Report on my own edx project

FUNG CHE HEI

8/19/2020

Abstract

This project The data describe the problem of high energy (higher than 10^4 J) seismic bumps forecasting in a coal mine. Data come from two of longwalls located in a Polish coal mine.

1. Introduction

1.1 Background

Mining activity was and is always connected with the occurrence of dangers which are commonly called mining hazards. A special case of such threat is a seismic hazard which frequently occurs in many underground mines. Seismic hazard is the hardest detectable and predictable of natural hazards and in this respect it is comparable to an earthquake.

In seismic hazard assessment data clustering techniques can be applied (Lesniak A., Isakow Z.: Space-time clustering of seismic events and hazard assessment in the Zabrze-Bielszowice coal mine, Poland. Int. Journal of Rock Mechanics and Mining Sciences, 46(5), 2009, 918-928), and for prediction of seismic tremors artificial neural networks are used (Kabiesz, J.: Effect of the form of data on the quality of mine tremors hazard forecasting using neural networks. Geotechnical and Geological Engineering, 24(5), 2005, 1131-1147).

1.2 Aim

The main aim of all seismic hazard assessment methods is to predict (with given precision relating to time and date) of increased seismic activity which can cause a rockburst.

With the information about the possibility of hazardous situation occurrence, an appropriate supervision service can reduce a risk of rockburst (e.g. by distressing shooting) or withdraw workers from the threatened area. Good prediction of increased seismic activity is therefore a matter of great practical importance.

1.3 Data set Information

The results obtained by mentioned methods are reported in the form of two states which are interpreted as 'hazardous' and 'non-hazardous'. Unbalanced distribution of positive ('hazardous state') and negative ('non-hazardous state') examples is a serious problem in seismic hazard prediction. Currently used methods are still insufficient to achieve good sensitivity and specificity of predictions.

In the paper (Bukowska M.: The probability of rockburst occurrence in the Upper Silesian Coal Basin area dependent on natural mining conditions. Journal of Mining Sciences, 42(6), 2006, 570-577) a number of factors having an effect on seismic hazard occurrence was proposed, among other factors, the occurrence of tremors with energy > 10ˆ4J was listed.

In the data set each row contains a summary statement about seismic activity in the rock mass within one shift (8 hours). If decision attribute has the value 1, then in the next shift any seismic bump with an energy higher than 10ˆ4 J was registered.

That task of hazards prediction bases on the relationship between the energy of recorded tremors and seismoacoustic activity with the possibility of rockburst occurrence. Hence, such hazard prognosis is not connected with accurate rockburst prediction.

The presented data set is characterized by unbalanced distribution of positive and negative examples. In the data set there are only 170 positive examples representing class 1.

```
##        id           seismic          seismoacoustic        shift
##  Min.   :   1.0   Length:2584        Length:2584        Length:2584
##  1st Qu.: 646.8   Class :character   Class :character   Class :character
##  Median :1292.5   Mode  :character   Mode  :character   Mode  :character
##  Mean   :1292.5
##  3rd Qu.:1938.2
##  Max.   :2584.0
##     genergy            gpuls           gdenergy          gdpuls
##  Min.   :    100   Min.   :   2.0   Min.   : -96.00   Min.   :-96.000
##  1st Qu.:  11660   1st Qu.: 190.0   1st Qu.: -37.00   1st Qu.:-36.000
##  Median :  25485   Median : 379.0   Median :  -6.00   Median : -6.000
##  Mean   :  90242   Mean   : 538.6   Mean   :  12.38   Mean   :  4.509
##  3rd Qu.:  52832   3rd Qu.: 669.0   3rd Qu.:  38.00   3rd Qu.: 30.250
##  Max.   :2595650   Max.   :4518.0   Max.   :1245.00   Max.   :838.000
##    ghazard              nbumps          nbumps2          nbumps3
##  Length:2584        Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  Class :character   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Mode  :character   Median :0.0000   Median :0.0000   Median :0.0000
##                     Mean   :0.8595   Mean   :0.3936   Mean   :0.3928
##                     3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
##                     Max.   :9.0000   Max.   :8.0000   Max.   :7.0000
##     nbumps4           nbumps5             nbumps6     nbumps7     nbumps89
##  Min.   :0.00000   Min.   :0.000000   Min.   :0   Min.   :0   Min.   :0
##  1st Qu.:0.00000   1st Qu.:0.000000   1st Qu.:0   1st Qu.:0   1st Qu.:0
##  Median :0.00000   Median :0.000000   Median :0   Median :0   Median :0
##  Mean   :0.06772   Mean   :0.004644   Mean   :0   Mean   :0   Mean   :0
##  3rd Qu.:0.00000   3rd Qu.:0.000000   3rd Qu.:0   3rd Qu.:0   3rd Qu.:0
##  Max.   :3.00000   Max.   :1.000000   Max.   :0   Max.   :0   Max.   :0
##      energy         maxenergy          class
##  Min.   :     0   Min.   :     0   Min.   :0.00000
##  1st Qu.:     0   1st Qu.:     0   1st Qu.:0.00000
##  Median :     0   Median :     0   Median :0.00000
##  Mean   :  4975   Mean   :  4279   Mean   :0.06579
##  3rd Qu.:  2600   3rd Qu.:  2000   3rd Qu.:0.00000
##  Max.   :402000   Max.   :400000   Max.   :1.00000
```

```
##   id seismic seismoacoustic shift genergy gpuls gdenergy gdpuls ghazard nbumps
## 1  1       a              a     N   15180    48      -72    -72       a      0
## 2  2       a              a     N   14720    33      -70    -79       a      1
## 3  3       a              a     N    8050    30      -81    -78       a      0
## 4  4       a              a     N   28820   171      -23     40       a      1
```

```
## 5  5       a             a     N   12640    57      -63    -52       a      0
## 6  6       a             a     W   63760   195      -73    -65       a      0
##   nbumps2 nbumps3 nbumps4 nbumps5 nbumps6 nbumps7 nbumps89 energy maxenergy
## 1       0       0       0       0       0       0        0      0         0
## 2       0       1       0       0       0       0        0   2000      2000
## 3       0       0       0       0       0       0        0      0         0
## 4       0       1       0       0       0       0        0   3000      3000
## 5       0       0       0       0       0       0        0      0         0
## 6       0       0       0       0       0       0        0      0         0
##   class
## 1     0
## 2     0
## 3     0
## 4     0
## 5     0
## 6     0
```

1.4 Arribute Information

- 1. seismic: result of shift seismic hazard assessment in the mine working obtained by the seismic method (a - lack of hazard, b - low hazard, c - high hazard, d - danger state);
- 2. seismoacoustic: result of shift seismic hazard assessment in the mine working obtained by the seismoacoustic method;
- 3. shift: information about type of a shift (W - coal-getting, N -preparation shift);
- 4. genergy: seismic energy recorded within previous shift by the most active geophone (GMax) out of geophones monitoring the longwall;
- 5. gpuls: a number of pulses recorded within previous shift by GMax;
- 6. gdenergy: a deviation of energy recorded within previous shift by GMax from average energy recorded during eight previous shifts;
- 7. gdpuls: a deviation of a number of pulses recorded within previous shift by GMax from average number of pulses recorded during eight previous shifts;
- 8. ghazard: result of shift seismic hazard assessment in the mine working obtained by the seismoacoustic method based on registration coming from GMax only;
- 9. nbumps: the number of seismic bumps recorded within previous shift;
- 10. nbumps2: the number of seismic bumps (in energy range [10^2, 10^3)) registered within previous shift;
- 11. nbumps3: the number of seismic bumps (in energy range [10^3, 10^4)) registered within previous shift;
- 12. nbumps4: the number of seismic bumps (in energy range [10^4, 10^5)) registered within previous shift;
- 13. nbumps5: the number of seismic bumps (in energy range [10^5, 10^6)) registered within the last shift;
- 14. nbumps6: the number of seismic bumps (in energy range [10^6, 10^7)) registered within previous shift;

- 15. nbumps7: the number of seismic bumps (in energy range [10^7, 10^8)) registered within previous shift;
- 16. nbumps89: the number of seismic bumps (in energy range [10^8, 10^10)) registered within previous shift;
- 17. energy: total energy of seismic bumps registered within previous shift;
- 18. maxenergy: the maximum energy of the seismic bumps registered within previous shift;
- 19. class: the decision attribute - '1' means that high energy seismic bump occurred in the next shift ('hazardous state'), '0' means that no high energy seismic bumps occurred in the next shift ('non-hazardous state').

1.5 Variables Information

There are totally 18 input variables (attributes) and 1 binary output variable (class) in the data set. The below table summarize the information of variables.

Table 1: Variable Summary Table

variable	Cardinality	Filled	Nulls	Total	Uniqueness
class	2	2584	0	2584	0.0
energy	242	2584	0	2584	0.1
gdenergy	334	2584	0	2584	0.1
gdpuls	292	2584	0	2584	0.1
genergy	2212	2584	0	2584	0.9
ghazard	3	2584	0	2584	0.0
gpuls	1128	2584	0	2584	0.4
id	2584	2584	0	2584	1.0
maxenergy	33	2584	0	2584	0.0
nbumps	10	2584	0	2584	0.0
nbumps2	7	2584	0	2584	0.0
nbumps3	7	2584	0	2584	0.0
nbumps4	4	2584	0	2584	0.0
nbumps5	2	2584	0	2584	0.0
nbumps6	1	2584	0	2584	0.0
nbumps7	1	2584	0	2584	0.0
nbumps89	1	2584	0	2584	0.0
seismic	2	2584	0	2584	0.0
seismoacoustic	3	2584	0	2584	0.0
shift	2	2584	0	2584	0.0

Although 'maxenergy' and 'nbumps' are numeric, they have a relatively small cardinality and result in zero Uniqueness (ratio of Cardinality to Total). Therefore they are classified into catagorical variables. Below is the table that summarizing the variable type of class and each attributes.

Table 2: Variable Type

Variable	Type
class	binary
energy	numeric
gdenergy	numeric

Variable	Type
gdpuls	numeric
genergy	numeric
ghazard	catagorical
gpuls	numeric
maxenergy	catagorical
nbumps	catagorical
nbumps2	catagorical
nbumps3	catagorical
nbumps4	catagorical
nbumps5	catagorical
nbumps6	catagorical
nbumps7	catagorical
nbumps89	catagorical
seismic	catagorical
seismoacoustic	catagorical
shift	catagorical

1.6 Key Steps

2. Data Analysis

2.1 Data Cleaning

2.2 Data Exploration and Visualization

2.3 Modeling Approach

3.Results and Discussion

4. Conclusion and Future Work

4.1 Limitation

More and more advanced seismic and seismoacoustic monitoring systems allow a better understanding rock mass processes and definition of seismic hazard prediction methods. Accuracy of so far created methods is however far from perfect. Complexity of seismic processes and big disproportion between the number of low-energy seismic events and the number of high-energy phenomena (e.g. > 10^4J) causes the statistical techniques to be insufficient to predict seismic hazard Therefore, it is essential to search for new opportunities of better hazard prediction, also using machine learning methods.

Aknowledgement

Marek Sikora^{1,2} (marek.sikora '@' polsl.pl), Lukasz Wrobel^{1} (lukasz.wrobel '@' polsl.pl) (1) Institute of Computer Science, Silesian University of Technology, 44-100 Gliwice, Poland (2) Institute of Innovative Technologies EMAG, 40-189 Katowice, Poland

Reference

https://archive.ics.uci.edu/ml/datasets/seismic-bumps#