Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 176 (2020) 300–309

Procedia Computer Science

www.elsevier.com/locate/procedia

24th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

RadRAR: A relational association rule mining approach for nowcasting based on predicting radar products' values

Gabriela Czibula, Andrei Mihai*, István Gergely Czibula

Department of Computer Science, Babeş-Bolyai University
1, M. Kogalniceanu Street, 400084, Cluj-Napoca, Romania

Abstract

The paper approaches the topic of *nowcasting*, one of the hottest topics in meteorology which deals with the problem of short-term forecasting of severe weather phenomena. Various types of meteorological data, including radar measurements, satellite data and weather stations' observations are currently used for forecasting severe weather events. Radar data is one of the important sources used by meteorologists for nowcasting and for providing alerts for severe weather events. We are proposing a new one-class classifier, named *RadRAR* (*Radar products' values prediction using Relational Association Rules*) for convective storms nowcasting based on radar data. More specifically, *RadRAR* is trained on radar data collected from normal weather conditions and learns to predict whether the radar echo values will be higher than 35dBZ, i.e. likely to indicate the occurrence of a storm. *RadRAR* is intended to be a proof a concept that relational association rule mining applied on radar data is helpful in discriminating between severe and normal weather conditions. Real radar data provided by the Romanian National Meteorological Administration is used for evaluating the effectiveness of *RadRAR*. A *Critical Success Index* of 61% was obtained, outperforming similar approaches from the literature and highlighting a good performance of *RadRAR*.

Keywords: Weather nowcasting; Data mining; Classification; Relational association rules
2000 MSC: 68T05; 68P15

1. Introduction

Nowadays there is a great opportunity to analyse very large volumes of meteorological data, such as radar measurements, satellite data and weather stations' observations. Applying *data mining* (DM) techniques for uncovering meaningful patterns in meteorological data would be effective for assisting meteorologists in their decision making processes and thus will contribute to optimise meteorological related tasks. The experience of operational meteorologists and their ability to analyse the available data and rapidly extrapolate from it is essential for most of their meteorological decision making processes, such as precipitation forecasting and issuing nowcasting alerts. DM based

* Corresponding author. Tel.: +4-264-405-327; fax: +4-264-591-906.
E-mail address: mihai.andrei@cs.ubbcluj.ro

1877-0509
10.1016/j.procs.2020.08.032

software solutions would be effective for accurately processing the huge volume of meteorological data and extracting relevant knowledge from it, thus serving to improve the efficiency of meteorologists' activity. One of the most challenging topics in meteorology is the forecasting of severe weather phenomena, as nowadays the number of such severe events is increasing in most regions of the world. Of particular interest is the short term weather forecast (up to 6 hours), known as *nowcasting* and having a significant role in risks prevention and crisis management. Despite its practical importance and relevance, the problem of weather nowcasting is difficult, due to the large volume of meteorological data which has to be analysed and which is highly dependent on various environmental conditions.

Radar data is one of the important sources used by meteorologists for nowcasting and for providing alerts for severe weather events. Particle *reflectivity* (R) and *velocity* (V) are the base radar products used for nowcasting. Values for particle reflectivity are mainly used by operational meteorologists for storms tracking. R values above a certain *threshold* (35 dBZ is generally used [8, 7]) are indications about potential moderate to heavy storms occurrences. The results of historical radar data analysis, as well as the estimation of radar products' values based on their historical values data would be important for the short term *quantitative precipitation estimation* (QPE) and *forecast* (QPF). QPE and QPF are relevant and complex activities in meteorology, as heavy rainfall is a common severe weather event which may cause damages and could affect the population safety. In spite of the great scientific and technological developments made over the past decades, a higher spatial and temporal accuracy of the precipitation forecasts is demanded by both meteorologists and society. Advances in high-performance computer systems led to the establishment of Numerical Weather Prediction (NWP) models [19] as the main technique for QPF, but difficulties in modelling cloud dynamics and microphysics still lead to significant errors in rainfall field forecasts.

DM has been employed in the literature for the prediction of meteorological phenomena or for meteorological data analysis. Bartok et al. [1] describe their research towards detecting and predicting meteorological phenomena such as fog and low cloud cover using various data sources including meteorological messages, satellite and radar images. An international project, "CrossGrid" intended to develop grid components for data intensive applications and one of the purposes was to apply these components for the analysis of databases of atmospheric circulation patterns [2]. Particularly, in the mentioned work, the authors present how DM techniques such as self-organizing maps and smoothing filters (Kalman) are employed to simplify the enormous quantities of data at their disposal to further allow other types of investigations. Kohain and El-Halees [3] applied ML and DM techniques for the analysis, classification of meteorological data and prediction of future meteorological conditions using a data set with data recorded from one weather station for nine years, for the Gaza Strip. Talib et al. [18] apply techniques such as clustering and decision trees for weather and climate change analysis using data collected over ten years at Faisalabad city, Pakistan and analyse a set of rules found within the data, which emphasize various correlations between weather conditions and the yearly time period.

Relational Association Rules (RARs) [15] extend the classical *association rules* by capturing relationships between values of attributes characterizing a data set. RARs mining have been applied in the literature for various data mining tasks, such as: software design defects detection [6], software defect prediction [13] or students performance prediction [5].

The contribution of the paper is twofold. First, we are investigating, as a proof of concept, the suitability of applying RAR mining for distinguishing between severe and normal weather conditions, with the aim of using these predictions for nowcasting. In addition, we aim to point out the relevance of the RARs mined from radar data, from a meteorological viewpoint. Thus, we are proposing a new one-class classifier, named *RadRAR* (*Radar products' values prediction using Relational Association Rules*) for convective storms nowcasting based on radar data. More specifically, *RadRAR* is trained on radar data collected from stormy weather conditions and will learn to predict weather the radar echo values will be higher than 35dBZ. Experiments conducted on real radar data provided by the Romanian National Meteorological Administration highlight that RARs are effective for estimating, based on historical data, if the radar echo values will be higher than 35dBZ (i.e. a storm is likely to happen). To the best of our knowledge, *RadRAR* is the first using RARs for nowcasting purposes.

The rest of the paper is organized as follows. We start by reviewing in Section 2 the recent machine learning based approaches used in nowcasting and by providing a brief background on RAR mining. The methodology applied for building *RadRAR* one-class classifier is introduced in Section 3. Section 4 presents the experimental results and their analysis, together with a comparison to related work. Section 5 concludes the paper and indicate directions to further improve and extend *RadRAR*.

2. Background

The section starts by presenting in Section 2.1 a literature review of recent approaches related to weather nowcasting based on radar data. Then, the fundamental concepts related to RAR mining are introduced in Section 2.2.

2.1. Literature review on machine learning based weather nowcasting

Various machine learning based solutions have been proposed in the literature for weather nowcasting. We are reviewing, in the following, several recent solutions for weather nowcasting applied on radar data. Han et al. [9] used Support Vector Machines (SVMs) for classifying whether or not a radar echo > 35 dBZ will appear on the radar within 30 minutes. The approach uses both temporal and spatial features, derived from vertical wind, and perturbation temperature. The obtained results were around 0.61 Probability Of Detection (POD), 0.52 False Alarm Ratio (FAR) and 0.36 Critical Success Index (CSI). Shi et al. [16] proposed an extension of the Long Short-Term Memory network (LSTM) model, named Convolutional Long Short-Term Memory network (ConvLSTM), appropriate for preserving spatiotemporal features due to the convolutional structure. Experiments were performed on the Radar Echo Dataset and the obtained results were compared with those obtained by the previous state of the art methods on the same data set. The proposed ConvLSTM model provided a Rainfall Mean Squared Error (an extension of the Mean Squared Error that converts from pixel-wise errors to radar echo intensities) value of 1.420.

Kim et al. [11] used a ConvLSTM-based model, named DeepRain, for precipitation prediction. Based on radar reflectivity data, DeepRain was used for predicting the amount of precipitation. The experimental evaluation of DeepRain highlighted that the Root Mean Squared Error (RMSE) is decreased with 23% compared to a linear regressor and that DeepRain is more performant then a fully connected LSTM. A RMSE value of 11.31 has been obtained on the test set. Sprenger et al. [17] used the AdaBoost algorithm for nowcasting prediction. The proposed learning model was trained with three years of hourly simulations data and achieved 0.88 sensitivity and 0.29 probability of false detection. Yan Ji [10] approached the problem of short-term precipitation prediction from radar observations using Artificial Neural Networks (ANNs) applied on radar raw data collected from China from 2010 to 2012. The reflectivity values were extracted from the raw radar and used for training the predictive model. A RMSE less than 5 was obtained, with a minimum of 0.97 and a maximum of 4.7.

For improving the performance of the learning, Tran and Song [20] proposed a new loss function for the Convolutional Neural Network (CNN) based models. The problem was approached from a computer vision perspective, using Image Quality Assessment Metrics as loss functions. The authors concluded that the Structural Similarity function performed better than Mean Squared Error (MSE) and Mean Absolute Error (MAE) and improved the quality of the output (predicted) images. The best performance was obtained by combining the Structural Similarity, MAE and MSE loss functions. A CNN model has been introduced by Han et al. [8] for convective storms nowcasting based on radar data. The model was designed for classification, more specifically for predicting whether radar echo values will be higher than 35dBZ in 30 minutes. In the proposed model, the output is a 3D-image, where each point of the image is 0 if radar echo is predicted to be $\leq$ 35dBZ in 30 minutes and 1 otherwise. A Critical Success Index (CSI) score of 0.44 was obtained.

2.2. Relational association rule mining

Association rules (ARs) represent powerful data analysis and mining tools useful for uncovering meaningful rule based patterns in various data sets. *Relational association rules* (RARs) [15] were introduced as an extension of classical ARs with the goal of expressing various types of relationships between data attributes. Considering that $A = (a_1, \ldots, a_m)$ represents the attribute set characterizing a data set D of instances and R is the set of possible binary relationships which may be defined between two attributes' domains, a RAR [15] has been introduced as a sequence of attributes from A and relations from R, $rule = (a_{i_1} r_1 a_{i_2} r_2 a_{i_3}, \ldots, r_{p-1} a_{i_p})$ such that between two consecutive attributes a_{i_j} and $a_{i_{j+1}}$ from *rule* the relationship r_j holds. The *support* of the *rule* was defined as the percentage of instances from D in which all the attributes from *rule* are non-missing, while its *confidence* was introduced as the percentage of instances from D in which all the relationships between the successive attributes from *rule* hold. The concept of *interestingness* [15] was associated to those rules having the *confidence* and *support* exceeding specified thresholds. For mining interesting RARs an Apriori-like algorithm called *DRAR* (Discovery of Relational Association Rules) [4] was introduced. .

For illustrating the concept of RARs mining, we consider a synthetically generated data set with ten instances and five integer valued attributes (i.e. $m = 8$). The data set is given in the left side table from Figure 1. We applied the *DRAR* mining algorithm with the following setting: the minimum support threshold of 1, the minimum confidence threshold of 0.9, $\leq$ and $\geq$ as binary relations between the attributes. The maximal interesting RARs mined from the sample data set are shown in the right side table from Figure 1.

No.	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
1.	8	8	9	4	5	10	7	8
2.	10	4	4	10	6	10	5	5
3.	7	9	4	10	10	3	5	3
4.	5	1	4	1	2	1	3	2
5.	10	10	6	3	3	10	4	3
6.	3	3	5	2	3	10	7	1
7.	10	8	8	2	3	4	8	7
8.	8	4	7	1	3	10	3	9
9.	5	1	1	1	2	1	3	1
10.	3	3	5	2	3	10	7	1

Length	**Rule**	**Confidence**
2	$a_1 \geq a_8$	0.9
2	$a_4 \leq a_5$	0.9
3	$a_5 \leq a_1 \geq a_2$	0.9
4	$a_6 \geq a_4 \leq a_1 \geq a_2$	0.9
4	$a_5 \leq a_1 \geq a_4 \leq a_6$	0.9

Fig. 1: Sample data set (left) and interesting maximal RARs mined for a minimum support threshold of 1 and a minimum confidence threshold of 0.9.

On each line from Table 1 a RAR of a certain length (i.e. the number of attributes contained by the rule) is depicted. The last column from Table 1 contains the confidence of the mined rules. For instance, the 3-length RAR depicted in the third line has the following interpretation: the value of the attribute a_5 is less than or equal to the value of a_1 which is greater than or equal to the value of the attribute a_2 in 90% of instances from the sample data, i.e. in 9 out of 10 instances.

3. Methodology

This section introduces the methodology we will use in building our *RadRAR* (*Radar products' values prediction using Relational Association Rules*) approach for convective storms nowcasting based on radar data. In proposing *RadRAR* one-class classifier we started from the result of our previous study [12] that similar values for a radar product in a certain location l at a given moment t are encoded in similar neighborhoods of l at time t-1. *RadRAR* is a one-class classifier which is trained on radar data collected from normal weather conditions and learns to predict if the radar echo values will be less than or equal to 35dBZ (i.e. very likely to indicate normal weather conditions). In our proposal, we started from the intuition that relationships between radar data attributes would be relevant for deciding weather a radar ata instance correspond to a normal or a severe weather event. We also note the generality of our RAR mining approach, as any type of binary relationships should be defined between data attributes, independent of the attributes' types.

We start by introducing in Section 3.1 the data model used in our approach. *RadRAR* one-classifier is then proposed in Section 3.2 which also details the evaluation methodology used for estimating the performance of *RadRAR*.

3.1. Data model

The radar data used in our experiments is provided by the WSR-98D weather radar [14] and is delivered to meteorological centers in the NEXRAD Level III format. The radar collects data with a frequency of about 6 minutes on a set of about 30 products, collected over 7 different elevations. The base products are particle *reflectivity* (R) and *velocity* (V) and is recorded for several elevation angles of the radar antenna, R01-R07 and V01-V07. The radar data is exported in the form of two-dimensional grids (matrices), with a point on the grid corresponding to a particular geographical location and storing the value of a radar product at a given time moment. Thus, for a given day, a sequence of matrices is available, each matrix storing the values for a certain meteorological product p (e.g. R01) for a time stamp t. In the current study we are focusing on a single meteorological product, namely $R01$. We decided to select R01, as it is one of the most relevant radar products used by operational meteorologists for issuing nowcasting warnings. Our previously performed statistical analysis of radar data [12] outlined that R01 is one of the most relevant data products correlated with the occurrence of severe meteorological phenomena.

In addition, we have shown in [12] that similar values for a radar product in a certain location l at a given moment t are encoded in similar neighborhoods of l at t-1. Accordingly, we assign to each location l from the analysed map (data grid) at timestamp t a high-dimensional vector whose elements are the values of R01 for the locations situated in a neighborhood of l at timestamp t-1. We are considering the neighborhood of l as the subgrid of diameter d centered on l. We note that the label of the d^2-dimensional instance previously described is the value of R01 for the geographical point l at timestamp t. For instance, let us consider a simplified example. Figure 2 depicts the radar data grid (containing the values for R01) at time stamp t and Figure 3 illustrates the radar data grid at time stamp t-1.

5	10	15	50	5
10	25	15	5	10
20	0	**15**	10	10
15	10	25	15	5
15	0	15	5	0

Fig. 2: The radar data grid at timestamp t.

20	30	10	15	10
15	25	10	20	5
20	10	**15**	20	25
10	10	15	10	20
15	5	10	15	20

Fig. 3: The radar data grid at timestamp t-1.

In Figure 2, in red, is the value for the R01 product at the location (3,3) at time t. In Figure 3, in blue are the values of R01 in the neighbourhood of the point (3,3) having a diameter of 2. For this example, the 9 dimensional vector associated to the data point (3,3) at time t is (25,10,20,10,15, 20,10,15,10) and is labeled as 15 (i.e. the value of R01 at location (3,3) and time t).

Considering the previously described vectorial representation for each geographic location at a certain time stamp, using a diameter of 13 for the neighborhood, a data set D is built from the 169-dimensional instances corresponding to each point from the data grid and all available time stamps. We note that a value of 13 has been selected for the diameter of the neighbourhood, since it represents about 5 kilometers in the physical world and this distance commonly determines small gradients of the meteorological parameters. As shown in Section 1, the reflectivity values higher than 35dBZ are indications for meteorologists for possible convective storms occurrences. Thus, we are diving the data set D in two classes of instances: the *positive* class (also denoted as "+") represents the instances labeled with R01 values higher than 35, while the *negative* class (denoted as "-") represents the instances labeled with R01 values less or equal to 35. The data set containing the positive instances is denoted as D_+, while by D_- we express the data set of negative instances. Obviously, D_- contains much more instances than D_+, as the number of locations from the considered geographical area in which severe storms occurred is significantly smaller than the number of locations without severe weather events.

3.2. RadRAR classifier

Let us consider the data sets D_+ and D_- built as shown in Section 3.1. We propose *RadRAR*, a one-class classifier which is trained on D_- (i.e. the radar data collected from normal weather conditions) and will learn to predict, based on the neighborhood of a certain location at time t, whether the radar echo value at time t+1 will be higher than 35dBZ. *RadRAR* is defined as a kind of anomaly detector, it is trained to recognize normal weather conditions and thus it will learn to predict if a certain instance correspond to stormy weather (i.e. is an anomaly with respect to the learned class D_-). The prediction is based on estimating the probability p_- that a certain 169-dimensional instance belongs to the "-" class and is computed considering the interesting RARs mined from D_- (the majority class).

As shown in Section 3.1, the instances from the data set D_- are 169-dimensional vectors. Accordingly, the set of *attributes* characterizing the instances is $A = \{a_1, a_2, \ldots, a_m\}$, where $m = 169$. The main idea behind our approach is briefly presented in the following. The classification process we propose takes place in two phases: *training* and *testing*. During the training, a classification model consisting of a set of interesting RARs from the set D_- (i.e. the set of instances representing normal weather conditions) will be built, and during testing, the model built during the training will be applied for deciding the class ("+" or "-") for a testing instance unseen during training.

RadRAR will be trained only on D_-, the subset of *negative* instances from D, representing the majority class. During training, the set of interesting RARs from the data set D_- are mined using the *DRAR* algorithm [4] and then, when a new instance s (169-dimensional vector represented as shown in Section 3.1) has to be classified (as "+" or "-"), the probability p_- expressing the likelihood that the query instance belongs to the *negative* class will be computed. Then the probability p_+ (the likelihood that the query instance belongs to the *positive* class) is calculated as $p_+ = 1 - p_-$.

Before training *RadRAR*, a **data preprocessing** step is first applied on the data set D_{-}. With the goal of enlarging the set of potential RARs which may be uncovered in the training data, the set of attributes A is extended with four additional attributes (expressing some meteorologically relevant thresholds for the value of R01): $a_{170} = 15$, $a_{171} = 25$, $a_{172} = 35$, $a_{173} = 40$ and $a_{174} = 50$. Since *DRAR* algorithm uncovers only binary relationships between attributes values, the newly added attributes will allow the discovery of relations between an attribute (R01) value and a threshold, as well. This way, more expressive and meaningful RARs will be detected.

RadRAR will be trained on D_{-}. During training, the set of interesting RARs from the data set D_{-} are unsupervisedly mined using the *DRAR* algorithm [4]. Let us denote by RAR_{-} the set of interesting RARs mined from D_{-}, with respect to the user-provided support and confidence thresholds. After *RadRAR* model has been built during the training step, it will be tested in order to evaluate its predictive performance.

3.2.1. Classification using RadRAR

At the classification stage, when a new query instance r has to be classified, we compute the probability $p_{-}(r)$ that the query instance belongs to the *negative* class. The underlying idea behind defining this probability is that the similarity of an instance r to the "-" class is very likely to be influenced by the number of RARs from RAR_{-} that are verified in the instance r.

First, we are computing μ and σ the mean and the standard deviation of the sequence $v_1, v_2, \ldots, v_n$, where n is the cardinality of D_{-} and v_i represents the number of rules from RAR_{-} which are verified in the i-th instance from D_{-}: $\mu = \frac{\sum_{i=1}^{n} v_i}{n}$, $\sigma = \sqrt{\frac{\sum_{i=1}^{n}(v_i - \mu)^2}{n}}$. The probability $p_{-}(r)$ for a testing instance r is computed as follows. Denoting by nr the number of rules from RAR_{-} which are verified in r, $p_{-}(r)$ is computed as shown in Formula 1.

$$p_{-}(r) = \begin{cases} 0.5 \cdot \frac{nr}{(\mu - 3\cdot\sigma)} & nr \in [0, \mu - 3\cdot\sigma) \\ 0.5 + \frac{nr - (\mu - 3\cdot\sigma)}{12\cdot\sigma} & nr \in [\mu - 3\cdot\sigma, \mu + 3\cdot\sigma] \\ 1 & otherwise \end{cases} \quad (1)$$

The intuition behind defining $p_{-}(r)$ is the so called *empirical rule* from statistics which state that about 99.73% of the values of a normally distributed variable lie within a band around the mean with a width of six standard deviations. Much more, the *three-sigma rule* states that even for variables which are non-normally distributed, at least 88.8% of values should lie within three-sigma intervals. Based on Formula 1, $p_{-}(r)$ will be 0.5 if $nr = \mu - 3\cdot\sigma$, less than 0.5 if $nr < \mu - 3\cdot\sigma$ and 1 if $nr \geq \mu + 3\cdot\sigma$. When instance r has to be classified, if $p_{-} \geq 0.5$, then instance r will be classified as a *negative* one, otherwise it will be classified as belonging to the *positive* class.

3.2.2. Testing

For evaluating the performance of the *RadRAR* model, it is tested on data sets containing both *positive* and *negative* instances which are completely disjoint from the training data set. For a testing data set, the *confusion matrix* consisting of four values is computed: True Positives – TP (instances predicted correctly as positive), True Negatives – TN (instances predicted correctly as negative), False Positives – FP (instances predicted **in**correctly as positive) and False Negatives – FN (instances predicted **in**correctly as negative). As evaluation measures, we are using four measures computed based on the values from the confusion matrix, used in supervised learning for assessing the performance of binary classifiers: *sensitivity* or *probability of detection* (**POD** = $\frac{TP}{TP+FN}$), *specificity* or true negative rate (**Spec** = $\frac{TN}{TN+FP}$), *false alarm rate* (**FAR** = $\frac{FP}{TP+FP}$) , overall *F-score* = $\frac{F_{+}+F_{-}}{2}$ (where $F_{+} = \frac{2TP}{2TP+FN+FP}$ and $F_{-} = \frac{2TN}{2TN+FN+FP}$) and *Area Under the ROC Curve* (**AUC** = $\frac{\textbf{POD}+\textbf{Spec}}{2}$) Additionally, we also consider the *Critical success index* (**CSI**) measure which is usually used for convective storms nowcasting based on radar data. CSI is computed as CSI=$\frac{TP}{TP+FN+FP}$. All the previously mentioned evaluation measures range in [0, 1]. Excepting FAR which has to be minimized, higher values for all other evaluation measures indicate better classifiers.

We note that for the weather nowcasting problem the main goal is to increase the *true positive rate* (*sensitivity*), while decreasing the *false alarm rate* (FAR). Overall, higher values for $\frac{POD}{FAR}$ indicate better classifiers.

4. Results and discussion

This section presents the experimental results obtained by evaluating *RadRAR* classifier, as well as a discussion on the obtained results and a comparison to related work.

4.1. Data sets

The case study used in our experiments is the radar data provided by the WSR-98D weather radar [14] located in Bobohalma, Romania, for the 5th of June 2017, a day with moderate atmospheric instability manifested through thunderstorms accompanied by heavy rain and medium-size hail. In the area from the central Transylvania region having the geographical coordinates (46.076N, 46.725N, 23.540E and 25.064E) there were two distinct episodes with intense meteorological events in June 5, 2017: the first one between approximately 09:00 and 11:00 UTC, and the second one between approximately 12:00 and 17:00 UTC, with the most severe events taking place between 14:00 and 15:00 UTC. In this paper, we restrict ourselves to a small size experiment, as our aim is to establish a proof of concept for the relevance of using RARs for nowcasting based on radar data. If this stands, the experiments will be further extended on a larger scale. The data used for training *RadRAR* is collected at approximately 14:37 UTC (in the middle of the severe event). The data sets D_+ and D_- collected as shown in Section 3.1 from the raw radar data consist of 1321 and 19991 instances, respectively. The imbalancement of D_+ and D_- is due to the fact that the number of geographical locations where the severe phenomenon occurred on 5-th of June, 2017, is much smaller than the number of locations with normal weather conditions.

4.2. RAR mining

After the data set D_- was obtained, as described in Section 3.2, a set of relations R between the integer valued features is established. In our experiments, two possible relations between the features' values are considered in the mining process: $\mathcal{R} = \{\leq, \geq\}$. After the relations were defined, the set RAR_- of interesting relational association rules of any length, having a minimum support 1 and various confidence thresholds (higher than 0.95) were discovered from D_-. A value of 1 was used for the minimum support threshold, as the data set does not contain instances with missing attribute values. For a better interpretation of the results, the *DRAR* mining algorithm has been also applied on the data set D_+ of non-zero labeled instances. The same minimum support of 1 was used for computing the set RAR_+, but considering minimum confidence thresholds higher than 0.89, as the number of uncovered interesting in D_- is much larger than the number of RARs discovered in D_+. Table 1 depicts the number of interesting RARs of a certain length (1-5) mined from the data sets D_+ and D_- for different confidence thresholds.

One observes from Table 1 that a significantly higher number of interesting rules are mined from D_- compared to the number of rules mined from D_+, for the same confidence threshold. For instance: (1) about 9350 interesting rules are uncovered in D_+ for a confidence threshold of 0.894 and in D_- for a confidence threshold of 0.991; (2) about the same number of interesting RARs (1750) are mined from D_+ for a confidence threshold of 0.95 and in D_- for a confidence threshold of 0.99525.

Analysing the sets RAR_+ and RAR_- mined from the data sets D_+ and D_- respectively, we observed distinct rules that characterize weather conditions on certain geographical locations and thus may be useful for discriminating between a normal and severe weather event. Figure 4 depicts such RARs extracted from the set RAR_+ (left table) and from the set RAR_- (right table).

Length	Rule	Confidence
4	$25 \leq a_{85} \geq 15 \leq a_{86}$	0.911
4	$25 \leq a_{85} \geq 15 \leq a_{98}$	0.91
4	$25 \leq a_{86} \geq 15 \leq a_{98}$	0.895
4	$25 \leq a_{97} \geq 15 \leq a_{98}$	0.92
5	$a_1 \leq 50 \geq a_3 \leq a_{85} \geq a_4$	0.896
5	$50 \geq a_4 \leq a_{85} \leq 15 \leq a_{98}$	0.894
5	$50 \geq a_5 \leq a_{85} \leq 15 \leq a_{98}$	0.895

Length	Rule	Confidence
5	$a_{72} \leq 35 \geq a_{84} \leq 25 \geq a_{85}$	0.971
5	$a_{72} \leq 35 \geq a_{97} \leq 25 \geq a_{98}$	0.972
5	$a_{84} \leq 35 \geq a_{85} \leq 25 \geq a_{86}$	0.971
5	$a_{84} \leq 35 \geq a_{98} \leq 25 \geq a_{99}$	0.971
5	$a_{85} \leq 35 \geq a_{97} \leq 25 \geq a_{98}$	0.973
5	$a_{85} \leq 35 \geq a_{98} \leq 25 \geq a_{99}$	0.971
5	$a_{86} \leq 35 \geq a_{97} \leq 25 \geq a_{98}$	0.972

Fig. 4: Interesting RARs of a certain length discovered in D_+ (left) and in D_- (right)

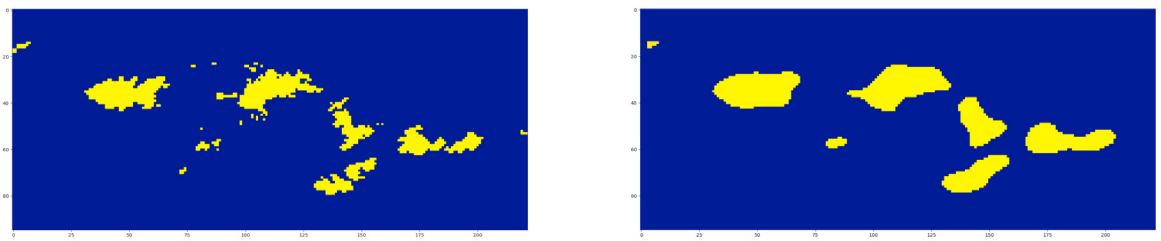
Table 1: The number of interesting RARs of a certain length (1-5) mined from D_+ and D_- for various minimum confidence thresholds.

Data set	Set of RARs	Minimum confidence	Number of interesting RARs				
			2-length	3-length	4-length	5-length	Total
D_+	RAR_+	0.95	96	1644	0	0	1740
		0.93	140	3698	0	0	3203
		0.91	192	6308	25	0	6525
		0.9	241	7871	106	6	8224
		0.894	280	8829	234	20	9363
D_-	RAR_-	0.99525	102	1673	0	0	1775
		0.992	169	7181	0	0	7350
		0.991	169	9181	0	0	9350
		0.99	169	11039	0	0	11208
		0.98	170	14197	8	0	14375
		0.97	199	14233	1494	121	16047

4.3. Results and comparison to related work

After the *RadRAR* classifier has been built applying the methodology described in Section 3, it is tested on data unseen during training, using the evaluation measures presented in Section 3.2.2. The testing is applied on eight time stamps between 13:39 UTC and 14:29 UTC and were selected around the beginning of the severe event. In all the testing sets we used both the positive and negative instances collected from the same geographic area as in the training set. Figure 5 shows the results of the predictions of *RadRAR* (using the best confidence threshold) compared to the real data, considering the time stamp 14:17 UTC. In bright yellow are depicted the cells where the values of the R01 product were greater than 35 DBZ (the positive class) and in dark blue the cells where the values of R01 were lesser or equal to 35DBZ (the negative class). On the left side is depicted the ground truth – the observations made by the radar; while on the right side are the predictions – the assigned class for each cell by *RadRAR*. It can be seen that the area covered by the yellow (representing the positive class) is noticeably smaller than the area in deep blue (the negative class), showing the significant imbalancement of the two classes.

If we compare the two images it can be observed that the prediction largely matches the real data. The yellow spots from the image with real data can also be found in the predicted image, having roughly the same shapes and sizes, showing intuitively that the prediction is relevant and in accordance with reality. The greatest flaw in the prediction is that it seems heavily smoothed. Meaning there will be false positives in places where there should be blue, but the shape is smoothed out and it was filled with yellow and, on the other hand, in places where there should be only few positives (small spots or points) will be smoothed out, resulting in false negatives.



(a) Real data – yellow represents cells where R01 product had values greater than 35 DBZ at 14:17 UTC.

(b) Predicted data – yellow represents cells where R01 product was predicted to have values greater than 35 DBZ at 14:17 UTC.

Fig. 5: Comparison of the predicted data with the real data collected by the radar for the product R01 at 14:17 UTC.

The left side table from Figure 6 summarizes the results obtained by applying *RadRAR* on the time stamp 14:17 UTC, using various confidence thresholds. The right side table presents the comparison between *RadRAR* (the best values obtained for a confidence threshold of 0.995) and the related work. The results reported for *RadRAR* in the

right side table from Figure 6 are averaged over the eight runs (time stamps used for testing) and a 95% confidence interval is given for the results. The best values for the evaluation measures are highlighted. As expected, the best values for the evaluation measures were obtained for higher confidence thresholds.

As shown in Section 2.1, there are numerous approaches developed in the nowcasting literature using machine learning techniques. From the existing related work, we found three approaches having similar goal to ours, that of predicting the values for the reflectivity (R) based on historical radar data. The approaches from the literature which are the most similar to ours are those proposed by Yan Ji [10] and Han et al. [9, 8]. Even if both the data sets and the testing methodology used in the previously mentioned papers differs from ours, we computed the evaluation measures reported in literature, trying to provide a comparison as accurate as possible. Still, it is difficult to provide a precise comparison, as: (1) the approaches from the literature use other data sets which are not publicly available in order to exactly reproduce the performed experiments; (2) the machine learning models used [9, 8] cannot be accurately reproduced due to the lack of information. Thus, we are going to compare *RadRAR* with the existing approaches considering only the magnitude of the evaluation measures (CSI, POD, FAR).

For better highlighting the effectiveness of *RadRAR* as an anomaly detector, we replaced it with an *autoencoder* (AE). The AE had been built using the Keras framework in Python, with Tensorflow backend; it had an encoding layer of size 25 (14.79% of input size) and employing 5 hidden layers before and after the encoding layer, having proportionally decreasing and, respectively, increasing sizes; the activation function used was ReLU for all layers with the exception of the encoding and output layers, where a linear activation function was used; the loss function used was Mean Squared Error and the optimizer was Adam. The training methodology applied for *RadRAR* was applied for AE as well, i.e the AE was trained on D_{-}. At the testing stage, Formula (1) was used for computing p_{-}, with the following modifications: (1) the v_i value (Section 3.2.1) now represents the loss value of the AE for the i-th training instance from D_{-}; and (2) nr is the loss value provided by the AE for the query instance r. The testing methodology described in Section 3.2.2 was applied for the AE experiment, as well, but with a slight modification: the threshold $\mu + 3 \cdot \sigma$ used in Formula (1) was reduced to $\mu + 0.75 \cdot \sigma$ in order to minimize the number of false negatives (otherwise, the AE would have been biased to predict the negative class). The results obtained by applying the AE on our data set are depicted on the last line from the right side table from Figure 1.

Confidence threshold	CSI	POD	FAR	Spec	*F-score*	AUC
0.95	0.52	0.67	0.31	0.981	0.83	0.83
0.97	0.57	0.68	0.21	0.987	0.86	0.83
0.98	0.56	0.67	0.23	0.986	0.85	0.83
0.99	0.59	0.67	0.18	0.989	0.86	0.83
0.995	**0.61**	**0.68**	**0.13**	**0.991**	**0.87**	**0.83**

Model	CSI	POD	FAR	POD/FAR
Our *RadRAR*	**0.56 ± 0.02**	**0.63 ± 0.01**	**0.17 ± 0.03**	**3.91 ± 0.78**
CNN [8]	0.44	–	–	–
SVM [9]	0.36	0.61	0.52	1.17
AE	0.21	**0.77**	0.78	0.99

Fig. 6: Experimental results obtained by *RadRAR* (left side table) and comparison to related work (right side table). A 95% confidence interval is given for the results of *RadRAR* reported in the right side table.

The comparison to related approaches depicted in the right side table from Figure 6 shows that our *RadRAR* proposal provides better results for the evaluation measures in 8 out 9 comparisons. We note than only the AE model slightly outperforms *RadRAR* in terms of **POD**, but its false alarm rate is very high, highlighting that AEs are not able to differentiate well enough between the positive and negative radar instances. Unlike the AE model, our *RadRAR* has a better balance between POD and FAR, providing a much higher value for $\frac{POD}{FAR}$ which denotes that it is able to better distinguish between the normal and severe weather conditions.

Analyzing the experimental results and the comparison to existing approaches from Figure 6 we may conclude that the RARs uncovered within the radar data are effective for predicting if the radar echo values are higher than 35dBZ, obtaining performances which are generally better than the results from the literature [8, 9].

While we tested *RadRAR* solely against radar data, *RadRAR* classifier is general and easily adaptable to different types of data. For example, satellite imaging data would be easy to model for use with *RadRAR*: each matrix would be an image from the satellite and instead of products there would be channels. Or, for using geographical data, adjacent geographical points can be arranged into a matrix and each matrix would contain data such as altitude or soil type. This 3D matrix model that *RadRAR* works with is very versatile and we would expect that the classifier to be applicable for different types of data.

5. Conclusions and further work

As a proof of concept, we introduced in this paper a novel one-class classification model *RadRAR* based on uncovering interesting *relational association rules* for estimating if the radar echo values will be higher than 35dBZ. Thus, based on the predicted values, the approach is useful for discriminating between normal and stormy weather conditions. Real radar data provided by the Romanian National Meteorological Administration have been used for assessing the performance of *RadRAR*. Future work will be performed for extending the experimental evaluation of *RadRAR* by incorporating more data various types of data, such as geographical features, and extending the number of radar products used in mining. In addition, we aim to generalize *RadRAR* towards a binary classifier, by mining the interesting RARs not only for the *negative* class (radar products' values below 35dBZ) but also for the positive class (radar products' values greater than 35dBZ). Alternative measures for defining the probability that a radar echo belongs to the *positive* class are also envisaged.

Acknowledgements

The authors acknowledge the assistance received from the NMA from Romania, for providing the meteorological data sets used in the experiments. Special tanks are due to Eugen Mihuleţ for the assistance and feedback.

References

[1] Bartok, J., Habala, O., Bednar, P., Gazak, M., Hluchy, L., 2010. Data mining and integration for predicting significant meteorological phenomena. Procedia Computer Science 1, 37 – 46. ICCS 2010.

[2] Cofino, A., Gutierez, J., Jakubik, B., Melonek, M., . Implementation of data mining techniques for meteorological applications. pp. 165–175.

[3] Cofino, A., Gutierrez, J., Jakubiak, B., Melonek, M., 2011. Implementation of data mining techniques for meteorological data analysis. International Journal of Information and Communication Technology Research 1, 96 – 100.

[4] Czibula, G., Bocicor, M.I., Czibula, I.G., 2012. Promoter sequences prediction using relational association rule mining. Evolutionary Bioinformatics 8, 181–196.

[5] Czibula, G., Mihai, A., Crivei, L., 2019. A novel relational association rule mining classification model applied for academic performance prediction, pp. 20–29.

[6] Czibula, G., Zsuzsanna, M., Czibula, I.G., 2015. Detecting software design defects using relational association rule mining. Knowl. Inf. Syst. 42, 545–577. URL: https://doi.org/10.1007/s10115-013-0721-z, doi:10.1007/s10115-013-0721-z.

[7] Dixon, M., Wiener, G., 1993. TITAN: Thunderstorm Identification, Tracking, Analysis, and Nowcasting - A radar-based methodology. Journal of Atmospheric and Oceanic Technology 10, 785–797.

[8] Han, L., Sun, J., Zhang, W., 2019. Convolutional Neural Network for Convective Storm Nowcasting Using 3D Doppler Weather Radar Data. arXiv e-prints , arXiv:1911.06185arXiv:1911.06185.

[9] Han, L., Sun, J., Zhang, W., Xiu, Y., Feng, H., Lin, Y., 2017. A machine learning nowcasting method based on real-time reanalysis data. Journal of Geophysical Research: Atmospheres 122, 4038–4051.

[10] Ji, Y., 2017. Short-term precipitation prediction based on a neural network, in: 3rd International Conference on Artificial Intelligence and Industrial Engineering, pp. 246–251.

[11] Kim, S., Hong, S., Joh, M., Song, S., 2017. Deeprain: Convlstm network for precipitation prediction using multichannel radar data. 7th International Workshop on Climate Informatics abs/1711.02316. URL: http://arxiv.org/abs/1711.02316, arXiv:1711.02316.

[12] Mihai, A., Czibula, G., Mihulet, E., 2019. Analyzing meteorological data using unsupervised learning techniques, in: ICCP 2019: IEEE 15th International Conference on Intelligent Computer Communication and Processing, IEEE Computer Society. pp. 1–8.

[13] Miholca, D., Czibula, G., Czibula, I.G., 2018. A novel approach for software defect prediction through hybridizing gradual relational association rules with artificial neural networks. Inf. Sci. 441, 152–170.

[14] NOAA's National Weather Service. Radar Operations Center, 2018. NEXRAD Technical Information. https://www.roc.noaa.gov/WSR88D/Engineering/NEXRADTechInfo.aspx.

[15] Serban, G., Campan, A., Czibula, I.G., 2006. A programming interface for finding relational association rules. International Journal of Computers, Communications & Control I, 439–444.

[16] Shi, X., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.k., Woo, W.c., 2015. Convolutional lstm network: A machine learning approach for precipitation nowcasting, in: Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1, MIT Press, Cambridge, MA, USA. pp. 802–810.

[17] Sprenger, M., Schemm, S., Oechslin, R., Jenkner, J., 2017. Nowcasting foehn wind events using the adaboost machine learning algorithm. Weather and Forecasting 32, 1079–1099.

[18] Talib, M.R., Ullah, T., Sarwar, M.U., Hanif, M.K., Ayub, N., 2017. Application of data mining techniques in weather data analysis. Int. Journal of Computer Science and Netw. Security 17, 22 – 28.

[19] Tan, C., Feng, X., Long, J., Geng, L., 2018. FORECAST-CLSTM: A New Convolutional LSTM Network for Cloudage Nowcasting, in: 2018 IEEE Visual Communications and Image Processing (VCIP), pp. 1–4.

[20] Tran, Q.K., Song, S.k., 2019. Computer vision in precipitation nowcasting: Applying image quality assessment metrics for training deep neural networks. Atmosphere 10, 244. URL: http://dx.doi.org/10.3390/atmos10050244, doi:10.3390/atmos10050244.