Using Case-based Reasoning and Artificial Neural Networks for the Efficient Prediction of Dust Storms

Tariq Saad Al Murayziq, Stelios Kapetanakis, Miltos Petridis

School of Computing, Engineering and Mathematics, University of Brighton, Moulsecoomb Campus, Lewes road, Brighton BN2 4GJ, UK
(tsa10, s.kapetanakis, m.petridis}@brighton.ac.uk

Abstract. A dust storm is a meteorological phenomenon that commonly occurs in desert regions where strong winds blow away the loose sand and dirt from dry surfaces. That could lead to transported soil from one place to another, through suspension and saltation. In this paper, we propose how past dust storm events could assist in forecasting of new dust storms as well as potential actions to mitigate their potential impact. To illustrate this concept a hybrid approach is proposed using a combination of case-based reasoning and artificial neural networks which indirectly estimate the similarity among past storm cases with any new ones. Our approach reveals that case-based reasoning may have a solid potential of being used in successful prediction and possible mitigation of extreme weather forecasting.

Keywords: Case-based Reasoning, Artificial Neural Network, Dust storm prediction, Weather forecast prediction, Artificial intelligence.

1 Introduction

Poor management and preservation measures of dry lands across the world seem to be the major causes of dust storms according to a number of researchers. For instance, abandonment of a fallow system can increase the frequency and size of a dust storm, thereby changing both local and global climate, which in turn may have great impact in agriculture and the overall global economy [19]. In addition, dusty weather can cause societal implications and changes in lifestyle, since by staying at home or generally indoors would be less likely to be at risk. Lack of efficient dust storm prediction at their early stages and preparing a set of actions to reduce their impact currently has severe consequences: it is recorded that the last dust storm that hit one of the major cities in Saudi Arabia in 2015 caused a number of accidents, had several flights cancelled and forced most of the public services and schools to close, due to the severeness of the storm [15].

Dust storms have been associated with the spread of diseases across the world since virus spores from the ground are blown up into the atmosphere and they then descent with acid rain or smog [21]. Dust can also cause impairment in the transportation and settlement centres of people living in the affected environments. As such, dust storm prediction helps groups and regions to prepare for their occurrence through prevention measures such as taking cover, vacating streets to avoid accidents, sealing windows and doors, and providing security to outdoor property, such as equipment and vehicles. It is also helpful in rural areas since farmers can have an early harvest, store any farm equipment and secure livestock. In this paper we will demonstrate a new approach to predicting dust storms, through focusing on case-based reasoning (CBR) and artificial neural network (ANN) techniques.

1.1 Motivation

Over the past few years, several dust storms have been recorded in a number of cities across the globe and their impact has been severe as mentioned above, like societal risks, health incidents and damage to crops, agriculture and landscapes. One of the countries that are frequently exposed to such extreme weather conditions throughout the year is Saudi Arabia. Therefore, in this study we propose to use a combination of artificial intelligence techniques (CBR and ANN) to identify dust storms, forecast the potential severity of dust events and formulate mitigation strategies for when they happen.

1.2 Rationale for using CBR and Artificial Neural Networks

The rationale behind choosing case-based reasoning to predict dust storms, is based on the nature of the application area. CBR seems appropriate to predict dust storms based on past storm events, since the fundamental hypothesis of CBR is that similar problems have similar solutions [11]. Additionally, case-based reasoning systems are able to make predictions along with indications of potential success of a chosen solution. This is done by measuring the level of similarity with evidence from past problems. When the solution is carried forward to a new case, the system is in a position to tell whether or not such approach will be successful [16]. One of the ways the system does this is by measuring the magnitude of the current query and weighing it against the past problem while measuring the effectiveness of the solution in the past problem by utilising any stored information.

Secondly, CBR is more flexible when the prediction system is updated by adding a number of past cases and use them in order to produce better explanations for any potential decision options [18]. CBR can deal with the uncertainty and the presence of incomplete information (e.g. cases with missing values or overwhelming number of features) by doing informed guessing on top of the closest neighbours of an investigated case. However, when the system is well integrated, especially when two techniques are combined, for example CBR is combined with ANN methods, they might produce more accurate and reliable results, because the limiting features of each technique can be mitigated upon integration [10]. The combination of ANN and CBR in this domain is in the early stages, but the advantages of this combination have been observed in different projects, such as fire alarms in early warning systems [10]. Moreover, the reason behind such integration is to mitigate the disadvantages of each technique. It is argued that the hybrid CBR-ANN shows better results than other combination methods in early warning systems, and the initial investigations into these combinations produce acceptable results [10]. In addition, research has claimed that combining two or more

AI methods, could raise the accuracy and the efficiency of a system, and could work better than using only one of the methods [5]. ANN will be applied in this project to recognise new dust storms from sensor data and enhance the CBR dataset with them.

2 Relevant Research work

Among the most successful techniques for predicting dust storms is the use of ANN models that use artificial intelligence for accurate forecasting. This method has been used to predict dust storms in various regions, and an example is Sistan in Southern Iran. Over the last few decades severe droughts in the Sistan region have reduced its vegetation cover, thereby decreasing the threat of dust storm occurrences. There are various studies by other scholars who have done research in terms of the importance of responsive systems [3]. This covers all the aspects of dust storm determination, including the effects on human life and the need for the best storm detection systems. Specifically, Chacon-Murgua et al. [3] proposed an approach for devising a dust storm detection system that is based on ANNs. Additionally, methods of managing ambiguous and undefined ANN outputs were presented in order to reduce the rate of false results. As a result, the researchers provide a well-formulated presentation of the capability and suitability of the proposed systems in detecting dust storm events [3]. Chacon-Murgua et al. claimed that a system using an ANN model could help to reduce the computational burden, as it removes cloud regions from the multispectral images received [3, 22]. Zakermoshfegh et al. [23] have shown encouraging results in using ANN to build a flexible model for predicting river flow [8]. It is stated that the ANN technique could an appropriate method to solve problems in water resources, since it has the ability to identify any non-linear relationships between the input and output using a cascade-correlation algorithm, which helps to understand better this natural phenomenon [6]. ANN techniques could be a useful technique in creating predicting tools to solve water problems and the investigations and analysis have proved this [23].

Currently, there is a significant body of research on the integration of artificial intelligence techniques in order to achieve better performance(s) in predicting systems. The application of ANN techniques has yielded encouraging results in prediction and it has the ability to learn and adapt its behaviour based on new cases. Additionally, ANN can be used for the “cleaning” of data. For example, if the input has missing values, ANN could estimate them by looking at the closest input cases to find the missing one [1].

CBR on the other hand can record successful attempts in building a predicting system. It is argued that CBR has the advantage of identifying and predicting successfully potential intrusion attempts of cyber attackers. The authors supported their claims with positive results [9]. Moreover, it is stated that case-based reasoning can be important in detection systems because it is a methodology that solves problems by utilising previous experiences, intrusions in case of [9], that may have occurred in the past in certain organisations [13]. The results from past research [9, 13] have strengthened the argument that Case-Based Reasoning may be more effective if combined with other methods. This could involve using other methods as support systems, integration for quicker problem solving, and utilising Case-Based Reasoning systems in supporting roles for other systems. An example is Case-Based design systems, which can be used as support and reasoning maintenance systems [17].

Aamodt and Plaza [2] proposed the steps for conducting Case-Based Reasoning in four steps. The first step, which is the retrieval step, aims at finding the most similar cases from a case base by using a weighted sum of properties and returning all similar cases which fit within certain agreeable parameters [14]. The re-use step involves adapting the retrieved solutions to an investigated case by using two methods, the transformational and the derivational method. The transformational method involves modifying the former solutions by using transformational operators that are specific to a certain domain, while the derivational method involves reusing the algorithms to generate the original solution so as to come up with a fresh one to fit the investigated case-problem.

The revision step involves evaluating the adapted solution, and if it is unsuccessful, repairing it by applying knowledge that is domain-specific. Finally, the solution is retained, and the new experience is stored as the new case by selecting the information to be stored and creating a new entry into the system's memory for future reference to emerging similar cases [12, 14].

Rule-Based Systems techniques have also shown ability to solve problems by deriving from expert knowledge. Conditions and actions are fed to the engine, which uses a working memory of information regarding the problem. The system adds actions to the rule and by applying and using a match-select-act cycle, the working memory and the knowledge base is repeated until no more relevant rules are found [4]. This technique could potentially be utilised in generating sets of affective actions and rules, to reduce the harm of the dust storms.

3 Proposed Research Methodology

We argue that CBR has the ability to predict sand storms; therefore, we propose to use a combination of CBR and ANN in order to obtain more accurate predictions of dust events and take the necessary actions to reduce their impact. ANN will be applied to obtain unclassified dust storm events from a satellite station and identify alert pattern in them. Any identified patterns will be stored in a CBR database, whereas CBR will be used to match these events with past knowledge to predict a dust storm in the near future. We may only need spectral or spatial data to improve the existing solutions. CBR is proposed since it seems the best approach since it could explain any case domain knowledge to civil engineers to engage with the best-worked solutions in order to improve their operational efficiency.

Our suggested approach aims to assist meteorological experts to determine any trends and the capability of a specific solution in order to improve its capacity for dust storm detection. Improving the solutions for an early forecast could help the public authorities to plan for economic activities, such as farming, in order to avoid financial losses when the event occurs unexpectedly. Working with analysed data helps experts to come up with a logical and quickly evaluated systematic solution. It is stated that variables are analysed to compute dust storm forecasts to include pressure gradients, temperature, humidity, sun radiation and soil moisture, among other elements [20].

In our proposed research reliance resides on both quantitative and qualitative methods. Quantitative methods will be used to measure any similarity using CBR, and to identify an unclassified dust storm from the radar station using ANN. In other words, CBR will be used to cleanse the provided dataset with past cases, sort all the dust storm events and apply filtering to reorder these events based on how similar the events are to each other.

The ANN technique will be applied to record new dust event patterns that will not exist in the provided dataset it the experts' station. Our proposed system aims to utilise this newly provided information with a classification using the CBR process upon any existing data. The qualitative method will be used to develop the actions that should be taken to avoid the impacts of dust storms. Simply, after the new dust storms are matched with old storms cases based on any similarities, a rule-based system is proposed to generate sufficient rules and actions, based on any old mitigation strategies or actions that were successfully taken during past storms, compliant with safety standards that apply in particular cities.

3.1 Proposed Research Methodology

Stage 1: A neural network will be applied to identify and monitor the weather forecast and determine the pattern of the dust storm from a radar station using a back-propagation algorithm, in order to obtain unclassified events (new cases).

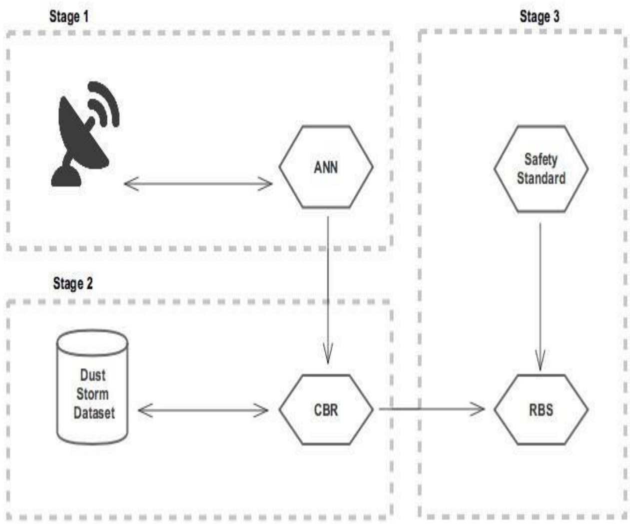

Fig. 1. Proposed research methodology

Stage 2: CBR will be applied to find a match between the new case and old cases in the database, in order to identify and predict the dust storm, before it begins.

Stage 3: A rule-based system will be used to update the database, taking into consideration the newly investigated case and the existing safety standards. The safety actions to be adopted will be listed, so the proper safety precautions can be taken to reduce the impact of the new dust event, based on the rules and safety standards.

4 Proposed Research Methodology

Our proposed methodology for dust storm classification comprises three stages:

4.1 Stage one: Find the similarity among dust storms cases using CBR.

Stage one of our methodology establishes the current benchmark of any existing cases for understanding better the current dataset of dust storm cases. For this stage we propose the following steps:

1. Find the similarity among the case-base cases and the investigated case (unclassified one) by using the simplified Euclidean distance equation. We apply this equation for all the case variable separately, and all variable weights are initialised to 1, including any new case as well.

$$\sqrt{\sum_{i=0}^{n} \left(x_i - y_i\right)^2} \qquad (1)$$

where $x=(x_1, x_2,\ldots,x_n)$ and $y=(y_1, y_2,\ldots,y_n)$ are two dust storm case vectors with i attributes and n+1 their maximum length of attributes.

Then similarity is calculated by aggregating all the results of the equation for each case individually. After that, all values are sorted in descending order, based on their similarity score. We called this way as the First Calculation. First Calculation shows us the similarity between a new case and any existing old cases, and establishes a neighbourhood of close cases to an investigated target case.

2. In the second calculation, we apply the same equation and we follow the same strategy with the same idea of the result, but this time we change the weight values of certain variables because some attributes will play a more important role than others, which will highlight the key attributes of the dust storm.

3. We import the dataset to myCBR tool, in order to measure and find the similarity between the cases themselves, by using weighted Euclidean distance equation measurements (2).

$$D(x,y) = \sqrt{\sum_{i=0}^{n} wi(xi - yi)^2} \quad (2)$$

4. From the observations into the dataset, we find out the most common characteristics of the dust storm, and from this point, we can see the reason for changing the values of selected variables, to see how they could predict if an event is either a dust storm or not.

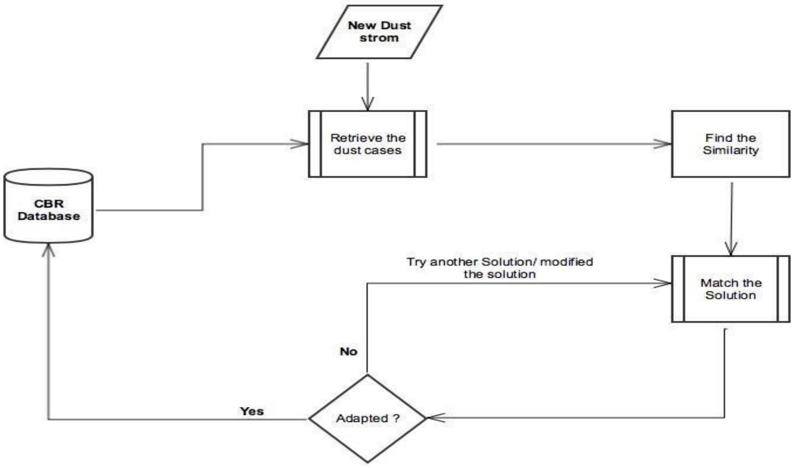

Fig. 2. CBR cycle in stage one

4.2 Stage two: Monitor the weather forecast to enhance the dataset with new dust cases using ANN techniques, and predict the dust storm using CBR.

1.After the key attributes of the dust storm have been identified, from the previous stage. ANN will be able to recognise the dust storm attributes from the weather forecast and provide them as unclassified cases input for further classification in our CBR system.
2.The CBR process will match the unclassified cases with old cases based on their similarity, to predict the new dust storm.
3. Evaluate the outcomes, by comparing the results with a weather forecasting satellite image.

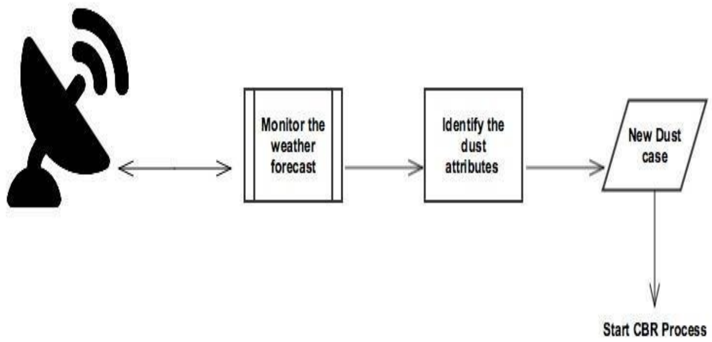

Fig. 3. Apply ANN in stage two

4.3 Stage three: Generate the proper actions, to reduce the impact of the dust storms using RBS technique.

1. In case of a successful dust storm classification using CBR, the Rule Based System can use any information associated to an old dust case that matches with the investigated dust storm case, to generate actions to reduce the harm of the forthcoming dust storm. RBS can also use any available safety standards to support the generated actions. The RBS will generate the rule based on the old solutions of the dust storm and the safety standard criteria

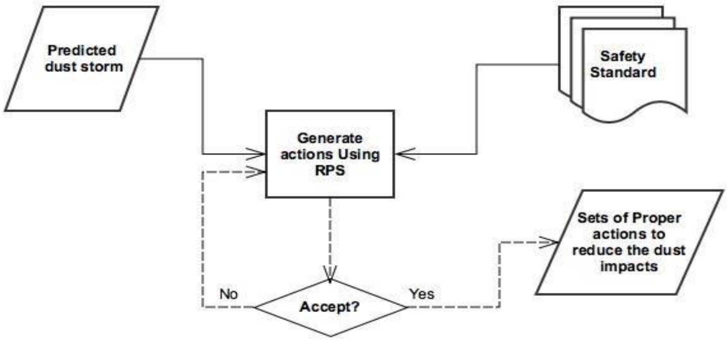

Fig. 4. Apply RPS in stage three

5 Dataset Accessibility

Through this study we apply the advantages of Artificial Neural Networks and Case-based reasoning to predict the magnitude of a dust storm before it happens, and the applicability of this research approach will be tested in the Kingdom of Saudi Arabia

(Riyadh). The weather experts in Saudi Arabia have agreed to provide us with a huge database for Riyadh city from 1999 to 2013.

In the last few years, Saudi Arabia has suffered from a number of huge sand storms, which have had a bad effect, as I mentioned above in the introduction. At the same time Saudi Arabia has a special and varied landscape. This provides an opportunity to measure the success of dust storm predictions.

6 Conclusions

As revealed in this paper, dust storms are likely to cause major damage to the environment, and spread diseases. Case- based reasoning will be used in our approach, to predict the magnitude of forthcoming dust storms. The CBR technique uses a similar past problem to solve a new one that is potentially similar in nature. Hence, it is very likely that, in such a case, any encountered dust storm case may have similar mitigation strategy to at least a past one. The combination of CBR and ANN in our approach could present effective results, because CBR has been proven successful in solving new problems, based on past experience in many projects and ANN has proven suitable for monitoring the weather forecast and satellite data. More investigations and analysis will be done for the next steps of this research to come up with applicable results in weather prediction.

7 References

1. Brockett, P. L., Cooper, W. W., Golden, L. L., Pitaktong, U. (1994). A neural network method for obtaining an early warning of insurer insolvency. Journal of Risk and Insurance, 61, pp. 402-402.

2. Aamodt, A., Plaza, E. (1994); Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. AI Communications. IOS Press, Vol. 7: 1, pp. 39-59.

3. Chacon-Murgua, M. I., Quezada-Holgun, Y., Rivas-Perea, P., Cabrera, S. (2011). Dust Storm Detection Using a Neural Network with Uncertainty and Ambiguity Output Analysis. Lecture Notes in Computer Science. Vol. 6718 (1): pp. 305-313.

4. Chen, J. H., and Hsu, S. C. (2007). Hybrid ANN-CBR model for disputed change orders in construction pro jects. Automation in Construction, 17(1), pp. 56-64.

5. Chen, S.H., Jakeman, A.J. & Norton, J.P. (2008). Artificial Intelligence Techniques: An Introduction to their Use for Modelling Environmental Systems, *Mathematics and Computers in Simulation* 78(2008),pp. 379-400.

6. Dibike, Y. B. and Solomatine, D. P. (2001). River flow forecasting using artificial neural networks. Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere, 26(1), 1-7.

7. Griffin, D. W., & Kellogg, C. A. (2004). Dust storms and their impact on ocean and human health: dust in Earth's atmosphere. *EcoHealth*, *1*(3), 284-295.

8. Imrie, C. E., Durucan, S. and Korre, A. (2000). River flow prediction using artificial neural networks: generalisation beyond the calibration range. Journal of Hydrology, 233(1), 138-153.

9. Kapetanakis, S., Filippoupolitis, A, Loukas, G., Al Murayziq, T. S. (2014). Profiling cyber attackers using Case-based Reasoning. In proceeding of the 19th UK Workshop on Case-Based Reasoning (UKCBR 2014), 9th December 2014, Cambridge, UK, pp.39-48.

10. Karlsson, E. Olsson, P. Funk, Hybrid Early Warning Systems, COMADEM, Osaka, June 2010

11. Kolodner, J. L. (1992). An introduction to case-based reasoning. Artificial Intelligence Review, 6(1), 3-34.

12. Lopez-Fernandez, H., Fdez-Riverola, F., Reboiro-Jato, M., Glez-Pena, D., & Mendez, J. R. (2011). Using CBR
as Design Methodology for Developing Adaptable Decision Support Systems. Efficient Decision Support
Systems Practice and Challenges from Current to future, Chiang J., (Ed.), USA: Intech Publishers,
123-126.

13. Main, J., Dilon, T. and Shiu, S. (2001). A Tutorial on Case-Based Reasoning, Soft Computing in Case- Based Reasoning, pp. 1-28.

14. Mascarenhas, S. (2010). *Case Based Reasoning & Fame IA,* Retrieved on 22 July 2014 from: https://fenix.tecnico.ulisboa.pt/downloadFile/3779574683590/ASS-Case-Based%20Reasoning.pdf

15. Polden, J. (2015). Engulfed by the sand storm: Man captures moment huge
cloud hit streets in Saudi Arabia. Retrieved April 8, 2015 from http://www.dailymail.co.uk/news/article-3029078/Sandstorm-wreaked-havoc- Saudi-Arabia-captured-camera-engulfing-road.html

16. Priddy, K. L., Keller, P. E. (2005). Artificial Neural Networks: An Introduction. SPIE Publications: Bellingham, WA, DC

17. Roth-Berghofer, T., Goker, M.H., Guvernir, H.A. (2006).*Advances in Case-Based Reasoning: 8th European Conference, ECCBR 2006, Fethiye, Turkey, September 4-7,2006 Proceedings.* Berlin, Germany: Springer-Verlag Publishers.

18. Shiu, S. C. K. and Pal, S. K. (2004). Case-Based Reasoning: Concepts, Features and Soft Computing, Applied Intelligence, 21, pp . 233-238.

19. Sivakumar, M. V. (2005). Impacts of sand storms/dust storms on agriculture. In Natural Disasters and Extreme Events in Agriculture (pp. 159-177). Springer Berlin Heidelberg.

20. Sivakumar,M.V.K. 2005 Impacts of Sand/Dust Storms on Agriculture. M.V.K. Sivakumar, R.P. Motha, and H.P. Das (Eds.). Natural Disasters and Extreme Events in Agriculture. Springer-Verlag, Berlin, Germany. 367p

21. Vidal, J. (2009). Dust Storms Spread Deadly Diseases Worldwide. London:
Guardian.co.uk. Retrieved April 23 2015 from
http://www.theguardian.com/world/2009/sep/27/dust-storms-diseases-sydney

22. Wei, K., Zhang, T., & He, B. (2008, May). Detection of sand and dust storms from meris image using FE-Otsu alogrithm. In Bioinformatics and Biomedical Engineering, 2008. ICBBE 2008. The 2nd International Conference on (pp. 3852-3855). IEEE.

23. Zakermoshfegh, M., Ghodsian, M. and Montazer, G. A. (2004). River flow forecasting using artificial neural networks. Hydraulics of Dam and River Structures, 90, pp . 425-430.