Assessing the weight of weather parameters on MW signal strength by descriptive data mining

Seyed Ahmad Mirbagheri[1] · Mostafa Mohammadi[1] · Somayyeh Bazrehkar[2]

Abstract

In this paper, weather parameters and electric field strength are measured in MW frequency band in Dashte_Ghazvin station. The experimental results were analyzed statistically as well as data mining techniques. Finally, the effects of weather parameters on signal strength were analyzed. In the far-field situations, the electric or magnetic field strength individually can be the representative of the electromagnetic wave radiation. In this paper, the electromagnetic radiation was assessed using measured electric field in terms of Volt per meter that changed from 8.50 to 3.32 V/m. The large amount of data provided the possibility of using data mining technique to find the exact relationship between the weather parameters and signal strength as well as the amount of their effects on signal strength. After the acquisition of data during 1 year, preprocessing of data was performed and weights were obtained through correlation. During a period of 1 year, 9,000,000 data was gathered in 2018 from January to December. The data includes the 5 weather parameters and signal strength. The results showed relationships between atmospheric parameters and the electric field strength. Additionally, the correlation analysis used to assess the relationship between each parameter and signal strength proved that the air humidity and air pressure have an inverted linear relationship with the signal level whereas the other variables have a direct linear relationship. The results are useful to design and develop the station before any construction.

Keywords MW band · Meteorological · Data mining · Electric field strength

1 Introduction

Seasonal variation may affect received signal strength. For example, the propagation of GSM frequency band is dependent on the propagation waves in the troposphere layer [8]. In this case, the intensity values of the electromagnetic waves can be used for different atmospheric conditions. Received signal strength is related to refractivity factor directly [5] as:

$$N = (n - 1) \times 10^6 \tag{1}$$

where n is the refractive index and N depends on atmospheric parameters like temperature, air pressure and vapor pressure by:

$$N = 77.6 \frac{P}{T} + 3.732 \times 10^5 \left(\frac{e}{T^2} \right) \tag{2}$$

where T is absolute temperature in Kelvin, P is air pressure in hpa and e is the vapor pressure. Refractive nature of the atmosphere is important to propagate the radio frequency [8]. A number of the weather parameters changed day by day or even hour by hour. This study aims to investigate the impact of atmospheric conditions on the signal strength due to their significant negative impact on the environment and human health. Previous studies have noted the effects of some parameters only and they did not widely measure any of these variables. Data mining is becoming an essential instrument for researchers to use methods and tools to analyze large quantities of data.

Data analyst use Data Mining to select, explore and model large amounts of data stored in databases to recognize unknown patterns or relationships [4].

In this paper, the goal of data mining is to assess the effects of weather parameters on signal strength in an MW station. Data mining which is known as knowledge discovery in databases (KDD) is the procedure of discovering useful knowledge from large quantities of data stored in data warehouses. Owing to the fact that such large amount of data had not been collected before, data mining was used to process the available data.

In this paper, the goal of data mining is finding the relationship between the electric field strength and weather parameters in an AM. In the following sections we will see that: Sect. 2 presents a review of related works. Section 3 describes the method of measurements and equipment used in the pilot district. Section 4 explains the methodology. Section 5 presents the results and analysis. Finally the conclusion is stated in Sect. 6.

2 Literature review

Although, many researchers have focused on the effect of mobile sources on human health [6], research on the effect of radio and TV stations are also significant. As well as the threat to the human body, the use of cellular phones and base stations, broadcasting transmitters may be dangerous [1], more than acceptable radiation exposure level. The related works are divided into two categories; a number of them relate to the influence of atmospheric parameters while others relate to the modeling of signal strength based on atmospheric parameters. This paper investigates the influence of atmospheric parameters on signal strength in MW frequency band using a data mining technique for a large quantity of data. However, although many researchers exist on data mining to process big data [7, 10, 11], the assessment of weather parameters changes and its effect on signal strength is normally not performed on data mining to process the data. The researchers showed that the change of atmospheric parameters is effective at the received signal VHF/UHF [9] whereas the weather parameters changes effect on the AM band requires more research. In this study, a large amount of data was gathered in relation to the AM frequency band. Since the main goal of data mining is to find hidden relationships between input and output parameters, data mining was used for data analysis in this paper.

3 Measurement method and Equipment

Dasht-E-Ghazvin is the most powerful MW station in Iran and one of the most powerful stations in the Middle East. The main transmitter in this site radiates 1 MW power in the omni-directional pattern. Due to the high-power radiation and avoiding interference with European countries, the mast height was reduced to 5/8λ of its frequency. With regard to low frequency assignment in this station (585 kHz), the vast area in the middle of Iran is covered. For analysis, data was collected during a period of 1 year. The materials used for the experiment were NARDA EHP200A and Combilog1022. The inputs and outputs were measured almost every 20 s on average. Hence, the number of recorded data collected in a day amounted to 4320 samples. Figure 1 shows NARDA EHP200A device to measure the electric field strength at the measuring point. This instrument is able to measure the electric filed for frequency spectrum from 9 kHz to 30 MHz. The probe was held at 1.5 m above the ground at each measurement and remained at the fixed position during the study period. The instrument for measuring atmospheric parameters has a specific sensor for each parameter. The weather parameters logger recorded data every 20 s and was exported to a PC. The weather parameters were used as inputs.

4 Methodology

Figure 2 shows the different steps of the data mining procedure. The KDD process [3] includes 9 steps. Brachman et al. [2] determined KDD process and asserted its' respective nature. These steps include: First, we should realize the application domain and understand the customer's object of the KDD process. This step involves the understanding of the application domain, knowing the relevant knowledge, and identifying the aim of process. Consequently, the target data can be chosen and the proper data samples and a relevant subset of variables can be selected.

Then we should construct a target data set, which has a data set, or combination on a subset of variables or discovery data samples. It means that the data turns into a propositional form, where each instance is represented by a feature vector. In addition, dimensionality reduction methods can also be applied in this step to improve the performance of subsequent data mining algorithms.

Fig. 1 NARDA EHP200A instrument measured the electric field strength in MW station

Fig. 2 Different steps of knowledge discovery in databases (KDD) [2]
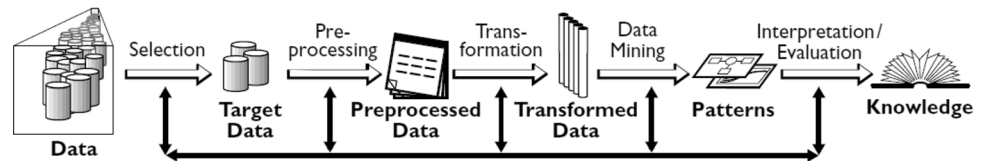
The next step is data preprocessing includes data cleaning, noise removing, to collect the essential information for model, we should select strategies for checking. This step includes the handling of missing values, the identification, the elimination of duplicates, and also the matching, fusion.

In fourth step, data reduction and extension that finds useful characteristics is done to represent the data. The large number of variables under consideration can be reduced. In fifth step, special are data-mining methods such as summarization, classification and regression and clustering are used to reach the goals of the KDD process data analysis, modeling and hypothesis selection will have done in the next step. Data mining algorithm is selected and data pattern method is chosen and the optimal data-mining method of the KDD process is selected with the overall criteria. Data mining is the seventh step, which looked forward for patterns of interest form classification rules or trees, regression and clustering. The main issue in this step is finding the appropriate models and parameters might be used, and a particular data mining method. After selecting the data mining method and algorithm, the data mining is carried out to search for patterns of interest in a particular representational form or a set of such representations, such as rule sets or trees.

Interpreting mined patterns is the eighth step, which repeats any of the steps 1 through 7 for more iteration. The patterns and models from data mining algorithm are examined. Furthermore, the user evaluate the usefulness of the found knowledge for the given application. In this step, the visualization of the extracted patterns and models is achieved. Discovering knowledge and incorporating it to act in other systems is the last step.

Documenting and reporting and checking for and resolving potential hazards with previously extracted knowledge can be completed in this step.

5 Results

Two categories of data were collected in this study. The first one related to the weather parameters; that were gathered by Comblig 1022 every 20 s and the second data that was signal level which was captured by NARDA EHP200A every 5 min. The large amount of prelimary data was ignored because of syncopation between two categories of data so that the amount of data reaches to 38,400.

Long-term data collected was stored accordingly in an SQL-based database. Each record contains 5 input parameters including temperature, air pressure, humidity, wind direction, wind speed and an electric field intensity variable in V/m as an output which stored according to the date and time. The measured data was tabulated in Tables 1, 2 as well as the minimum, maximum, standard deviation of each parameter. The descriptive statistics computed using Rapid miner version 7.1. The histograms are the output of software that show the parameter value in X axes and the y axes states the frequency of parameter. The units of each parameter are stated in each row of the Table 1.

Table 1 The unit and histogram of measured data

Parameters	Unit	
Air pressure	Pa	
Humidity	%	
Air temperature	°C	
Wind direction	Degree	
Wind speed	m/s	
Output	V/m	

Table 2 The descriptive statistics of measured data

	Output (V/m)	Wind direction (degree)	Wind speed (m/s)	Air temp (C°)	Air humidity (%)	Air pressure (pa)
Statistics						
N						
Valid	38,418	38,418	38,418	38,418	38,418	38,418
Missing	0	0	0	0	0	0
Mean	27.18	177.31	2.29	16.47	69.49	881.67
Median	28.14	180	2	15.50	74	881
Mode	29.15	266	1	10.8	96.2	881
SD	3.89	96.50	2.40	6.24	24.27	3.13
Variance	15.14	9312.37	5.77	38.95	589.02	9.82
Minimum	8.51	2	0.0	6.1	8.6	874
Maximum	37.32	360	25	39.4	100	890
Sum	1,044,202.67	6,811,961	88,134	632,835.8	2,669,588	33,872,124

Fig. 3 Weight by Correlation model in rapid miner
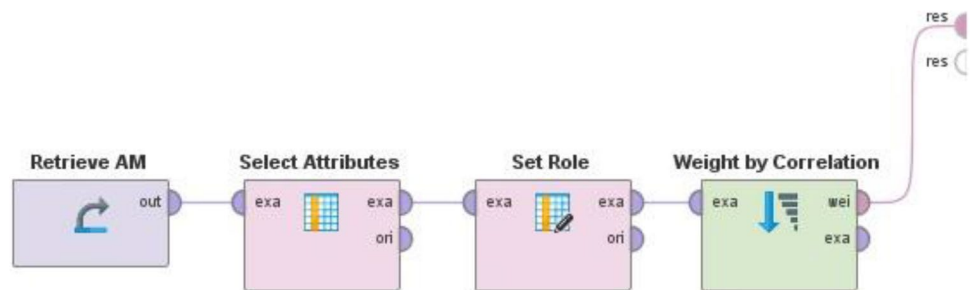
Table 3 Weight by correlation

Attribute	Weight
Wind speed	0.027
Air pressure	0.030
Wind direction	0.118
Humidity	0.250
Temperature	0.262

air pressure and signal strength. It means that increasing the air pressure or humidity decreases the signal strength. Therefore, the best time to measure the maximum electric filed strength is when the air pressure and humidity have maximum values. In fact, it concluded that the time of measuring in 1 year can be effective on the measuring results and assessment.

Afterwards Mahalanobis distance method was used to delete the outlier whose number was 889 records. Figure 3 shows the weighting procedure by correlation process of the rapid miner software.

Table 3 shows the output of weight by correlation operator. The Weight by Correlation operator calculates the relation between the attributes by computing the value of correlation for each attribute of the inputs with respect to the output. The weighting scheme is based on correlation and it returns the squared value of correlation as attribute weight. According the Table 3, the weight of the temperature is 0.262. It means that there is a relation between the temperature and signal strength. In addition, the correlation coefficients were yielded by correlation matrix show that there is an inverse relation between the

6 Discussion

We supposed firstly, the weather parameters would have to impact on signal level. The main question was about this impact on signal level. For this end, the weather parameters were measured in different period of time in the year from January to December of 2018. To be exact, a large amount of data was collected by two equipment; first, one involved the measuring of weather parameters that are air humidity, air temperature, air pressure, wind speed and wind direction. The second one measured the signal level at the specific point in an AM radio station. After collecting the data, the analysis was carried out by statistical aspect using Rapid miner software. In fact, the amount of impact of each weather parameter on signal

level was stated from the statistical analysis using correlation quantity.

7 Conclusion

In this paper, data on atmospheric parameters and the electric filed strength was monitored for 1 year. The acquired data was subsequently analyzed by Data Mining and the relationship between these variables were determined. The results indicated that the air humidity and air pressure have an inverted linear relationship with the signal level while the other variables have a direct linear relationship. The signal strength and exposure level affected by detected parameters include air temperature, air humidity, wind direction, air pressure and wind speed respectively. Henceforth, to optimally analyze the signal strength in broadcast stations, weather parameters have shown to be highly influential. Furthermore, according to the results, the best time to measure the electric field strength is when the humidity and air pressure were at the minimum and the temperature, wind speed and wind direction at the maximum.

Compliance with ethical standards

Conflict of interest The authors have no conflict of interest to declare.

References

1. Boz K et al (2018) Spatial electromagnetic field intensity modelling of global system or mobile communication base stations in the Istanbul Technical University Ayazaga campus area. Geospat health 13:527
2. Brachman RJ et al (1996) The process of knowledge discovery in databases. In: Fayyad UM, Piatetsky-Shapiro G, Smyth P, Uthurusamy R (eds) Advances in knowledge discovery and data mining. American Association for Artificial Intelligence, Menlo Park
3. Chau K-W et al (2007) Data mining and multivariate statistical analysis for ecological system in coastal waters. J Hydroinform 9:305–317
4. Giudici P (2005) Applied data mining: statistical methods for business and industry. Wiley, New York
5. Hall MPM (1996) Effects of the troposphere on radio communication. Peter Peregrinus Ltd, London
6. Hamnerius Y et al (2016) Microwave exposure from mobile phones and base stations in Sweden. In: Proceedings of the international conference on cell tower sitting
7. Liu X et al (2016) Review on the data mining technology and the applications on financial analysis area. In: International conference on communication and electronics systems (ICCES). IEEE, Coimbatore, India
8. Mat R et al (2016) Temperature effect on the tropospheric radio signal strength for UHF band at Terengganu, Malaysia. Int J Adv Sci Eng Inf Technol
9. Schmiedel S et al (2008) An evaluation of exposure metrics in an epidemiologic study on radio and television broadcast transmitters and the risk of childhood leukemia. Bioelectromagnetics 30:81–91
10. Suresh R et al (2017). Data mining and text mining—a survey. In: 2017 international conference on computation of power, energy information and communication (ICCPEIC). IEEE, Melmaruvathur, India
11. Wang J et al (2017) The application of data mining technology to big data. In: 2017 IEEE international conference on computational science and engineering (CSE) and ieee international conference on embedded and ubiquitous computing (EUC). IEEE, Guangzhou, China

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.