GPLSI at TREC 2019 Incident Streams Track

Javi Fernández*, Fernando Llopis,
Patricio Martínez-Barco, José M. Gómez
{javifm,llopis,patricio,jmgomez}@dlsi.ua.es

Department of Software and Computing Systems
University of Alicante

Abstract

In this paper we present our contribution to the *TREC 2019 Incident Streams* track. We submitted four runs to the *2019-B* edition of this task. Our main goal is to evaluate the effectiveness of sentiment analysis and information retrieval techniques to automatically detect and prioritize incidents on social media streams. Here, we describe these techniques and show the results obtained.

1 Introduction

Emergency situations must be responded as quickly and effectively as possible. Social networks allow anyone to contact emergency operators, so they can respond almost instantly. However, due to the large volume of information published on these platforms, it is very difficult to categorize, cross and verify that information in real time. This is why automatic tools that effectively speed up monitoring this data are gaining interest. The *Incident Streams*[1] (IS) track [6] of the *Text Retrieval Conference*[2] (TREC) focuses on this need.

In this paper we describe the techniques employed to identify and prioritize relevant tweets, among which we can highlight the use of *sentiment analysis* and *information retrieval* to help this task. The detailed approaches are described in Section 2. Subsequently, in Section 3 we show the assessment of our model. Finally, the conclusions and future work are presented in Section 4.

*Corresponding author
[1] http://trecis.org
[2] https://trec.nist.gov

2 Methodology

Baseline (*run1_baseline*)

In summary, this approach detects which information types are mentioned in a tweet, giving a greater score to those tweets with a greater number of information types related. All our executions have this this base in common. The detailed process is described as follows. For each tweet t in the collection of a topic, it assigns a weight $b_{t,i}$ to each information type i in the set of information types I, depending on their similarity. This weight is calculated as shown in Equation 1, where $lls_{t,i}$ is the similarity of the tweet t and the low level types of the information type i, $tis_{t,i}$ is the similarity of the tweet t and the words in the id of the information type i and $tds_{t,i}$ is the similarity of the tweet t and the description of the information type i. Finally, all these weights are merged into one single score B_t, which will be assigned to the tweet t, using the formula in Equation 2, where $norm(x)$ is a function that normalizes the value of x in the range $[0, 1]$ using the formula $norm(x) = x/(x + 1)$.

$$b_{t,i} = 0.7 \cdot lls_{t,i} + 0.2 \cdot tis_{t,i} + 0.1 \cdot tds_{t,i} \quad (1)$$

$$B_t = norm(\sum\nolimits_{i \in I} b_{t,i}) \quad (2)$$

Negative Filter (*run2_negative*)

This run adds *sentiment analysis* to the baseline run. The addition consists on detecting the sentiment of the tweets, and assigning a score of 0 to those that have a positive polarity. This is, only negative tweets can have a score greater than 0. The hypothesis of this execution is that when people have an emergency, the polarity of their texts tends to be more negative, so it could be a good filter to get the most priority tweets. The sentiment analysis approach used in this run is the one described in [1, 2, 3, 4] due to its good results in previous works.

IR-n Weighting (*run3_irn*)

Again, this run uses the baseline techniques described previously, but uses the weighting schema of the information retrieval system *IR-n* [5, 7]. The weight $b_{t,i}$ for each tweet t and information type i depending on their similarity, is improved adding the similarity given by this system. The formula in Equation 1 is then modified to include this similarity, shown in Equation 3, where $irns_{t,i}$ is the similarity of the tweet t and the information type i given by the IR-n system.

$$b_{t,i} = 0.35 \cdot lls_{t,i} + 0.1 \cdot tis_{t,i} + 0.05 \cdot tds_{t,i} + 0.5 \cdot irns_{t,i} \quad (3)$$

Negative Filter and IR-n Weighting (*run4_all*)

This run combines the techniques used in *run2_negative* and *run3_all*, scoring only negative tweets, and calculating the similarity using Equation 3.

3 Results

These are the measures used in the track:

- *AAWH*. Accumulated Alert Worth - High Priority
- *AAW*. Accumulated Alert Worth - All
- *ITFA*. Information Type F1 - Actionable
- *ITF*. Information Type F1 - All
- *ITA*. Information Type Accuracy
- *PEH*. Priority RMSE - High Priority
- *PE*. Priority RMSE - All

The results obtained by our system are shown in Table 1. As all our runs have slight differences with the baseline, they obtain very similar results (all of them on the track median or below).

	AAWH	AAW	ITFA	ITF	ITA	PEH	PE
Track median	-.9197	-.4609	.0386	.1055	.8583	.1767	.1028
run1_baseline	-.9784	-.4895	**.0386**	.0000	.8753	**.2067**	**.1150**
run2_negative	-.9784	**-.4894**	.0377	.0000	**.8758**	.2075	.1154
run3_irn	**-.9794**	-.4897	**.0386**	.0000	.8753	.2132	.1175
run4_all	**-.9794**	-.4897	.0377	.0000	**.8758**	.2138	.1177

Table 1: GPLSI TREC 2019-B Incident Streams Track results

4 Conclusions and Future Work

This was the first participation of the *GPLSI* of the *University of Alicante* in the Incident Streams Track, submitting four runs to the *2019-B* edition. Our main goal was to evaluate the effectiveness of sentiment analysis and information retrieval techniques in the detection and prioritization of incidents on social media streams.

Our runs did not obtain the results we expected. We believe that the main reason for this result is to have given more importance to those tweets that had a greater number of related information types. Instead, we observed that a tweet can have a high priority even if it is related to only one information type. In future editions of this track (and similar tracks), we plan to change the baseline to correct this mistake. But we still believe that there is room for improvement by using sentiment analysis in this kind of tasks.

Acknowledgements

This research work has been partially funded by the University of Alicante (Spain), Generalitat Valenciana and the Spanish Government through the projects *Tecnologías del lenguaje humano para una sociedad inclusiva igualitaria y accesible* (PROMETEU/2018/089), *LIVING-LANG: Modelado del comportamiento de entidades digitales mediante tecnologías del lenguaje humano* (RTI2018-094653-B-C22) and *INTEGER: Intelligent Text Generation*, *Generación inteligente de textos* (RTI2018-094649-B-I00).

References

[1] Fernández, J., Gutiérrez, Y., Gómez, J. M., Martínez-Barco, P., Montoyo, A., and Munoz, R. Sentiment Analysis of Spanish Tweets using a Ranking Algorithm and Skipgrams. *XXIX Congreso de la Sociedad Española de Procesamiento de lenguaje natural (SEPLN 2013)* (2013).

[2] Fernández, J., Llopis, F., Gutiérrez, Y., Martinez-Barco, P., and Diez, A. Opinion Mining in Social Networks versus Electoral Polls. In *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017* (2017), pp. 231–237.

[3] Fernández, J., Llopis, F., Gutiérrez, Y., Martínez-Barco, P., Gómez, J. M., and Muñoz, R. GPLSI at TREC 2018 RTS Track. In *TREC* (2018).

[4] Fernández, J., Llopis, F., Martínez-Barco, P., Gutiérrez, Y., and Díez, Á. Analizando opiniones en las redes sociales. *Procesamiento del Lenguaje Natural 58* (2017), 141–148.

[5] Llopis, F. IR-n un sistema de Recuperación de Información basado en pasajes. *Procesamiento del Lenguaje Natural*, 30 (2003), 127–128.

[6] McCreadie, R., Buntain, C., and Soboroff, I. TREC Incident Streams: Finding Actionable Information on Social Media. *Proceedings of the 16th International Conferenc e on Information Systems for Crisis Response and Management, Valencia, Spain* (2019).

[7] Vicedo, J. L., Llopis, F., and Ferrández, A. University of Alicante Experiments at TREC 2002. *NIST SPECIAL PUBLICATION SP*, 251 (2003), 595–602.