Proceedings of the 2013 Federated Conference on
Computer Science and Information Systems pp. 107–110

Recommender System for Ground-Level Ozone Predictions in Kuwait

Mahmood A. Mahmood[1,*], Eiman Tamah Al-Shammari[2], Nashwa El-Bendary[3,*], Aboul Ella Hassanien[4,*], Hesham A. Hefny[1]

[1]*ISSR, Computer Sciences and Information Dept., Cairo University, Cairo - Egypt*
mahmood.moneim@egyptscience.net, h.hefny@ieee.org
[2]*Faculty of Computing Science and Engineering, Kuwait University, Kuwait*
eiman.tamah@gmail.com
[3]*Arab Academy for Science,Technology, and Maritime Transport, Cairo - Egypt*
nashwa.elbendary@ieee.org
[4]*Faculty of Computers and Information, Cairo University, Cairo - Egypt*
aboitegypt@gmail.com
[*]*Scientific Research Group in Egypt, (SRGE), http://www.egyptscience.net*

***Abstract*—This article presents a recommender system based on rough mereology for predicting Ozone concentration in Kuwait through testing the data gathered from Al-Jahra station. The proposed recommender system consists of three phases; namely pre-processing, classification, and recommendation phases. To evaluate the performance of the presented recommender system, fifteen parameters were used. Those parameters were developed and validated between Jan. 2006 and Sept. 2010. The obtained results demonstrated the effectiveness and the reliability of the proposed recommender system.**

***Index Terms*—recommender system, rough mereology, air pollution, ground-level Ozone**

I. Introduction

INDUSTRIALIZATION, technical growth and over population in urban areas of Kuwait have resulted in increased air pollution [1], [2]. This has deteriorated the quality of fresh air. Toxic air pollutants in close proximity to populate areas can have adverse health effects. In this respect, surface Ozone can become a serious problem in the urban areas of Kuwait - if it frequently occurs in sufficient concentration to threaten human health and environment [3].

Ground-level Ozone (O3) is formed by chemical reactions between nitrogen oxides (NOx) and volatile organic compounds (VOC) in the presence of heat and sunlight. It is difficult to exactly define the formation and destruction mechanism of Ozone. This is because Ozone is an extremely reactive pollutant and can be scavenged by its precursors [2], [4]. As a result, the area of air pollution forecasting through empirical methods has gained importance with the availability of sufficient data. Earlier forecasting models were based on simple empirical data correlations, but the availability of a large amount of information has resulted in development of complex air pollution simulations for forecasting [2], [5]. Management of public warning strategies for Ozone levels in densely populated areas require accurate forecasts of ambient levels. Although Ozone prediction models exist or have been proposed at several cities [6]–[9], [16], [17], they have not been assessed in realistic conditions.

Recently, there has been an increase in air pollution in urban areas of Kuwait. This is the result of rapid industrialization, technical growth and over population. In the past, it has been observed that toxic air pollutants in close proximity of populated areas can have adverse health effects. Ozone is one such pollutant that can become a serious problem in urban areas of Kuwait if it occurs in sufficient concentration. Thus, accurate forecasting of surface Ozone is required as it can help with successful implementation of public warning strategies during episodic days in Kuwait.

This article presents a recommender system based on rough mereology for predicting Ozone concentration in Kuwait through testing the data gathered from Al-Jahra station. The proposed system firstly maps the Ozone dataset into a normalized dataset of ground-level Ozone predictions. Then, rough mereology and rough inclusion techniques are applied for clustering and classifying the normalized Ozone dataset into sets of granules with different radius. Voting by objects approach is subsequently applied in order to select the optimized granules. Finally, normalized rating matrix is acquired, then the predicted ground-level Ozone is recommended. To evaluate the performance of the presented recommender system, fifteen parameters were used. Those parameters were developed and validated between Jan. 2006 and Sept. 2010. The initial three years of data are used to develop the predicting models and the remaining data is used for testing and verifying these models.

The rest of this article is organized as follows. Section II presents the basic concept of rough mereology. Section III describes the different phases of the proposed recommender system; namely pre-processing, classification, and recommendation phases. Section IV introduces experimental results via firstly discussing the tested dataset in addition to the details of

978-1-4673-4471-5/$25.00

the applied Air Quality Index (AQI), then presenting statistical analysis of the obtained experimental results. Finally, Section V presents and discusses conclusions.

II. Rough Mereology: An Overview

Rough mereology proposed by Lesniewski in [10] as the theory of concept. The relation of mereology is a part of relation, e.g. x mereology y means x is a part of y. According to Polkowski [11], the mereology relation described in equation (1), where $\pi(u, w)$ is a partial relation (proper part) and $ing(u, w)$ is ingredient relation means an improper part.

$$ing(u, w) \Leftrightarrow \pi(u, w) \ or \ u = w \tag{1}$$

$\mu(x, y, r)$ means rough mereology relation x is part of y at least degree r, also described as shown in equation (2).

$$\mu(x, y, r) = sim_\delta(x, y, r) \Leftrightarrow \rho(x, y) \leq (1 - r) \tag{2}$$

Computing the indiscernibility relation to get the object can be done using rough inclusion, which is of less complexity time than the indiscernibility relation computed by rough set technique. Rough inclusion from metric according to Polkowski in [11] computed by the Euclidean metric space or Manhattan space, where,

$$\mu_h(x, y, r) \Leftrightarrow \rho(x, y) \leq 1 - r \tag{3}$$

Then, the indiscernibility relation Ind can be computed as shown in equation (4).

$$Ind(x, y) = \frac{|IND(x, y)|}{|A|} \tag{4}$$

Then, equation (2) becomes equations (5) and (6).

$$\mu_h(x, y, r) \Leftrightarrow Ind(x, y) \geq r \tag{5}$$

$$IND(x, y) = a \in A : a(x) = a(y) \tag{6}$$

Where, a is an attribute(s) in an information system A, and $|A|$ is the cardinality of a set A.

III. The proposed recommender system

The proposed Ozone recommender system tested for ground-level Ozone data collected at Al-Jahra city in Kuwait during the time from January/2006 to September/2010. The architecture of the proposed system consists of three phases; namely pre-processing, classification, and recommendation phases.

A. Pre-processing Phase

In this phase, the proposed recommender system receives the ground-level Ozone data as an input, then it generates a normalized rating matrix. As shown in equation (7) with X representing the original value, the input data were mapped into a normalized dataset of the range [0,1], where the values 0 and 1, represent the smallest and the largest values in each dataset attribute, respectively.

$$Normalized = \frac{X - min.value}{max.value - min.value} \tag{7}$$

B. Classification Phase

In this phase, the proposed recommender system receives the normalized Ozone dataset generated from the pre-processing phase. Rough mereology and rough inclusion approaches were applied in order to produce rough inclusion table that reflects similarity degree among parameters. Then, the granulation mechanism that reflects a given set of inclusion data into collection of granules will be applied via voting by training objects in order to produce the optimal similarity measurement.

1) Ground-level Ozone Clustering: In this phase, the granular computing formalized within the theory of rough mereology - as proposed by Polkowski; as an application of the idea of a granular reflection of data and of classifiers induced from it [5] is used to classify the normalized data into group of similar data according to the definition of rough mereology. In this section, a brief description of the theory of rough mereology and rough inclusion will be presented.

2) Rough Mereology: Indiscernibility relations in rough set theory represents the core problem in this theory that it takes a long of time to get the classification of data. Rough mereology theory used the flexible similarity relations, which allow for huge data to be classified. The similarity relations that will be used must be satisfy properties are MON, ID, EXT. Rough inclusion technique satisfy their three properties of similarity relations; namely Monotonic (MON), Identity (ID), Extreme, or proportionality (EXT) [5].

- (MON) if similarity (x, y, 1) then for each z, from similarity (z, x, r) it follows that similarity (z, y, r).
- (ID) similarity (x, x, 1) for each x.
- (EXT) if similarity (x, y, r) and $s \leq r$ then similarity (x, y, s).

3) Rough Inclusion: Rough Inclusion is a technique that uses the Reduced Hamming Distance [11] equation to compute the similarity between vector u and v, where u represents a user and v represents an item, as shown in equation (8) [5], where, $IND(u, v) = \{a \in A : a(u) = a(v)\}$, and $|A|$ denotes the cardinality of set A.

$$ind(u, v) = \frac{|IND(u, v)|}{|A|} \tag{8}$$

After applying equation (8) on the normalized rating matrix, similarity table is produced by using rough inclusion approach. The proposed recommender system classifies the attributes of the dataset using *Gödel* t-norm technique [6], as shown in equation (9).

$$T_{min}(a, b) = min\{a, b\} \tag{9}$$

The granulation mechanism reflects a given inclusion data into collection of granules, each granule has a fixed granule radius value r, where $r \in$ the interval [0,1] [7]. Accordingly, the proposed recommender approach applies the voting by training objects to produce prediction/recommendation regarding the ground-level Ozone.

4) Voting by Training Objects: Voting by objects stage takes as an input the list of similarity measure tables for each radius resulted from the clustering stage and computes the accuracy rate for each radius as shown in equation (10). Then, the optimum radius that represent the largest accuracy measure is selected. This stage is divided into two steps: 1) accuracy measure computation, which uses equation (10) to compute the accuracy for each table representing the similarity measure at radius r and 2) optimization step, where the table containing the largest accuracy measure at radius r is selected, so the radius r called the optimum radius r_{opt}.

$$Accuracyrate = (\frac{T}{N}) * 100 \tag{10}$$

Where, T is the rate of number of valid tuples and N is the total number of tuples in each row.

Let a is an attribute, u is a test object, v is a training object, and ϵ is a selected number in an interval $[0, 0.1]$ taken every 0.01. So, the final value of ϵ is taken according to the equation of the factor, shown in equation (11).

$$q_a(u, v) = \frac{|a(u) - a(v)|}{diam(a)} \tag{11}$$

Where, diam(a) is computed as shown in equation (12).

$$diam(a) = Max(a) - Min(a) \tag{12}$$

and $a(u)$ can be computed as follows:

$$a(u) = \begin{cases} min & \text{if } u < min, \\ max & \text{if } u > max, \\ u & \text{if } otherwise \end{cases}$$

For each selected ϵ, $sel_u(c, t) = \frac{\sum w_u(v,t)}{sizeofc}$ is computed, where c is the decision category, t is selected t-norm, and $w_u(v, t) = ind_\epsilon(u, v)$. Then, u is assigned to the category with maximal $sel_u(c, t)$. The end value of ϵ is selected when $q_a(u, v) < \epsilon$. Based on the optimum radius selected from the voting by object stage, rules are generated from the rough inclusion table of selected optimum radius.

C. Recommendation Phase

In this phase the system receives the testing Ozone dataset as an input, then it outputs a recommendation/prediction value of ground-level Ozone as shown in equation (13) according to the formula that used to recommendation/prediction in collaborative filtering technique.

$$prediction = \frac{\sum_i \frac{w_{u,i} - mean}{\sigma_u} \times w_{a,u}}{\sum_i w_{a,u}} \times \sigma_a + mean \tag{13}$$

Where, $w_{u,i}$ is the Pearson's correlation coefficient shown in equation (14).

$$w_{a,u} = \frac{\sum_{i=1}^{m} (r_{a,i} - mean) \times (r_{u,i} - mean)}{\sigma_a \times \sigma_u} \tag{14}$$

IV. Experimental analysis and discussion

A. Dataset

The study used hourly air pollutants data from January 2006 through September 2010 gathered by the Environmental Monitoring Information System of Kuwait (eMISK, working under the Environment Public Authority of Kuwait). The data were collected from Al-Jahra fixed surface station. The initial three years of data was used to develop the forecasting models and the remaining data was used for testing and verifying these models. Table I represents the generated rules of Al-Jahra dataset.

B. Air Quality Index (AQI)

The Air Quality Index (AQI) is a key tool for making information about outdoor air quality as easy to find and understand as weather forecasts. It is an index for reporting daily air quality via updating how healthy or unhealthy the air is. Equation (15) shows the AQI conversion formula for the ground-level Ozone.

$$AQI = \frac{I_{HI} - I_{LO}}{BP_{HI} - BP_{LO}} \times (C_{O3} - BP_{LO} + I_{LO}) \tag{15}$$

Where, I_{HI} is the index value at the upper limit of the AQI category, I_{LO} is the index value at the lower limit of the AQI category, BP_{HI} is the break-point concentration at upper limit of the AQI category, BP_{LO} is the break-point concentration at lower limit of the AQI category, and C_{O3} refers to 1-hour Ozone concentration. Also, depending on the value of the Ozone concentration C_{O3}, the values of the I_{HI}, I_{LO}, BP_{HI}, and BP_{LO} parameters are being identified. For example, if the detected value of the C_{O3} is within the range [0.00, 0.059], then the values of the I_{HI}, I_{LO}, BP_{HI}, and BP_{LO} parameters will be 50.00, 0.00, 0.059, and 0.00, respectively [12]. The AQI is considered a standard measurement that runs from 0 to 500, where the higher the AQI value, the greater the level of air pollution and the greater the health concern [13].

C. Statistical Analysis

Evaluating the recommendation quality of the proposed recommender system is mainly based on statistical precision

TABLE I
RULE GENERATION TABLE OF AL-JAHRA DATASET

Rule	Description	Accuracy
Rule(1)	if SO2 is healthy then Ozone is healthy	79.1087
Rule(2)	if NO is healthy then Ozone is unhealthy	29.5273
Rule(3)	if NOX is healthy then Ozone is healthy	62.2358
Rule(4)	if NO2 is healthy and Wind-deg is healthy then Ozone is unhealthy	26.7496
Rule(5)	if PM10 is healthy then Ozone is healthy	79.3381
Rule(6)	if CO is healthy and Wind-deg is healthy then Ozone is unhealthy	26.7496
Rule(7)	if CH4 is healthy then Ozone is healthy	76.1768
Rule(8)	if NCH4 is healthy then Ozone is healthy	73.8360
Rule(9)	if WS is healthy then Ozone is healthy	83.3083
Rule(10)	if WG is healthy then Ozone is healthy	83.8488
Rule(11)	if SOLAR is healthy then Ozone is healthy	80.2869
Rule(12)	if TEMP-IND is healthy then Ozone is unhealthy	44.2708
Rule(13)	if TEMP-AMB is healthy then Ozone is healthy	83.3141
Rule(14)	if CO2 is healthy then Ozone is healthy	69.2238

and decision supporting precision measurement methods [14]. Statistical precision measurement method adopts the MAE (Mean Absolute Error) in order to measure the recommendation quality [15]. MAE is a commonly used recommendation quality measurement method. MAE calculates the irrelevance between the recommendation value predicted by the recommender system and the actual evaluation value. Each pair of interest predicted rank is represented as $< p_i, q_i >$, where p_i is the system predicted value and q_i is the user evaluation value. Based on the entire set of $< p_i, q_i >$ pairs, MAE calculates the absolute error value $|p_i - q_i|$ and the sum of all the absolute error value. Then, the average value is calculated. Small MAE values represent a good recommendation quality indicators. The predicted values of rating sets can be represented as $p_1, p_2, ..., p_N$ and the corresponding actual testing rating set can be represented as $q_1, q_2, ..., q_N$. The MAE can be defined as shown in equation (16).

$$MAE = \frac{\sum_{i=1}^{N} |p_i - q_i|}{N} \tag{16}$$

Based on experimental results, the ground-level Ozone depends on the values of NO, (NO2, CO), Wind direction degree, and Industrial temperature. Figure 1 depicts the monthly actual values, predicted values by the proposed recommender system, and the mean absolute error (MAE) of the ground-level Ozone in Kuwait (AL-Jahra city station) during the time from January-2009 to September-2010. As shown in figure 1, the curve presenting the values of ground-level Ozone predicted by the proposed recommender system has a similar behavior as the curve presenting the actual values of the ground-level Ozone dataset. Also, both curves existed in the healthy region of the O3 value, which is less than 0.165. For the MAE curve, it almost matches the predicted curve for the months (10/2009, and 11/2009) as the data readings gathered in the tested dataset during those months was zeros, the MAE curve may be in some points drawn over the actual values of the ground-level Ozone dataset because the actual dataset attribute's has sparsity problem and the rough mereology classifier predict this value as the lowest value in this attribute.

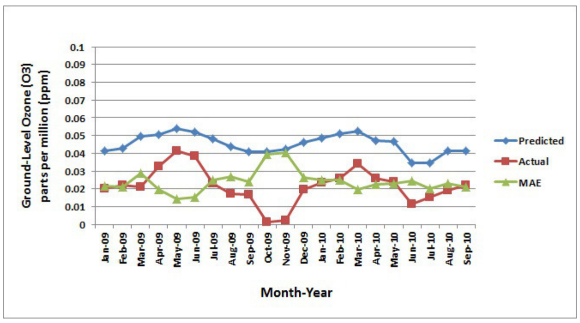

Fig. 1. Monthly values of actual, predicted, and MAE values of ground-level Ozone in Kuwait (AL-Jahra) [Jan-2009 to Sept-2010]

V. Conclusions

In this article, based on ground-level Ozone concentration data gathered from Al-Jahra station in Kuwait, a rough mereology based recommender system was presented for the prediction of ground-level Ozone pollution. The obtained results demonstrate the effectiveness and the reliability of the proposed recommender system. Resulted experimental values of ground-level Ozone predicted by the proposed recommender system showed similar behavior as the actual tested values of the ground-level Ozone dataset. Also, both experimentally resulted and actual dataset values existed in the healthy region of the O3 value, which is less than 0.165 ppm according to the reference AQI.

References

[1] S. Abdul-Wahab, W. Bouhamra, H. Ettouney, B. Sowerby, B. D. Crittenden, "Analysis of Ozone Pollution in the Shuaiba Industrial Area in Kuwait," *Int. J. Environ. Studies*, vol. 57, no. 2, pp. 207–224, 2000.

[2] E. T. Al-Shammari, "Public warning systems for forecasting ambient ozone pollution in Kuwait," *Environmental Systems Research*, vol. 2, no. 2, 2013.

[3] K. N. Jallad, C. E. Jallad, "Analysis of Ambient Ozone and Precursor Monitoring Data in a Densely Populated Residential Area of Kuwait," *J. Saudi Chem. Soc.*, vol. 14, pp. 363–372, 2010.

[4] B. Dimitriades, "Photochemical Oxidant Formation: Overview of Current Knowledge and Emerging Issues," *Atmospheric Ozone Research and its Policy Implications*, vol. 35, Studies in Environmental Science, Elsevier Science Publishers, Amsterdam, pp. 35–43, 1989.

[5] B. Telenta, N. Alfksic, M. Dacic, "Application of the Operational Synoptic Model for Pollution Forecasting in Accidental Situations," *Atmos. Environ.*, vol. 28, pp. 2885–2891, 1995.

[6] S. M. Robeson, D. G. Steyn, "Evaluation and Comparison of Statistical Forecast Models for Daily Maximum Ozone Concentrations," *Atmos. Environ.*, vol. 24B, pp. 303–312, 1990.

[7] D. Elsom, "Smog Alert: Managing Urban Air Quality," *London: Earthscan Publications Limited*, 1996.

[8] H., Noordijk, "The National Smog Warning System in the Netherlands; a Combination of Measuring and Modeling," *Air Pollution*, vol. 2, Pollution Control and Monitoring, WIT Press: Southampton, pp. 169–176, 1994.

[9] J. Yi, V. R. Prybutok, "A Neural Network Model Forecasting for Prediction of Daily Maximum Ozone Concentration in an Industrialized Urban Area," *Environ. Pollut.*, vol. 92, pp. 349–357, 1996.

[10] S. Lesniewski, "On the foundations of set theory," *Topoi*, vol. 2, pp. 7–52, 1982.

[11] L. Polkowski and P. Artiemjew, "Granular Computing in the Frame of Rough Mereology. A Case Study: Classification of Data into Decision Categories by Means of Granular Reflections of Data, rq' *International journal of intellegent systems*, vol. 26, no. 6, pp. 555–571, 2011.

[12] http://www.ncair.org/airaware/Ozone/codecalc.shtml (Accessed: March,2013)

[13] U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Outreach and Information Division, "Air Quality Index (AQI) - A Guide to Air Quality and Your Health, EPA-456/F-09-002, August 2009.

[14] R. S. Ettouney, F. S. Mjalli, J. G. Zaki, M. A. El-Rifai, H. M. Ettouney, "Forecasting of Ozone Pollution using Artificial Neural Networks," *Mgmt. Environ. Quality: An Int. J.*, vol. 20, no. 6, pp. 668–683, 2009.

[15] S. Abdul-Wahab, W. Bouhamra, H. Ettouney, B. Sowerby, and B. D. Crittenden, "Predicting Ozone Levels: A Statistical Model for Predicting Ozone Levels," *Environ. Sci. Pollut. Res.*, vol. 3, pp. 195–204, 1996.

[16] A. Ali, S. E. Amin, H. H. Ramadan, and M. F. Tolba, "Ozone monitoring instrument aerosol products: Algorithm modeling and validation with ground based measurements over Europe," *Proceedings of the International Conference on Computer Engineering and Systems, (ICCES'2011),*, pp. 181–186, 2011.

[17] A. Ali, S. E. Amin, H. H. Ramadan, and M. F. Tolba, "Data assimilation of ozone monitoring instrument images for improving aerosol optical depth prediction," *8th International Conference on Informatics and Systems (INFOS 2012)*, pp. BIO131–BIO139, 2012.