possible that a frost is going to occur. One aspect to consider when analysing and assessing possible frosts is the area where the crop is located. The weather prediction varies depending on the different plots, therefore it is necessary to have temperature prediction using weather stations as close as possible to the plot/plots where the anti-frost techniques are used. To address this problem, the internet of things (IoT) and cloud computing paradigms can be of great help.

IoT and cloud computing paradigms make a wealth of information available in any real-world environment [16]. This information is captured through sensors and depending on the scope of application, many of the data are numerical values with a time dependency between them. For these problems, the available information is obtained from different stations by several meteorological variables. These variables correspond to time series where there is a dependence between one value and the previous and/or following value. Studies have shown that decision support systems in agriculture can make a significant contribution to long-term sustainable development. However, not all decision support systems use their full potential and are not adapted to farmers' trade-offs and/or needs [15].

This paper focuses on the problem of frost prediction in stone fruit trees. The idea is to obtain as accurate a prediction as possible of whether a frost will occur in the next hours in order to be able to activate the anti-frost technique of sprinkler irrigation in time. This technique has the disadvantage of wasting water if a frost does not occur, but it is one of the most effective and least polluting techniques if a frost does occur. It is important to note that any anti-frost system needs to be put into operation well in advance. To have the trees sufficiently moist, in the case of sprinklers, to have the staff at work to plug in the fans, to switch on the windmills to prepare the environment, etc.

The solution proposed in this paper is the design of a novel methodology to create a decision support system using machine learning techniques embedded in a mobile application where the complexity of use is fully transparent and flexible to the farmer. When we want to extract knowledge from the data using machine learning techniques, it is necessary to perform a previous preprocessing on them. This preprocessing will depend on the specific application in which the data is going to be used [4, 10]. In the case we are dealing with, on the one hand, the creation of instances through lagged attributes to collect the time dependence of the values will be approached. On the other hand, the creation of imprecise attributes to capture the true nature of the data will be performed. Finally, an instance selection technique is used to choose the most representative instances and not to have a heavy decision model, so that this model can be executed through cloud computing and be easily stored in a mobile application. The used instance selection technique in the case study of this paper will be Condensed Nearest Neighbour. The model with the reduction of instances will be used as input to the classifier that will predict whether or not a frost occurs. As a component of the methodology, any classifier capable of handling imprecise data can be used. In the study case of this paper, an extension of k-Nearest Neighbour technique (kNN) is used, [7]. By reducing the size of the data and consequently the size of the model, the possibility of updating the model in the application with the collection of new data is more feasible without the need for a high performance mobile device. Thus the farmer's opposition to decision support system is diminished.

As an example of the importance of the application addressed in this work, we can refer to the situation that has occurred during the month of January 2021 in Spain, in which a heavy snowfall and subsequent frost has led to the loss of many crops with the consequent shortage of supplies for a large part of Europe and the large increase of prices of the little produce available. Furthermore, experts indicate that these more extreme situations are going to become more and more frequent as a consequence of the global warning [11]. Thus, with the new proposed methodology, we are tackling the problem of frost on crops, on the one hand saving harvests and on the other hand saving significant amounts of water, which leads to the development of sustainable agriculture, reducing the environmental impact.

In order to develop the objective of this work, the paper is organised as follows. In Section 2 reviews in a non-exhaustive way several proposals found in the literature to carry out the instance selection in both standard and time series datasets. Section 3 presents the proposed methodology for carrying out instance selection in time series datasets. In Section 4 the proposed methodology is applied to early crop frost forecasting, and some preliminary results are shown and analysed. Section 5 presents the main ideas to develop a mobile phone app to help farmers in their decisions in preventing crop frost. Finally, Section 6 summarises the reached conclusions.

2 Background

Instance selection has been extensively studied in the literature. Without being exhaustive, this section gives a brief presentation of different works in the framework of this data preprocessing.

Given a set of instances E, the goal of any instance selection technique is to obtain a subset $RE \subset E$ with an accuracy similar to the E one.

Instance selection methods can be categorised into Wrapper methods and Filter methods [17]. Wrapper methods have their selection criterion based on the accuracy obtained by a classifier “C” (usually instances that do not contribute to the classifier’s accuracy are removed from the set of instances). Filter methods take as their selection criterion a function that is not based on the accuracy of a classifier.

Many of the Wrapper methods that have been proposed in the literature are based on kNN classifier [9]. One of the first methods is the Condensed Nearest Neighbour (CNN) [12]. This method is incremental, and consists of two steps: in step 1, one instance of each class is included in the set RE; in step 2, each instance $e \in E$ is classified by the model obtained by the classifier “C” using RE as training set and the e instances that produce errors are included in the set RE. Different extensions of CNN have been proposed: SNN (Selective Nearest Neighbour rule) [18] or GCNN (Generalized Condensed Nearest Neighbour rule) [8] among others.

Although instance selection has been widely studied, it has been less so for time series data. Among the works for time series data, those focused on the kNN technique and the DTW (Dynamic Time Warping) distance stand out. Among these works, an instance selection technique called INSIGHT is developed in [5]. Specifically, this technique assigns a score to each instance and selects the ones with the highest scores. To assign these scores they take into account how many times an instance appears as a good neighbour (same classes) of other instances and how many times as a bad neighbour (different classes).

In [21] the AWARD algorithm is proposed which is an iterative procedure based on Nearest Neighbour classifier with DTW distance. In each iteration, the times series are ordered based on their contribution to classify the training set and eliminated from lower ranking. An optimal warping window size is used which depends on the number of time series that are kept. That is, as they remove instances, they update the window size to use for the DTW distance.

The authors of [20] propose a methodology for instance selection consisting of a recursive prediction strategy together with an advanced instance selection criterion using a Nearest-Neighbour based mutual information estimator.

In the next section, our proposed methodology is presented to tackle a forescasting problem from time series datasets by carrying out a preprocessing stage based on instance selection that will allow working with values expressed with imprecise values.

3 A methodology based on instance selection with imprecise values for time series datasets

As discussed above, IoT and cloud computing paradigms make a large amount of information available to applications in any real-world domain. This information, in many cases, is formed by measurements observed in consecutive time instants, giving rise to time series datasets. When we want to extract knowledge from these datasets using machine learning techniques, it is necessary to perform a previous preprocessing on them [4, 10]. Among the most frequent preprocessing tasks that are usually carried out are: transforming the series to the frequency domain using the Fourier Transform, averaging consecutive values to assign them a discrete value (symbolic aggregate approximation), transforming the values into others that describe their evolution over time, etc.

In this section, we define a methodology for the time series datasets preprocesing to be used with machine learning techniques so that 1) the quality of the original knowledge contained in the initial data is maintained and 2) the final dataset is computationally adequate in size for efficient use by machine learning techniques.

The main phases that compose the methodology are:

1. Pre-process the dataset by simplifying the number of attributes through a fuzzy representation, while maintaining all the initial information. Also, this pre-processing transforms the instances by means of lagged attributes to maintain the time series of data.
2. Use of machine learning techniques that work di-

rectly with fuzzy information to select the most relevant instances when building models that are simple, portable and helpful.

An illustrative and more detailed diagram of the phases of the proposed methodology is shown in Figure 1.

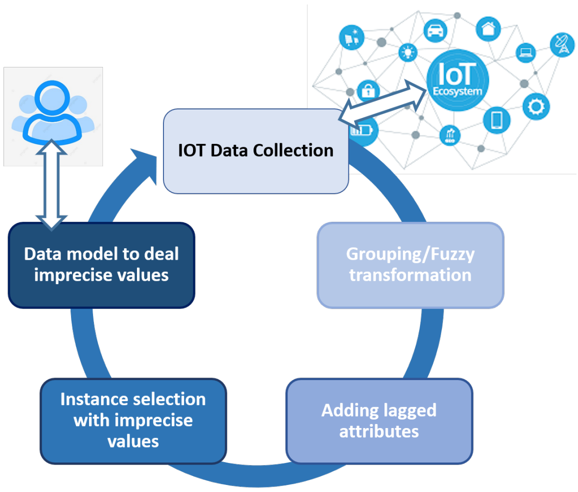

Figure 1: Steps sequencing of the proposed methodology.

The following subsections explain in detail the two phases of the proposed methodology.

3.1 Preparing the dataset

3.1.1 Instance grouping

In a first step, a certain number of consecutive instances is grouped and representative information about these grouping is added to the dataset such as the minimum, maximum, mean and standard deviation values, among others. This grouping produces an increase in the number of attributes describing each time series and this can lead to a considerable increase in problems that require long-term forecasting.

3.1.2 Interval/fuzzy transformation

The previous problem can be improved by replacing the set of attributes with crisp values associated to a time series with a single attribute whose values will be expressed by interval/fuzzy values and will collect the information contained in all the attributes it replaces. This allows grouping to be carried out without generating an increase in the number of attributes in the dataset. This improvement is illustrated in Figure 2.

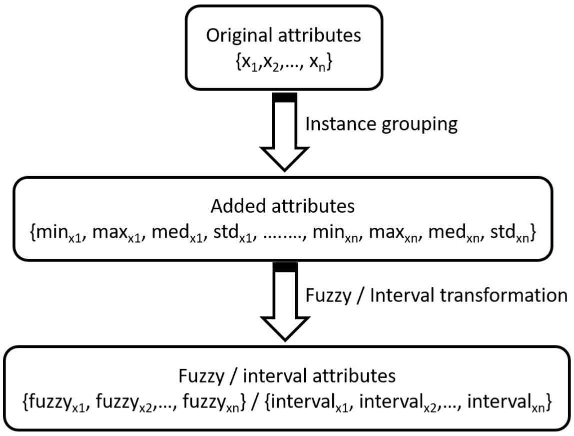

Figure 2: Methodology fuzzy/interval transformation.

3.1.3 Adding lagged attributes

After the previous transformations, the correlation between consecutive instances of the dataset is captured by adding lagged attributes. These added attributes allow us to capture a prediction horizon that will depend on the problem we are dealing with. So, we will group in the same instance as many lagged attributes as the size of the prediction horizon. For example, in a problem with hourly information and a prediction horizon of "*x*" hours, we will have to add "*x*" lagged attributes.

3.2 Instance selection and classification model

The next step in the proposed methodology is to apply an instance selection process. The instance selection technique must be a preprocessing technique which can deal with heterogeneous attributes (numeric and nominal) and with crisp and imprecise values.

Once the number of instances of the dataset has been decremented, the model that will be used to classify new instances is obtained. This is the model that will be available to the end user of the application to obtain answers to their requests. This model can be stored in the cloud or locally on the user's device. Again, the classification techniques that can be part of this methodology must allow working with heterogeneous attributes and crisp/imprecise values.

As we can see in Figure 1, the proposed methodology is part of a cyclical process that allows keeping the final model updated based on new data, which is of particular importance when the problem is placed in the framework of environments that evolve over time. This will make it possible to capture evolving trends. This update must be efficient to avoid high computational cost in the face of a high frequency of updates. Therefore, the choice of the final technique/model must be made with these requirements in mind.

4 Study case: early crop frost forecasting

The objective in this section is to apply the methodology presented in the previous section to the specific situation of early crop frost forecasting. The aim is to obtain as accurate prediction as possible of whether a frost will occur in the next hour and thus to be able of activating the anti-frost technique of sprinkler irrigation in time.

For this purpose, we obtain weather information collected by The Murcian Institute of Agricultural and Food Research and Development (IMIDA), [13], through weather stations located in different areas of the Murcia region (located in the southeast of Spain). Since these stations cover large areas of fruit tree crops, early frost detection aims to warn the farmer of possible frost on his/her fruit trees.

4.1 Instance Grouping

Each station is equipped with the following sensors and ephemeris: weather vane, radiometer, rain gauge, data-logger and thermo-hygrometer. The collected information corresponds to “Humidity”, “Radiation“, “Accumulated radiation”, “Wind speed”, “Wind direction”, “Rainfall”, “Vapor pressure deficit”, “Dew point” and “Temperature”. The IMIDA stores the information captured every 5 minutes and it is grouped 12 by 12 so that each instance reports the values of the sensors every hour. The granularity of this time grouping is due to the conditions of the frost prediction problem. It does not make sense to predict a frost within 5 minutes because in that case there is no time to activate the frost protection systems with any guarantee of success in order to save the crop. The obtained hourly values are summarised in Table 1 and constitute the starting time series datasets.

Type of information	
Min. relative humidity (%)	Mean relative humidity (%)
Max. relative humidity (%)	Mean radiation (W/m^2)
Accum. radiation (W/m^2)	Max. radiation (W/m^2)
Mean wind speed (m/s)	Max. wind speed (m/s)
Mean wind direction (^{o}C)	Rainfall (mm)
Vapor pressure deficit (kPa)	Dew point (^{o}C)
Min. temperature (^{o}C)	Mean temperature (^{o}C)
Max. temperature (^{o}C)	

Table 1: Information collected every hour for each station – IMIDA.

In these datasets, all instances with “Min. temperature”>7 are removed because they are not relevant to the problem we are modelling.

4.2 Transformation by Imprecise values

On these data, a fuzzy transformation that collects in a single fuzzy attribute the available information of each time series is performed. For example, the available information for the time series “relative humidity” in the attributes “Minimum relative humidity”, “Mean relative humidity” and “Maximum relative humidity” is collected in a single fuzzy attribute “RH$_f$” (Table 2). This way, from the 15 attributes shown in Table 1 we move to the 9 attributes shown in Table 2.

Attr	Description	Attr	Description
RH$_f$	Relative humidity	R$_f$	Radiation
AR	Accumulated radiation	WS$_f$	Wind speed
WD	Mean wind direction	RF	Rainfall
VPD	Vapor pressure deficit	DE	Dew point
T$_f$	Temperature		

Table 2: Transformed Attribute Description.

Each fuzzy attribute is represented by a trapezoidal fuzzy number $[v_1, v_2, v_3, v_4]$. For attributes with three measurement values (*min*, *med*, *max*) it is constructed as $v_2 = med - 5\%med$, $v_3 = med + 5\%med$, $v_1 = min$ and $v_4 = max$. For attributes with two measurement values (*med*, *max*) it is constructed as $v_2 = med - 5\%med$, $v_3 = med + 5\%med$, $v_1 = 2med - max$ and $v_4 = max$. In the latter case, if $v_1 \leq \text{min}_{\text{global}}$ then v_1=min$_{\text{global}}$. In order to model all available information, when a value is missing it is represented by the interval $[min_{\text{global}}, max_{\text{global}}]$.

4.3 Adding lagged attributes

In order to establish a prediction horizon that allows us to know the frost risk with the data obtained during the previous two hours, for each time series of the Table 2 a lagged attribute is added. This allows us to capture the data time dependence of two consecutive hours ($H-1$ and H).

In addition, and as intermediate step, this set of attributes is augmented to add the target attribute T$_{MIN}$ (“Min. Temperature”) of the instance of the time H+1 giving rise to a dataset where each instance is described with the set of attributes of the Table 3. From this dataset the attribute T$_{MIN}$ is replaced by what will be the attribute to predict and which will have two possible values FROST/NOFROST. The value of this attribute for each instance is obtained from the value of the attribute T_{MIN}. If $T_{MIN} > 0$ the instance is labelled with the value NOFROST and otherwise with FROST.

Hour $H-1$								
RH_f	R_f	AR	WS_f	WD	RF	VPD	DE	T_f
Hour H								
RH_f	R_f	AR	WS_f	WD	RF	VPD	DE	T_f
Hour $H+1$								
T_{MIN}								

Table 3: Constructed Instance to reflect the relationship between what happened in hour $H-1$ with hour H to predict the temperature in hour $H+1$.

The time series datasets obtained for each weather station and their description are shown in Table 4. Table 4 shows the acronym of each dataset, $|E|$ indicates the number of instances, Attr the number of attributes, nu and no indicate the number of numeric and nominal attributes respectively, I is the number of class values, %MV percentage of missing values, %FV percentage of fuzzy values and %IMFV indicates the percentage of instances with missing and/or fuzzy values.

Acron	$\|E\|$	Attr	Nu	No	I	%MV	%FV	%IMFV
CR12	15185	19	18	1	2	0.02	44.4	100
CR32	11855	19	18	1	2	0.01	44.4	100
CI52	9094	19	18	1	2	0.02	44.4	100
JU71	8470	19	18	1	2	0.01	44.4	100
JU81	8296	19	18	1	2	0.01	44.4	100

Table 4: Features of the different datasets for the different stations.

4.4 Instance selection and preliminary classification results

On the datasets described in Table 4, an instance selection technique capable of working with heterogeneous values that can be expressed by precise/imprecise values is applied. In this work, CNN_{imp} technique [6] is applied. This technique is an extended version of CNN technique able to work with this kind of data. The CNN_{imp} technique pre-orders the initial set of instances based on their imprecision, so that the instances with more imprecise values will be considered last by the technique and will be less likely to be part of the final training dataset. This avoids that instances with very abrupt changes of values within an hour (which may indicate temporal errors in the sensors) are included in the set of selected instances.

To perform the instance selection and obtain some preliminary classification results, each dataset of the Table 4 is divided into a train dataset with 80% of instances and a test dataset with 20%. The instance selection technique will be applied to the train dataset and the classification of the test dataset will be performed using the original and condensed train dataset. To perform the classification of the test dataset, the kNN_{imp} technique is used, [7].

Table 5 shows the accuracy obtained using the train dataset with all the instances (DS_O). It also shows the percentage reduction obtained with CNN_{imp} technique using the values k=5 and k=1 (%R_5, %R_1 respectively); and finally it shows the accuracy obtained when classifying the test dataset with the condensed datasets (DS_C). Subscripts in the accuracy values indicate the k value that is used in the classification. The averaged values of the different calculated measures are also shown (Av).

	Dataset (DS)					
	CR12	CR32	CI52	JU71	JU81	**Av**
$\mathbf{DS}_O$	95.49_7	95.28_7	96.37_7	96.22_7	97.11_7	**96.09**
%$\mathbf{R}_5$	72.95	73.00	80.00	74.69	83.56	**76.84**
$\mathbf{DS}_C$	95.39_7	95.23_7	96.48_7	96.28_7	96.93_7	**96.06**
%$\mathbf{R}_1$	87.08	87.45	80.00	88.20	92.00	**86.95**
$\mathbf{DS}_C$	95.13_{15}	95.02_{15}	96.48	96.22_{15}	96.68_{15}	**95.91**

Table 5: Accuracy with original/condensed datasets including fuzzy and missing values.

As can be seen in the results of Table 5, the reductions obtained with the CNN_{imp} technique are very relevant and the reduced datasets maintain the accuracy of the results (even improving slightly in some cases). This indicates that the relevant information contained in the data has been captured in the reduced datasets. Therefore, these preliminary results are promising.

As mentioned above, in this study case, the classification task for early frost forecasting has been carried out with the kNN_{imp} technique. The model therefore is formed by instances. The gradual update of the model to capture the evolution of the environment can be carried out by assigning each instance a weight (initially at value 1) that is decremented in relation to the passage of time and frequency with which that instance participates in the nearest neighbour set to rank new instances. In this way, instances that become obsolete due to the appearance of new trends can be eliminated, for example, in the analysed study case, temperature relationships could change due to climate change and global warming or simply due to the change of climatic seasons.

5 Towards an eco friendly App close to farmer

Farmers have been very reluctant to the technological transformations of the last decades. For farmers to benefit from IoT, they need to be able to do so using simple smart farming gadgets like small local weather station or a simple mobile phone to get the information from their field. A mobile phone App will make use of this information and the proposed methodology. The App provides, in a very simple way, information about possible frost situations allowing farmers to anticipate and take the most appropriate decisions to protect their crops. The App is integrated in an IoT system using the FIWARE infrastructure [22] (see Figure 3).

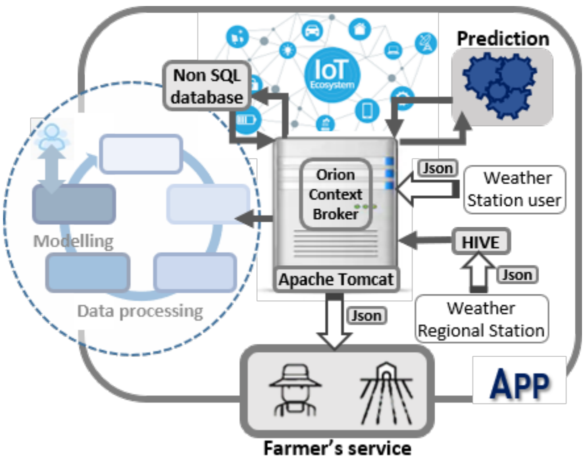

Figure 3: Preliminary design of the structure of an APP in an IoT system with FIWARE infrastructure.

In the IoT system, the proposed methodology constitutes the data analytics module. The FIWARE server is in charge of receiving the information from the sensors of the weather stations, storing this information and providing the appropriate web services that will be used by the user interface to visualise this information, as well as serving as a data source for the data analytics module. To interact with the App, data are collected through IoT devices (weather stations) located on the farmers' plots. This context information is stored in non-SQL databases with the help of the Orion Context Broker which provides the function to manage, query and update this information. The server integrates the HIVE infrastructure for data warehousing, built on Hadoop, to provide clustering, querying and data analysis on data available from weather stations at regional level. Json agents are used to communicate IoT devices using the Json protocol. This IoT agent translates the messages sent by the devices into the appropriate format.

6 Conclusions

Decision support systems in agriculture must be user-friendly and mobile-friendly applications that are easy and accessible to all farmers. Climate change and new technologies are creating a trend towards sustainable agriculture. This paper has focused on the problem of frost in agriculture by proposing a methodology to tend to the design of a farmer-friendly application that helps farmers to make decisions when activating anti-frost techniques as accurately as possible (sprinkler irrigation technique). By having greater precision, water resources are saved and both staff and resource costs are reduced, creating a more sustainable trend. This methodology integrates an instance selection model and a classifier that allow working with imprecise data and lagged attributes. The methodology is implemented in a real study case in the Region of Murcia, Spain. The techniques used for instance selection and classification are the CNNimp and KNNimp techniques capable of handling imprecise data. The results show an instance reduction of more than 75% and a classification success rate of 95%. These initial results show that a lightweight model with instance selection can be integrated into the decision support system and periodically updated. Farmers would then have an accessible decision support system application with very satisfactory results, thus moving their work towards a more environmentally and economically sustainable environment. The preliminary results have been satisfactory and will be extended and analysed for more areas and stations in the Region of Murcia in order to have a more robust system. It will also be generalised to any other region.

Acknowledgement

Supported by the projects TIN2017-86885-R (AEI/FEDER, UE), RTC-2017-6389-5 and RTC2019-007159-5, granted by the Ministry of Economy and Competitiveness of Spain (including FEDER support).

References

[1] S. Balit, M. Acunzo, A Changing World: FAO Efforts in Communication for Rural Development, in: J, Servaes (Eds.) Handbook of Communication for Development and Social Change, Springer, 2020, pp. 133–156.

[2] B. Basso, J. Antle, Digital agriculture to design sustainable agricultural systems, Nature Sustainability 3(4) (2020) 254–256.

[3] M. Bastan, R.R. Khorshid-Doust, S.D. Sisi, A. Ahmadvand, Sustainable development of agriculture: a system dynamics model, Kybernetes 47(1) (2018) 142–162.

[4] K. Buza, Time Series Classification and its Applications, in: Proc. 8th International Conference on Web Intelligence, Mining and Semantics, New York, USA, 2018, pp. 1—4.

[5] K. Buza, A. Nanopoulos, L. Schmidt-Thieme, INSIGHT: Efficient and Effective Instance Selection for Time-Series Classification, in: J.Z. Huang, L. Cao, J. Srivastava J. (Eds.), Advances in Knowledge Discovery and Data Mining. PAKDD 2011. Lecture Notes in Computer Science, Vol. 6635, 2011, pp. 149–160.

[6] J.M. Cadenas, M.C. Garrido, R. Martínez, A. Martínez, Regla k_M-vecinos más cercanos en Bases de Datos de Baja Calidad, in: Proc. VI Simposio de Teoría y Aplicaciones de Minería de Datos, Madrid, Spain, 2013, pp. 1303–1312.

[7] J.M. Cadenas, M.C. Garrido, R. Martinez, E. Muñoz, P.P. Bonissone, A fuzzy k-nearest neighbor classifier to deal with imperfect data, Soft Computing 22(10) (2018) 3313–3330.

[8] C.H. Chou, B.H. Kuo, F. Chang, The generalized condensed nearest neighbor rule as a data reduction method, in: Proc. 18th International Conference on Pattern Recognition Vol. 2, Hong Kong, China, 2006, pp. 556–559.

[9] T.M. Cover, P.E. Hart, Nearest neighbor pattern classification, IEEE Transactions on Information Theory 13(1) (1967) 21-–27.

[10] A. González-Vidal, F. Jiménez, A.F. Gómez-Skarmeta, A methodology for energy multivariate time series forecasting in smart buildings based on feature selection, Energy & Buildings 196 (2019) 71–82.

[11] FreshPlaza, Frosts in Murcia will reduce the supply of vegetables in the coming weeks, https://www.freshplaza.com/article/9281419/, February 2021.

[12] P. Hart, The condensed nearest neighbor rule (corresp.). IEEE Transactions on Information Theory 14(3) (1968) 515-–516.

[13] IMIDA, Murcia Agricultural Information System, Meteorological data for the Region of Murcia. http://siam.imida.es/.

[14] P. Koohafkan, M.A. Altieri, Globally important agricultural heritage systems: a legacy for the future. Food and Agriculture Organization of the United Nations, Rome, 2011.

[15] J. Lindblom, C. Lundström, M. Ljung, A. Jonsson, Promoting sustainable intensification in precision agriculture: review of decision support systems development and strategies, Precision Agriculture 18(3) (2017) 309–331.

[16] M.S. Mekala, P. Viswanathan, A Survey: Smart agriculture IoT with cloud computing, in: Proc. 2017 International Conference on Microelectronic Devices, Circuits and Systems, Vellore, India, 2017, pp. 1–7.

[17] J.A. Olvera-López, J.A. Carrasco-Ochoa, J.F. Martínez-Trinidad, J. Kittler, A review of instance selection methods, Artificial Intelligence Review 34(2) (2010) 133—143.

[18] G.L. Ritter, H.B. Woodruff, S.R. Lowry, T.L. Isenhour, An algorithm for a selective nearest neighbor decision rule. IEEE Transactions on Information Theory 21(6) (1975) 665-–669.

[19] S.L. Snyder, J.P. Melo-Abreu, Frost protection: Fundamentals, Practice and Economics Vol. 1, FAO, Roma, 2005.

[20] M.B. Stojanovic, M.M. Bozic, M.M. Stankovic, Z.P. Stajic A methodology for training set instance selection using mutual information in time series prediction, Neurocomputing 141 (2014) 236–245.

[21] X. Xi, E. Keogh, C. Shelto, L. Wei, C.A. Ratanamahatana, Fast Time Series Classification Using Numerosity Reduction, in: Proc. 23nd International Conference on Machine Learning, Pittsburgh, PA, USA, 2006, pp. 1033—1040.

[22] T. Zahariadis, A. Papadakis, F. Alvarez, J. Gonzalez, F. Lopez, F. Facca, Y. Al-Hazmi, FI-WARE lab: Managing resources and services in a cloud federation supporting future Internet applications, in: Proc. 7th International Conference on Utility and Cloud Computing, London, 2014, pp. 792–799.