ROMANIAN JOURNAL OF INFORMATION SCIENCE AND TECHNOLOGY
Volume 23, Number S, pp. S41–S52, 2020

The Social Networks information noise in crisis situations

Mironela Pirnau[1]
[1]Titu Maiorescu University, Faculty of Informatics, Bucharest, Romania
mironela.pirnau@prof.utm.ro

Abstract. In the last few years, Social Networks (SNs) have been spreading information and messages with a dynamics that has been changing continuously. The present research highlights the fact that it is very important to know the dynamics of information in SNs, especially in crisis situations. The data used in the present research have been collected from Twitter in January 2020. These data show the fact that it is extremely useful to spread the information referring to the present and future critical and catastrophic events as soon as possible. The results of the data collected from SNs, which have been analysed as soon as possible, can lead to making more efficient the specific activities of the intervention teams. At the same time, they can lead to increasing the awareness of people in reference with future dangers. The data analysis patterns within SNs are quite varied, containing specific words of the analysed field, using both neural networks and AI systems. Irrespective of the analysis method, the informational noise remains significant and must be traced and identified in each specific case. The present study focuses on the contexts above.

Keywords: information noise, social networks, logical conditions

1. Introduction

It has become a common fact that during disasters, social media and Twitter in particular, [1][2][3][4] contribute to spreading information which are useful for the monitoring and rescue actions. In the last decade, numerous researches regarding the understanding process of complex social networks have been conducted [5][6][7][8].

In order to analyse the SNs data flow that moves during the periods of times affected by disasters, earthquakes and epidemics, it is recommended that the processing process of collected posts be done together with the process of collecting these posts in view of accelerating the process of identifying the relevant information in SNs [9]. Social media plays a vital role in accounting for the persons disappeared in disasters [10]. The resistance to disasters can be measured by developing the reference indicators for measuring and monitoring the places resistance to disasters.

The complexity of the message flow within SNs during disaster periods is a challenge for the learning techniques of machines [12][13][14][15], especially for the ones that use supervised learning. The classification methods based on neural networks used for classifying the binary and multiple tweets are more performant than the previous generation methods [16][17].

The expansion of the message size on Twitter from 140-character limit to 280 characters, along with the possibility of including emoticons, significantly contribute to improving the classification of feelings in traditional algorithms. The approach of feelings classification [18][19] contributes to knowing the feelings of people, including of those who are in a very difficult situation.

The present study is in line with the researches [20][21][22][23] and aims to identify the information noise in the flow of collected tweets. The analysis approach of the present paper is divided into six sections distributed as follows: introduction, data flow in SNs, the events occurrence probability, the filtering process of posts for identifying the information noise, the analysis of the information noise, discussions and conclusions.

2. Data flow in SNs

Because of the increasing number of social media users in the last decade, the selection of relevant messages in social networks has become quite a difficult procedure because, on the one hand, the data flow is large and, on the other hand, it repeats itself. The usage of noise filters is important so that a piece of information be relevant for a certain event or topic.

The noise filters are combinations of key words that allow the SNs monitoring information system to identify the information of the type to be found and that is relevant for the targeted topic and purpose. These specific word combinations eliminate the irrelevant results. In disaster situations, the filters applied to the messages flows in social media help both knowing the real state and identifying the place of the disaster and its possible consequences. In these situations, without the help of noise filters, social media information contains thousands of inefficient posts.

When working with noise filters, it is useful to use simple formulas in order to obtain relevant and efficient interrogations. Many researchers [7][8][9][18][21] have shown that, during crisis situations, there are many messages in social media, so that it is difficult for human operators to manage this situation. For filtering the SN messages in view of eliminating the information noise in crisis situations, it is recommended a better knowledge of disaster typology [24] and classification to further correctly elaborate information filters or even specific word dictionaries [16][22][23][24][25][26]. For this purpose, a classification of the natural dangers is presented in Table 1. These natural dangers represent the basis for the word dictionary used to filter messages in view of removing the information noise. There are numerous examples described in the specialized literature [27][28][29], which extract relevant social media information in crisis situations.

AIDR (Artificial Intelligence for Disaster Response) represents an interesting example. It is a platform created specially to perform automatic classifications of the crisis situation microblogs [28]. The objective of AIDR is represented by the classification of messages posted by people during disasters in a set of well-defined categories of information, specifically classified according to the type of the emergency: biological, geophysical, meteorological, hydrological [30] [31][32] [33], see Table 1.

Table 1. Typology and classification of natural disasters

Type of danger	Main event	Side dangers
Biological	Epidemic	viral/ bacterial/ insect/ parasitic / fungal infectious diseases.
Geophysical	Earthquake	Nervous population, High fire risk, Explosions, Tsunami.
	Volcanic eruption	Volcanic eruption, Nervous population, High fire risk.
	Landslides	Landslides
Meteorological	Storm	Tropical cyclone, Extratropical cyclone, Hailstorm, Lightnings, Tornado, Windstorms, Sandstorms, etc.
	Wildfire	Wildfire
	Extreme temperatures	Heatwaves, Cold waves and frost waves, Extreme winter conditions.
Hydrological	Flood	Floods, flash floods, Huge waves, avalanches, landslides.

3. The Event Occurrence Probability

The Words [] vector was defined taking as a starting point the concepts in paper [23] which highlight both that fact that each field has its own polar words and that the strong polar words are relatively frequent for each domain. It is used when the flow of extracting Twitter posts is opened, with the help of the application elaborated for [9].

*Wor*ds [] = {earthquake, seism, magnitude, cutremur, tsunami, volcano, terremoto, coronavirus}.

The tweets probability corresponding to the vector of the key words, *Wor*ds [], is calculated according to equation 1:

$$P(word) = \frac{Number\ of\ favorable\ Tweets}{Number\ of\ possible\ tweets} \qquad (1)$$

Let's give the name S to the source of sending information, represented by the probabilistic area {S, s, p(s)} and the scheme of distribution:

$$S = \begin{pmatrix} s_1 & s_2 \dots \dots \dots & s_n \\ p(s_1) & p(s_2) \dots \dots & p(s_n) \end{pmatrix} \qquad (2)$$

where s is the letter of the source alphabet, $p(s_n)$ the probability that in the source to appear the symbol s_n and $\sum_{i=1}^{n} p(s_i) = 1$ (*n* represents the number of symbols from the source). Then the values that characterize the source of information are:

- Shannon Entropy according to equation 3:

$$H(s) = -\sum_{i=1}^{n} p(s_i) * log\, p(s_i) \qquad (3)$$

When Entropy is bigger it implies a greater diversity, but when Entropy is zero, then the result is certain.

- Maximum Entropy $H(s)_{max} = log_2 n$;
- Source efficiency= $\frac{H(s)_{max}}{H(s)}$;
- Source redundancy = $H(s)_{max} - H(s)$.

For the calculation process of these values, the code sequence designed on php and displayed in listing 1.

Listing 1. The calculation of the values/measures/amounts that characterize the source of information

```
<?php
…
function entropy($tweet)
 {
  $tweet=strtoupper($tweet);
  $H=0;$contor=0;
  $size = strlen($string);
  foreach (count_chars($string, 1) as $s)
    {
     $P = $s/$size; $contor++;
     $H -= $P*log($P)/log(2);
    }
  $Hmax=log($contor)/log(2);
  $efficient_source= $H/$Hmax;
  $redundant_source=$Hmax-$H;
  return array($H, $Hmax,$efficient_source, $redundant_source) ;
  }
$result = entropy($string);
echo "<br> entropia =".$result[0];
echo "<br> entropia maxima =".$result[1];
echo "<br> efficient_sources =".$result[2];
echo "<br> redundant_source =".$result[3];
?>
```

4. The filtering process of posts for identifying the information noise for the collected data

In the present study, a part of the data flow (the tweets) used for the analysis was extracted with two independent soft systems: the application presented in paper [9] and using AIDR (Artificial Intelligence for Disaster Response http://aidr.qcri.org/) which is a free of charge and also an open source software that automatically collects and classifies the tweets that are posted during the humanitarian crises, by combining human intelligence with artificial intelligence. In order to identify the information noise, after extracting and selecting the characteristics of the messages, specialized automatic processing algorithms are applied. The AIDR system continuously extracts data from Twitter using classification techniques of automatic learning. The categories used for filtering the Twitter posts are those in Listing 2.

Listing 2. Types of classifications for the tweets extracted by using AIDR

Natural Hazard: Geophysical: Earthquake and/or Tsunami
Natural Hazard: Geophysical: Volcano
Natural Hazard: Geophysical: Dry mass movement (rockfall, avalanche, landslide, subsidence)
Natural Hazard: Meteorological: Storm (tropical, cyclone, tornado, blizzard, dust storm)
Natural Hazard: Hydrological: Flood (river-, flash-, coastal)
Natural Hazard: Hydrological: Wet mass movement (rockfall, avalanche, landslide, subsidence)
Natural Hazard: Climatological: Extreme temperature (heat/cold wave)
Natural Hazard: Climatological: Drought
Natural Hazard: Climatological: Fire (forest, bush, grass, wild)
Natural Hazard: Biological: Epidemic (diseases, insects, animals)
Natural Hazard: Other
Human Induced: War or armed conflict, incl. acts of war
Human Induced: Terrorist attack against civilians
Human Induced: Demonstrations (peaceful or violent, riots)
Human Induced: Pollution (hazardous material, oil spill, etc.)
Human Induced: Radiation, including nuclear explosion
Human Induced: Transportation accident (train, boat, plane)
Human Induced: Other

Taking as a starting point the concepts discussed in [23], which highlight both the fact that each specialty has its own polar words and that the strong polar words are relatively frequent in each field, the word vector *Wor*ds [] was defined. It was used when opening the flow of extracting the Twitter posts, by means of the application designed for the research [9].

*Wor*ds [] = {earthquake, seism, magnitude, cutremur, tsunami, volcano, terremoto, coronavirus}

Taking as a starting point the vector Words [], two vectors with adjacent words were defined:

- **Ad_words1** {earthquake, magnitude, M number, KM, Cuba, Jamaica, tsunami. seism, volcano, disaster, terremoto, dead} and
- **Ad_word2** {coronavirus, disaster, virus, panic, illness, dead, infected, fears, spread, people}

by means of which logical conditions for filtering [7][8][9] were created. They were also used for identifying both tweets without information noise and tweets containing information noise. In order to find unique tweets with information noise and making use of the theoretical concepts mentioned in the papers [7][8][9][25], we set forth the interface used to define the conditions (shown in Fig. 1). The information must be extracted from SM using logical conditions, to eliminate the information noise of the posts.

Examples of logical conditions used in the analysis of the daily flow of posts:

R1 = earthquake **and** magnitud
R2 = earthquake **and** (Cuba or Jamaica)
R3 = earthquake **and** tsunami
R4 = earthquake **and** km
R5 = earthquake **and** m+numar

R6 = earthquake **and** disaster
R7 = earthquake **and** volcano
R8 (Informational noise) = earthquake **and not** (R1 **or** R2 **or** R3 **or** R4 **or** R5 **or** R6 **or** R7)

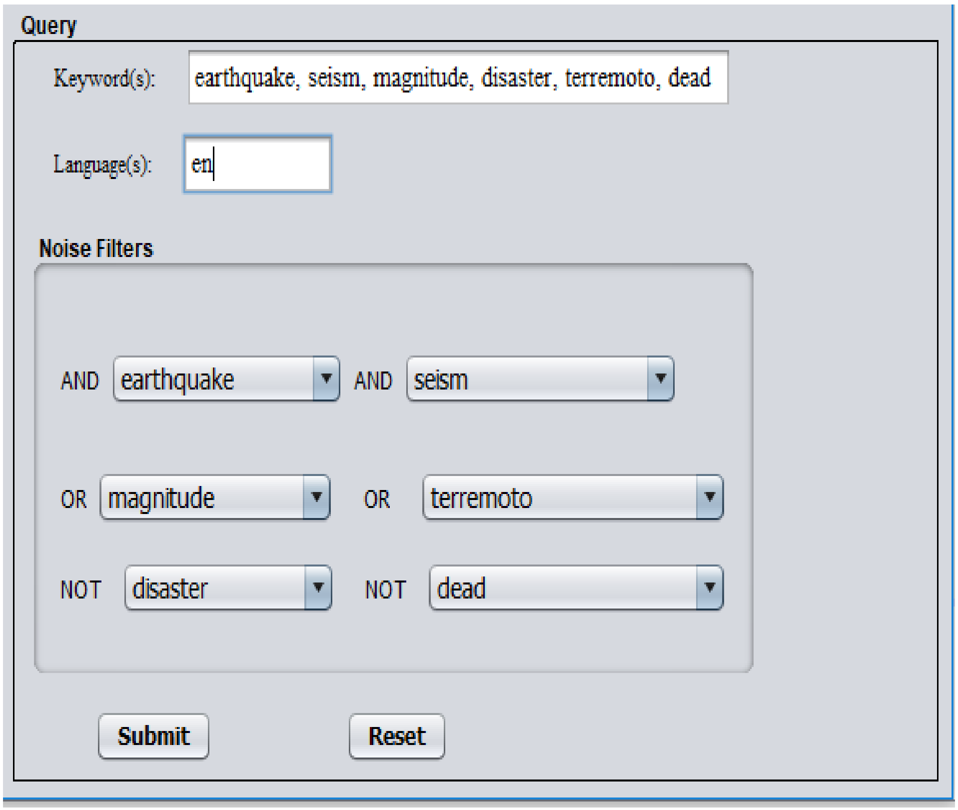

Fig. 1. Graphical interface with logical conditions

However, sometimes the logical conditions for the filtering process become too complex to be implemented. This fact happens because of the messages with imprecise meaning in the social network; the grammar structures; the wrong use of information in contexts; slang; different types of metaphors and emoticons [25].

A real example in which it is difficult to use the logical conditions to identify information noise, is the following: "*Coronavirus, Mexico earthquake, Volcano eruptions, Australia fire, PSL 5 Anthem, WWIII, Kobe Bryant. 2020 getting Deadly each week*". In such difficult cases, other methods are needed, for example the semantic analysis, the feeling analysis etc., which must be used together with the lexical analysis based on key words.

5. The analysis of the information noise

In the time slot [13:32:24 to 22:23:40], on 29th January 2020, tweets were extracted using the following word vector {cutremur, earthquake, seism, magnitudine, terremoto, coronavirus}, in order to emphasize the information noise. Because on 28th of January 2020, an earthquake having the magnitude of 7,7 occurred between Cuba and Jamaica and the National Weather Service of USA initially sent a tsunami warning that was later stopped, we have interested in the analysis of information noise in the Twitter posts related to this event. A number of 301,389 of tweets and retweets were collected within this period. Only 106,638 of them had unique content. There were 38,094 retweets with unique content out of a total of 231,545 retweets. There were 68,544 unique tweets analyzed which do not contain retweets. Table 2 shows the occurrence probability and the collecting speed of tweets and Table 3 shows the frequency of "earthquake" word related to the logical condition.

Table 2. Distribution of collected tweets

Type of tweets [unique/ without retweets/ with retweets]	Number of tweets	Probability of tweets	Speed [tweets per second]
tweets	301,389	100%	4.47
tweets without retweets	69,844	23%	1.04
tweets with retweets	231,545	77%	3.44
unique tweets	106,638	35%	1.58
unique tweets without any retweets	68,544	23%	1.02
unique tweets with retweets	38,094	13%	0.57

The corresponding diagrams (probability of collected tweets and the spread of identified/analyzed logical conditions) are presented in Fig. 2 and Fig 3.

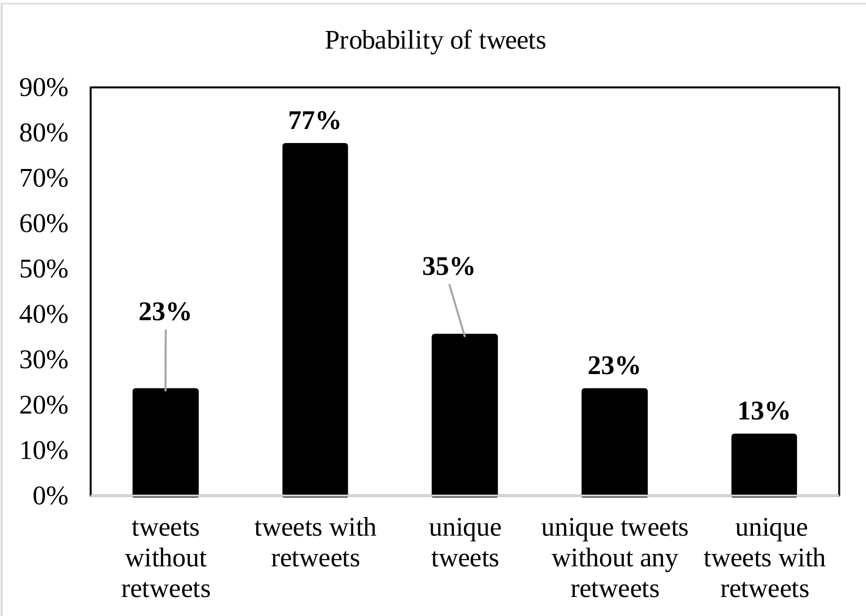

Fig. 2. Probability of collected tweets

If the following are considered as noise: tweets that contain "RT @"; tweets that do not have unique content then the probability for one tweet not to represent content noise is about 12% and the occurrence speed for a possible tweet without noise is 0,5 tweets per second. Therefore, a number of 29,827 tweets which contain none of the elements of the word vector **Ad_words1** was identified. It results that the probability for unique tweets to be information noise is about 44%. Nevertheless, this statement must be correlated with the content analysis of short links in these posts, using a content parsing tool [34] which is applied to the short tweets in the posts. Because Twitter doubled the lengths of posted messages, the classical methods of identifying the information noise have to take into account the content parsing of short links in messages. For the following condition: *… where tweets like '%earthquake%' or tweets like '%seism%' and tweets not like '%coronavirus%'* 1433 posts were found in the unique set of tweets analysed in the present study. Using the program with the code sequence displayed in listing 1, we have determined the entropy of the information source with its characteristic values, for the R1-R7 rules, in Table 4.

Table 3. The frequency of "earthquake" word related to the logical condition

Rule	No. of posts	Probability % of tweets that contain earthquake
R1	315	21.98%
R2	292	20.38%
R3	62	4.33%
R4	265	18.49%
R5	42	2.93%
R6	65	4.54%
R7	74	5.16%
R8	318	22.19%
Total	1433	100%

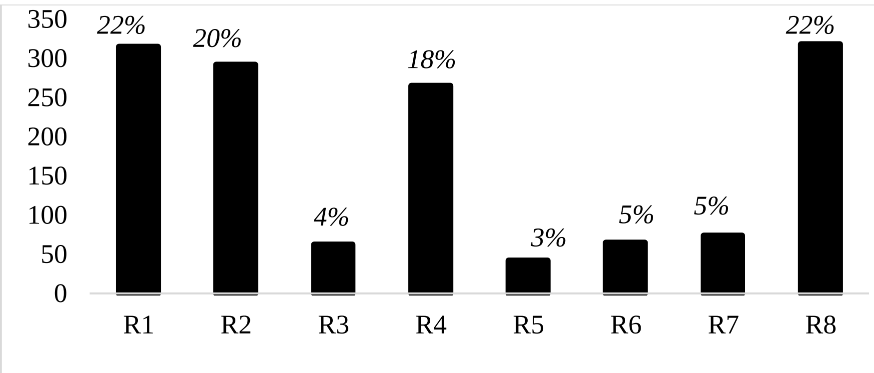

Fig. 3. The spread of identified/analyzed logical conditions

Table 4. Sizes that characterize the information source

Rule	H (entropy)	H max	Source efficiency	Redundant source
R1	3.64	3.81	96%	0.17
R2	3.886	4.248	91%	0.362
R3	3.630	3.807	95%	0.177
R4	3.455	3.585	96%	0.130
R5	3.787	3.907	97%	0.120
R6	3.534	3.700	96%	0.166
R7	3.732	3.907	96%	0.175

The size of 'source efficiency' for R1-R7 rules is shown in Fig. 4. One can observe that the values obtained are over 90% for all the identified rules, which points to their precision. In accordance with these values, while the source efficiency increases and is closer to value 1, the redundancy decreases and is closer to 0, see Table 4.

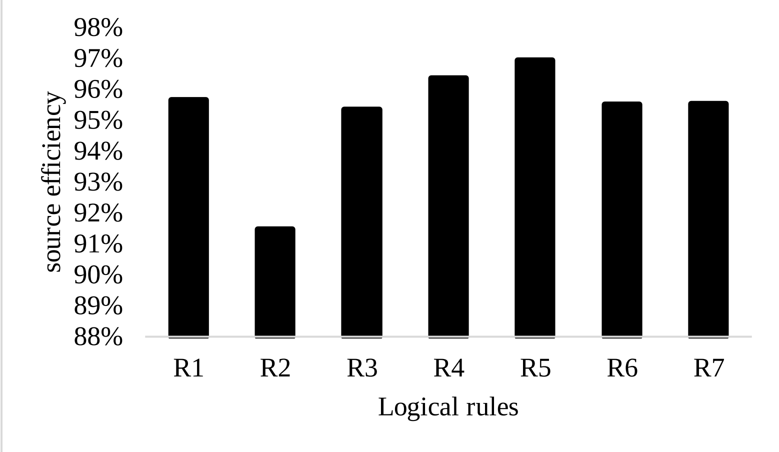

Fig. 4. Logical rules and the source efficiency

Starting from the rule R1 {*earthquake and magnitude*}, only the 7.7 magnitude tweets from the collected data were selected. Thus, a number of 48 tweets were found. Taking into account the fact that the similarity of reported contents for a period of time [35] is a problem in the case of Social Media data analysis, these posts were divided into two equal groups in chronological manner, with the aim of identifying the editing distance among posts. In relevant literature, this process is known as the Levenshtein distance, which represents a measure of similarity among strings [36]. For these two groups of posts, the similarity was calculated using an open source instrument - Fuzzy Lookup Add-In for Excel [37]. According to the calculations graphed in Fig. 5, there are 10 posts for which the value of the similarity is 0 but there are also 2 posts for which the value of the similarity is 1, even if the posts are unique.

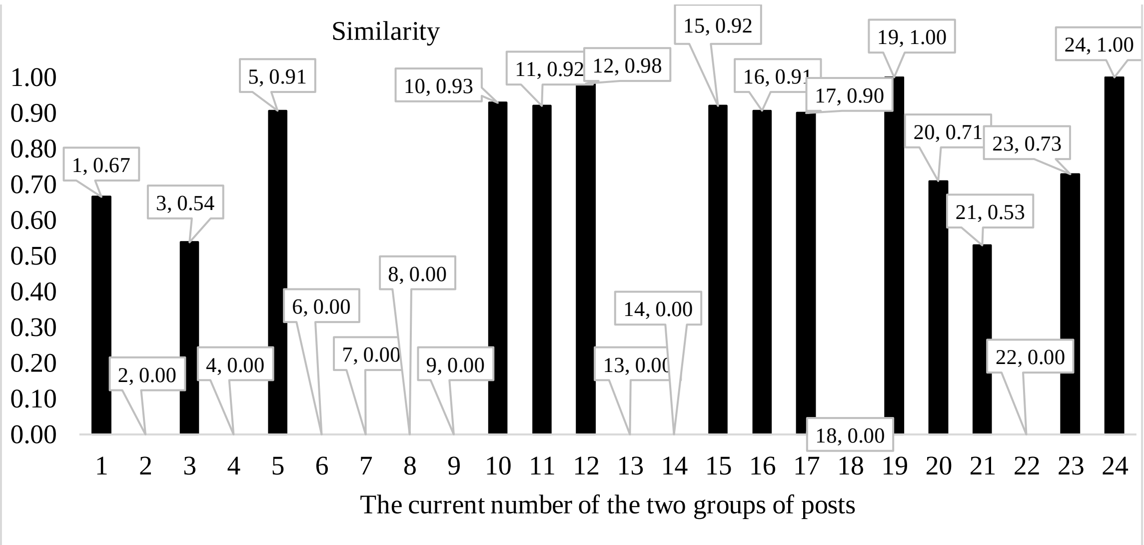

Fig. 5. Similarity of posts for rule R1 and Magnitude 7.7

Even if the posts were selected as being unique, in Table 5 it is noticed that the similarity of posts can be considered as noise. In a disaster situation, the time of posts analysis increases.

Since the crisis situations do not necessarily include only earthquakes and natural calamities, the AIDR (Artificial Intelligence for Disaster Response http://aidr.qcri.org/) was used to extract posts from Twitter. In this case, the main targeted word was 'coronavirus', and, according to listing 2, the category, was: Natural Hazard: Biological: Epidemic (diseases, insects, animals). In this case, the application ran for 508 sec, and the speed for extracting the 1802 posts was approximately 3.5 posts/second. Out of all these posts, only 995 contain the word 'coronavirus', which implies the idea that the speed for delivering posts without noise could have been of 1.5 posts per second.

Table 5. Selective results from the calculation of posts similarity

Tweet	Created posts (dates)	Similarity
No injuries reported after magnitude 7.7 earthquake shakes Jamaica, Cuba	29 Jan 2020 15:30:06	0.728889
Magnitude 7.7 #earthquake hits between Cuba and Jamaica, no injuries	29 Jan 2020 16:30:05	
Powerful 7.7 magnitude earthquake strikes between Cuba and Jamaica	29 Jan 2020 14:13:23	0.928571
Earthquake strikes 7.7-magnitude between Jamaica and Cuba	29 Jan 2020 19:44:18	
7.7 magnitude earthquake strikes between Cuba and Jamaica	29 Jan 2020 14:32:00	1.00
Earthquake strikes 7.7-magnitude between Jamaica and Cuba	29 Jan 2020 19:44:18	

For the elements in the Ad_word2 {coronavirus, disaster, virus, panic, illness, dead, infected, fears, spread, people} vector, the occurrence probability in combined word rules was identified according to Table 6.

Table 6. Probability from the total tweets that contain coronavirus

Rule	No. of posts	Probability %
Coronavirus and infect	35	3.52%
Coronavirus and disaster	3	0.30%
Coronavirus and dead	24	2.41%
Coronavirus and panic	23	2.31%
Coronavirus and illness	3	0.30%
Coronavirus and fears	29	2.91%
Coronavirus and spread	7	0.70%
Coronavirus and virus	54	5.43%
Coronavirus and people	75	7.54%
coronavirus and not vector Ad_word2	742	74.57%
Total posts that contain 'coronavirus'	995	

One can notice that the probability for a post that contains 'coronavirus' and no other element of the Ad_words2vector is about 75%. This means that almost 75% of tweets may be information noise. If the posts that do not contain combinations of the elements belonging to Ad_word2 vector are information noise, then the speed with tweets that are not information noise and that are being delivered by Twitter is about 0.49 posts/second.

6. Discussion and Conclusions

This present paper demonstrates that the posts in the social networks during crisis periods produce a lot of information noise, regardless the situations that generate the crisis. These situations may include: earthquakes, epidemics etc. The calculations in the present paper show that Twitter information noise is not caused only by “RT @”. In real crisis situations, the information noise becomes considerably bigger than the real information flow. The noise decreases the message analysis speed because it increases the volume of information very much. For the period of time taken into account in the present paper, according to Table 2, it can be noticed that the possibility of one post to be noise is strong, so the elimination process of the information noise in posts represents a considerable effort and it must be performed before the analysis of useful tweets. The fact that the posts are similar must also be taken into account, alongside information noise. The similarity of posts makes longer the time needed for identifying the data to be analyzed. In crisis situations, it is interesting to analyze, together with the values identified in the present paper, the feeling and the opinion of the messages which may contribute to detecting some noise messages (the increase of the detected information noise) and to diminishing the real flow of data to be analyzed.

Acknowledgements. I thank Prof. H.N Teodorescu for his ideas on the manner of analyzing the data in this paper, as well as for the resources he made available for me.

References

[1] Laylavi F., Rajabifard A., Kalantari M. *'Event relatedness assessment of Twitter messages for emergency response'. Information Processing and Management.* 2017; vol. 53, no. 1, pp. 266–280.

[2] Takahashi B., Tandoc E.C., Carmichael C. *'Communicating on Twitter during a disaster: an analysis of tweets during typhoon Haiyan in the Philippines'. Computers in Human Behavior.* 2015; vol. 50, pp. 392–398.

[3] Houston J.B., Hawthorne J., Perreault M.F. et al. *'Social media and disasters: a functional framework for social media use in disaster planning, response, and research'. Disasters.* 2014; vol. 39, pp. 1–22.

[4] Eriksson M., Olsson E.K. *'Facebook and Twitter in crisis communication: A comparative study of crisis communication professionals and citizens', J. Contingencies and Crisis Management*. 2016; vol. 24, no. 4, pp. 198–208.

[5] Lü L, Zhou T. *'Link prediction in complex networks: A survey'. Physica A: Statistical Mechanics and its Applications*. 2011; vol. **390, no.** 6, pp. 1150-1170.

[6] Banujan K., Banage Kumara T. G. S., Paik I. *'Twitter and online news analytics for enhancing post-natural disaster management activities'. Proceedings of the 9th International Conference on Awareness Science and Technology (iCAST), Fukuoka*. 2018; pp. 302-307.

[7] Teodorescu H.N.L., Bolea S.C. '*On the algorithmic role of synonyms and keywords in analytics for catastrophic events'. Proceedings of the 8th International Conference on Electronics, Computers and Artificial Intelligence. ECAI;* Ploiesti, Romania, Jun 30-Jul 02, 2016, pp. 1–6.

[8] Teodorescu H.N.L. *'Emergency-Related, Social Network Time Series: Description and Analysis'*, Rojas I., Pomares H. (Eds.) *Time Series Analysis and Forecasting. Contributions to Statistics*. Springer, Cham, 2016, pp. 205–215.

[9] Teodorescu H.N.L., Pirnau M. '*Analysis of requirements for SN monitoring applications in disasters - a case study'. Proceedings of the 8th International Conference on Electronics, Computers and Artificial Intelligence. (ECAI);* Ploiesti, Romania, Jun 30-Jul 02, 2016, pp. 1–6.

[10] Alexander D. 'Disaster and emergency planning for preparedness, response, and recovery'. Oxford Research Encyclopedia of Natural Hazard Science. *Oxford University Press*, Oxford, 2015, pp. 1–20.

[11] Cutter S. L., Burton C. G., Emrich C. T., *'Disaster resilience indicators for benchmarking baseline conditions'. Journal of Homeland Security and Emergency Management*. 2010; vol. 7, no. 1, pp. 1-24.

[12] Asencio-Cortés G., Martínez-Álvarez F., Morales-Esteban A., Reyes J., 'A sensitivity study of seismicity indicators in supervised learning to improve earthquake prediction. Knowledge-Based Systems'. 2016; vol 101, pp. 15–30.

[13] Najeebullah, Zameer A., Khan A., Javed S.G., *'Machine Learning based shortterm wind power prediction using a hybrid learning model'*. Comput Electr Eng. 2015; vol. 45(C), pp. 122–33. 10.1016/j.compeleceng.2014.07.009.

[14] Cortes C., Vapnik V., *'Support-vector networks. Machine learning'*. 1995; vol. 20, no. 3, pp. 273–297.

[15] Song X., Liu Z., Yang X., Yang J., *'A fuzzy supervised learning method with dynamical parameter estimation for nonlinear discriminant analysis', Computers & Mathematics with Applications, 2013, vol 66, no. 10, pp. 1782-1794.*

[16] Mahmoudi N., Docherty P., Moscato P., *'Deep neural networks understand investors better'*. Decis. Support Syst. 2018; vol. 112, pp. 23–34.

[17] Rahman N.H.A., Lee M.H., Latif M.T., *'Artificial neural networks and fuzzy time series forecasting: An application to air quality'*. Qual. Quant. 2015; vol. 49, pp. 1–15.

[18] Bolea S.C., *'Vocabulary, synonyms and sentiments of hazard-related posts on social networks'. Proceedings of the 8th Conference Speech Technology and Human-Computer Dialogue (SpeD);* Bucharest, Oct 14-17, Romania 2015, pp. 1-6.

[19] Teodorescu H.N.L., Saharia N., *'An Internet slang annotated dictionary and its use in assessing message attitude and sentiments'. Proceedings of the 8th Speech Technology and Human-Computer Dialogue (SpeD).* Bucharest, Romania, 2015, pp. 1-8.

[20] Pirnau M., *'Analysis of data volumes circulating in SNs after the occurrence of an earthquake'*, *Romanian Journal of Information Science and Technology*. 2017; vol. 20, no. 3, pp. 286–298.

[21] Teodorescu H.N.L., *'Social signals and the ENR index—noise of searches on SN with keyword-based logic conditions*'. In: Proceedings of the IEEE Symposium ISSCS, Iasi (2015) (978-1-4673-7488-0/15, 2015 IEEE).

[22] Singh J. P., Dwivedi Y. K., Rana N. P., Kumar A., Kapoor K., *'Event classification and location prediction from tweets during disasters*'. Annals of Operations Research, 2017; pp. 1–21.

[23] Haghighi A., Barzilay R., *'Event discovery in social media feeds'*. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Volume 1 of HLT '11, Association for Computational Linguistics, Stroudsburg, PA. 2011; pp.389-398.

[24] Burrell E.M , Graham A.T , Ronald R.H, *'Natural Hazards', Second Edition: Explanation and Integration,* Publisher: The Guilford Press; Second edition (April 26, 2017). ISBN-10: 1462529178 , ISBN-13: 978-1462529179.

[25] Teodorescu H.N, Saharia N., *'A semantic analyzer for detecting attitudes on SNs*', International Conference on Communications (COMM 2016) Book Series: International Conference on Communications (ICC), Bucharest, 2016; pp. 47-50. DOI: 10.1109/ICComm.2016.7528201 .

[26] Teodorescu H.N.L., 'SN voice and text analysis as a tool for disaster effects estimation - a preliminary exploration'. Proceedings of the 7th Conference Speech Technology and Human-Computer Dialogue (SpeD); Cluj Napoca, Romania, Oct 16-19, 2013; pp. 1-8. DOI: 10.1109/SpeD.2013.6682650.

[27] Kleinberg J. M., *'Navigation in a small world'*. *Nature*, 24 Aug 2000, 406, p. 845.

[28] D. T. Nguyen, K. Al-Mannai, S. R. Joty, H. Sajjad, M. Imran, and P. Mitra, *'Rapid classification of crisis-related data on social networks using convolutional neural networks'*, CoRR, vol. abs/1608.03902, 2016.

[29] Imran M., C Castillo., Diaz F., Vieweg S., *'Processing social media messages in mass emergency: A survey*', ACM Computing Surveys (CSUR), 2015; vol. 47, no. 4, pp. 1-38.

[30] Jonsson C.B., Figueiredo L.T., Vapalahti O., *'Aglobal perspective on hantavirus ecology, epidemiology, and disease'*. Clinical Micro-biology Reviews, 2010; vol, 23, pp. 412–441.

[31] Vaheri A., Henttonen H., Voutilainen L., Mustonen J., Sironen T., Vapalahti O., *'Hantavirus infections in Europe and their impact on public health*'. Rev Med Virol. 2013; vol. 23, pp. 35–49.

[32] Kilbourne E.D., *'Influenza pandemics of the 20th century*'. Emerging Infectious Diseases, 2006; vol. 12, pp. 9-14.

[33] Below R., Wirtz A., Guha-Sapir D., '*Disaster category classification and peril terminology for operational purposes'*. Common accord Centre for Research on the Epidemiology of Disasters (CRED) and Munich Reinsurance Company (Munich RE). 2009; Working Paper 264, UCL.

[34] Pirnau M., *'Tool for monitoring Web sites for emergency-related posts and post analysis'. Proceedings of the 8th Speech Technology and Human-Computer Dialogue (SpeD)*; Bucharest, Oct 14-17, Romania 2015; pp.1-6.

[35] Gusfield D., *'Algorithms on Strings, Trees, and Sequences'*, Computer Science and Computational Biology'. Cambridge, U.K.: Cambridge Univ. Press, 1997.

[36] Navarro G., *'A Guided Tour to Approximate String Matching*', ACM Computing Surveys, 2001; vol. 33, no. 1, pp. 31-88.

[37] https://www.microsoft.com/en-us/download/details.aspx?id=15011, *'Fuzzy Lookup Add-In for Excel'*.