Automatic Identification of Assistance Needs in Disaster Situations Using Hybrid Word Embedding Techniques

Sarra Chaiir
ENSI, RIADI LR99ES26,
Campus Univ Manouba,
2010, Tunisia
sarra.chaiir@ensi-uma.tn

Malika Charrad
EFREI Paris,
Panthéon-Assas,Univ France,
94800, France
malika.charrad@gmail.com

Narjès Bellamine Ben Saoud
ENSI, RIADI LR99ES26,
Campus Univ Manouba,
2010, Tunisia
narjesbbs@gmail.com

Abstract

Social media platforms have become increasingly important channels for affected people to request urgent assistance during disaster situations. However, given the huge amount of data shared on these platforms, it has proved challenging to identify the specific types of assistance needed. The reason behind this challenge lies in the complexity of the task of designing and representing features from existing data. Herein, we propose to combine two types of feature extraction techniques, namely syntactic and semantic, using TF-IDF, Glove, and FastText, in order to extract highly discriminating features from tweets relating to requests for assistance. The purpose of this study is to evaluate the effectiveness of different feature extraction techniques and their combination when applied to various classifiers. We will perform this evaluation using the Purohit dataset, which consists of Hurricane Sandy tweets classified into six different classes. Experimental results show that the TF-IDF combined with FasTtext, fed into an SVM classifier outperformed the other models in terms of accuracy (86.11%) and F1-score (84.38%). According to this study, combining feature extraction techniques can improve machine learning performance.

1 Introduction

Over the past few decades, we have witnessed an appalling increase in the frequency of natural disaster events around the world. According to the UN Office for Disaster Risk Reduction, more than 4.5 billion people have been affected by natural disasters between 1998 and 2017 [1]. During such events, social media platforms has been playing a crucial role to help survivors communicate their needs. People can resort to these platforms to share damaged situations and make urgent requests for help. According to (Villegas et al., 2018), more than 5,200 rescue requests made on social media were missed by the first responders, as well as 46% of critical damage information posted on social media was overlooked by FEMA in their first damage assessments during the Harvey Hurricane.

Assistance identification involves the recognition and categorization of help requests expressed in natural language. Within the field of Natural Language Processing (NLP), this task can be treated as a text classification problem that aims to assign predefined categories or labels to text, such as tweets. In this context, specific categories can be defined to represent various types of assistance requests, including 'shelter', 'food', 'clothing','money', etc. With the ever-increasing volume of information being shared, government agencies are struggling to respond to the public's needs in a timely manner. As a result, there is a pressing need to develop systems that can identify the appropriate types of support needed to enable a quick and effective response. However, designing a features that distinguishes help assistance tweets from other types of tweets is a difficult task. This challenge stems from the need to explore and use improved feature representation techniques to effectively differentiate between different categories of tweets. Several techniques have been proposed in the literature, but their combination has not been fully exploited.

Our proposal involves using two different types of information, syntactic and semantic, to contribute to a more accurate understanding and representation of tweets. In order to gather syntactic information, we use the TF-IDF technique, which involves analyzing the occurrence patterns of keywords in the training data. This technique extracts valuable statistical information that helps us to better understand the structural aspects of the text. While TF-IDF is primarily concerned with statistical properties, it can also indirectly capture some syntactic information. For instance, frequent nouns or verbs within a document may provide clues about the document's content and its likely syntactic structures.

[1] https://www.unisdr.org/

Additionally, IDF assigns higher weights to terms that are rare or distinctive across the corpus, which is crucial for capturing syntactic nuances. On the other hand, for semantic information, we employ two well-defined techniques which are Glove and FastText. These particular techniques capture contextual information, enabling us to better determine the meaning and context of the text.

In this study, we evaluate the effectiveness of different feature extraction techniques using four well-known machine learning algorithms. The objective here is to identify the most effective feature extraction techniques and the best machine learning algorithms for hybrid feature extraction.

The remainder of this article is structured as follows. Section 2 outlines the related work. Section 3 describes our framework for the assistance need prediction problem. Section 4 presents the experiments and the evaluation results. Finally, the conclusion and future works are given in section 5.

2 Related works

Recently, there has been a growing interest in leveraging social media data to enhance disaster response efforts. Several studies have been undertaken to effectively identify and detect rescue requests, as well as categorize the types of assistance needed in the aftermath of disaster events. For example, Devaraj et al.(Devaraj et al., 2020) conducted a study on identifying urgent requests published during Hurricane Harvey. They trained several machine learning models, including Convolutional Neural Network (CNN), SVM, MLP, AdaBoost, Logistic regression, Naive Bayes, Decision tree, and Ridge classifier. The results obtained showed that CNN and SVM were the most accurate and precise models to identify urgent tweets from other disaster messages.

Wang et al. (Wang et al., 2022) developed a new method for identifying rescue requests from Twitter messages using geographic features, specifically ZIP codes. They discovered that ignoring the significance of ZIP codes as a feature could hinder the development of a better machine learning model for accurately detecting rescue requests and the retrieval of training data. This research has also revealed that all machine learning classifiers, except kNN and NB, were effective in distinguishing rescue requests from other messages.

Purohit et al. (Purohit et al., 2014) developed a machine learning method to automatically identify and match needs and offers for resources published through social media. In their work, they used a binary classifier that leverages n-grams and a set of regular expressions to classify tweets as either requests or offers. Then, they match requests with sitable offers by computing the cosine similarity of the tf-idf term vectors.

Ullah et al. (Ullah et al., 2021) proposed a system coined RweetMiner which aims to identify and categorize rescue requests on Twitter during catastrophic events. The proposed system used rule-based approach and machine learning algorithms to classify tweets that contain requests into six categories: food, medical care, shelter, clothing, money, and volunteer services. The authors of this work, emphasized that RweetMiner can be useful in disaster relief operations, as it effectively detects and organizes aid requests on Twitter.

Kabir and Madria (Kabir and Madria, 2019) created a method to improve disaster rescue scheduling. They used a combination of an attention-based Bi-directional LSTM and CNN with feature engineering to classify Twitter data into six categories such as rescue needed, water needed, injured, sick, flood, and disabled elderly children and women. They also developed a hybrid scheduling algorithm that takes into account resource limitations and different rescue priorities to efficiently manage rescue operations during a disaster.

Nurdeni et al.(Nurdeni et al., 2022) presented a system that can identify the types of assistance for victims after a natural disaster in Indonesia, and display their location on a map-based dashboard. This study used text mining techniques and multiple machine learning algorithms, including Naive Bayes, Support Vector Machine, and Logistic Regression, to analyze Twitter data and extract geographical information. The best model is achieved using a combination of Support Vector Machine and Classifier Chain with UniGram+BiGram features extraction, which achieved 82% precision, 70% recall, and 75% F1-score.

3 Methodology

The purpose of this work is to precisely identify the types of assistance requests published on social media platforms. As depicted in Figure 1, our proposed framework involves four distinct stages. In the first stage, we begin by collecting the Purohit dataset. Next, we perform data preprocessing using the natural language processing toolkit (NLTK).

Following this, in the feature extraction stage, we use a series of well-established word representation techniques, including TF-IDF, GloVe, and FastText, as well as a hybrid combination between them. The outcomes of this stage are subsequently fed into multiple classifiers for evaluation.

3.1 Data Preprocessing

Data preprocessing is a crucial step before training any model. In this work, this step involves several important tasks to clean and standardize the data, including removing non-ASCII characters and non-English tweets, converting all text to lowercase, removing punctuations and stop words, generalizing tags such as usernames, URLs, and hashtags, applying tokenization and term lemmatization to reduce words to their base or root form, then eliminating duplicate tweets. These steps help to ensure that the data is consistent, free of noise and unnecessary words, and ready to be trained by machine learning algorithms.

After processing the data, a set of 1480 tweets were obtained which are then utilized for our experiments. Given that the dataset is heavily unbalanced, we have applied an oversampling technique, which is SMOTE (Synthetic Minority Oversampling Technique) to deal with this issue. This technique generates new synthetic samples for the minority class by using a k-nearest neighbor algorithm.

3.2 Feature extraction methods

In this paper, we explore the use of two different feature extraction methods, including syntactic method (Tf-idf) and semantic methods (Glove, FastText), to train various machine learning models such as Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), and XGBoost classifer (XBoost). In addition, two combinations of TF-IDF and word embeddings were also evaluated against different models. A brief overview of these techniques is provided below:

- **Tf-IDF**: TF-IDF stands for Term Frequency-Inverse Document Frequency. It is a statistical measure used in natural language processing and information retrieval to determine the importance of a particular word in reference to a given corpus. The implementation of TF-IDF involves a two-step process. Initially, the frequency of each word in the document is determined, which is also termed as term frequency (TF). Then, the occurrence frequency of each word in the entire document, known as Inverse Document Frequency (IDF), is identified in the subsequent step. Finally, the TF-IDF value for each word is obtained by multiplying its corresponding TF value with IDF value.

- **Glove**: Glove (Pennington et al., 2014) is a type of word embedding that uses an unsupervised learning algorithm to map words into a space where the distance between them reflects their semantic similarity. This technique generates a word embedding matrix by capturing the feature-feature co-occurrence from a large corpus of text. Within our work, we used 300-dimensional GloVe word embeddings pre-trained on a large corpus of 2 billion tweets.

- **FastText**: FastText (Joulin et al., 2016) is a word embedding technique developed by Facebook that uses a skip-gram model to convert words into N-grams characters. In this study, we used a pre-trained FastText vectors to create 300-dimensional token vectors for each word in tweet. The model we used had 1 million word vectors, which were trained on Wikipedia 2017, a dataset of 1 billion tokens.

4 Experimentations and results

In this section, we evaluates the performance of the different proposed models. We first begin by describing the dataset and the performance metrics used for the evaluation. Then, we present the experimental setting and the achieved results.

4.1 Data Description

In this study, we used the dataset presented in the work of Purohit (Purohit et al., 2014) to verify the effectiveness of the proposed models. The authors of this work build a high-quality corpus containing tweets mining help requests and offers made by individuals or organizations seeking or providing assistance during the Hurricane Sandy event. This dataset was collected using the Twitter Streaming API based on a set of keywords and hashtags (e.g. #sandy), and afterward manually labeled by the Crowdflower crowdsourcing platform. The collected tweets were firstly annotated to identify whether they contained a request for help or an offer of assistance and furthermore categorized based on the type of resource involved into six different categories, including shelter, food, clothing, money, medical, and volunteer. Table 1 shows exemples of

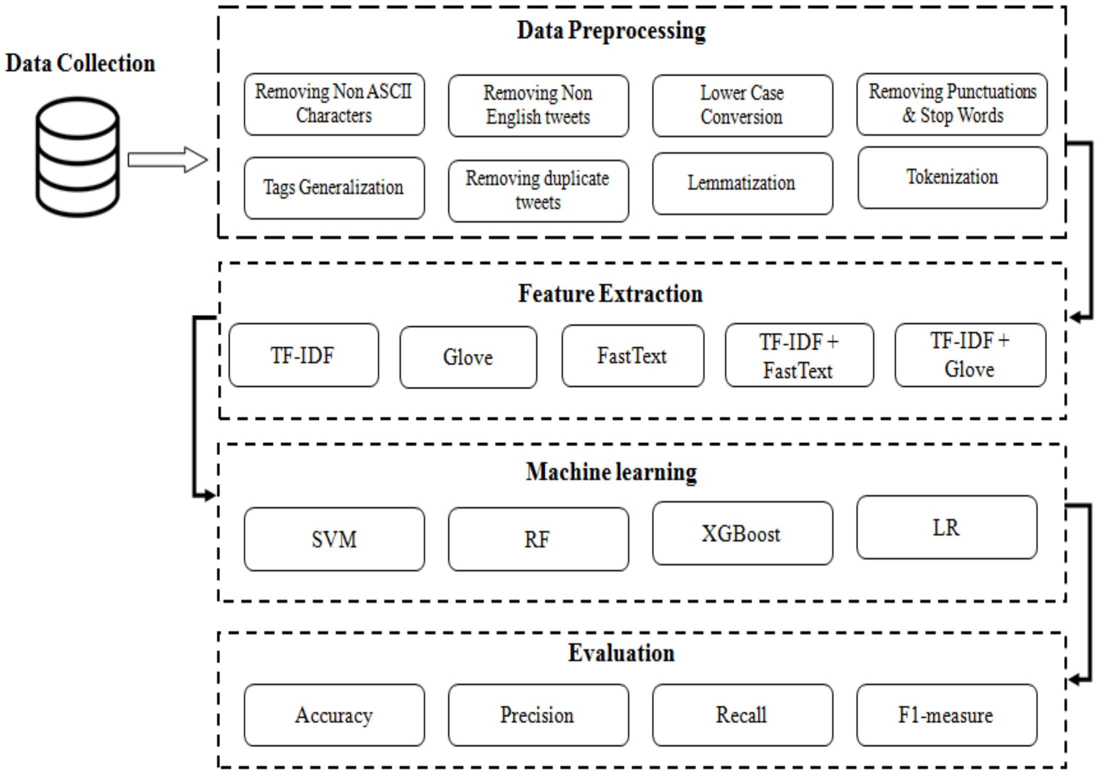

Figure 1: The proposed framework for need assistance prediction.

tweets belonging to each category. We randomly divided the dataset into training and testing sets, with 80% (1158) of the data allocated for training and the remaining 20% (290) for testing.

4.2 Performance metrics

To evaluate the performance of the proposed models, various metrics have been utilized, such as accuracy, precision, recall, and F1-mesure, which are calculated respectively as follows:

- Accuracy is the percentage of correctly classified tweets.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

- Precision is the percentage of positively classified tweets that are actually correct.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

- Recall score indicates the ability of the classifiers to classify all positive instances correctly.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

Category	Tweet
Money	"I'm raising money for Operation Hurricane Rescue. Click to Donate: http://t.co/rpscPr94 #gofundme"
Clothing	"If you have extra clothes and wish to donate to the victims of Hurricane Sandy, @Apt78, @ Apt http://t.co/RRFcDRrO"
Blood	"Sign up to donate blood to help victims of Hurricane Sandy here: http://t.co/pqpn0oNX"
Food	"Help! Hurricane victims need grub. You can buy food on http://t.co/56m51cS7 and have it delivered to Jacoby Church, 5406 4th ave., Brooklyn"
Medical	"Donate to Hurricane Sandy Victims tomorrow @ church across from Ocean Medical Center #restoretheshore #sandy #sandyhelp"
Volunteer	How to volunteer for and/or donate to the Hurricane Sandy relief effort: http://t.co/kuYlo5s8

Table 1: Examples of tweets from each category

- F1-score indicates the weighted harmonic mean of both precision and recall.

$$F1\!-\!score = 2\times\frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

All the metrics used in the evaluation are described in terms of TP, TN, FP and FN. In our experiments, we consider true positives (TP) and true negatives (TN) as the number of correct prediction of positive class and the number of correct positive predictions of negative class, respectively. Conversely, false positives (FP) represent the number of incorrect positive predictions of a class, while false negatives (FN) represent the number of the incorrect negative predictions number of a class.

4.3 Experimental Setting

We conducted our experiments using Python on a Google Colab environment with GPUs. The scikit-learn package in Python was used to implement the different predictive models. Our study employed four machine learning classifiers, namely: Logistic Regression (LR), Random Forest (RF), Support Vector Machines (SVM) and XGBoost classifier. These models have hyperparameters that can be adjusted to improve their prediction performance. To optimize these models, we performed a grid search using 10-fold cross-validation to determine the best combination of parameters with respect to the F1 metric. This study also investigated various word representation methods, such as Term Frequency Inverse Document Frequency (TF-IDF), Glove, and FastText. These methods are used to represent words in a numerical format that can be further used as input to all machine learning models. For all experiments in this research, the training and testing percentages were set to 80% and 20% respectively.

4.4 Results

We carried out a comparative study on the use of different feature extraction methods to predict support needs in disaster situation. The results obtained by all the machine learning models are presented in table 2. This table shows the accuracy, precision, recall and F1 measurement value of each model using five different feature extraction methods. Following these results, we can conclude that:

- The best results were obtained with the support vector machines (SVM) in terms of accuracy and F1-measure score. This classifier performed much better than other supervised models using various feature representation techniques. For example, a score of 85.71% was achieved using TF-IDF, 84.12% using Glove, 85.71% using TF-IDF and Glove combined, and 86.11% using TF-IDF and FastText combined.

- By integrating word embedding techniques such as Glove and FastText into the classifier, we observed a substantial enhancement in its performance.

- Combining both word embeddings (TF-IDF + Glove) and (TF-IDF + FastText) yielded slightly better performance than using single-word representation techniques.

- Hybrid TF-IDF based FastText outperforms hybrid TF-IDF based Grove for most machine learning algorithms. This is because FastText is good at grasping the meaning of words in a sentence, and combined with TF-IDF, it can represent words better.

- Combining syntactic and semantic information improves the performance of machine learning models.

5 Conclusion and future work

In this work, we conducted an extensive experiment using different feature extraction techniques to identify the types of assistance needs posted on social media platforms during disasters.

Our results show that using hybrid feature extraction techniques can improve the performance of machine learning models. And, the best result were obtained by combining TF-IDF and FastText word embedding. Among the different models evaluated, the SVM model gives the best result in terms of accuracy and F1-mesure.

As a future work, we plan to explore the use of different deep learning models for help assistance detection. Additionally, we will investigate the potential benefits of combining other feature extraction techniques such as Bag of Words, BERT, RoBERTa, DistilBERT, XLNet, and Roberta among others.

	Metrics	RF	LR	SVM	XGBoost
TF-IDF	Accuracy	82.53%	75.79%	85.71%	80.55%
	Precision	85.58%	89.05%	83.05%	85.39%
	Recall	82.53%	75.79%	83.05%	80.50%
	F1-score	83.63%	79.86%	81.85%	82.19%
Glove	Accuracy	84.12%	77.38%	84.12%	83.73%
	Precision	81.59%	85.28%	81.77%	82.25%
	Recall	84.12%	77.38%	84.12%	83.73%
	F1-score	81.05%	80.26%	82.64%	82.20%
FastText	Accuracy	82.53%	76.98%	82.55%	80.55%
	Precision	79.02%	74.39%	79.40%	77.75%
	Recall	82.53%	72.98%	80.32%	80.55%
	F1-score	79.14%	78.34%	79.82%	78.89%
TF-IDF + Glove	Accuracy	85.71%	79.76%	85.71%	84.12%
	Precision	85.52%	81.58%	83.53%	83.30%
	Recall	85.71%	79.76%	85.71%	84.12%
	F1-score	82.30%	80.45%	83.61%	82.53%
TF-IDF + FastText	Accuracy	86.11%	82.53%	86.11%	84.52%
	Precision	86.32%	84.04%	84.38%	82.48%
	Recall	86.11%	82.53%	86.11%	84.52%
	F1-score	82.97%	83.14%	84.38%	81.68%

Table 2: Model evaluation results with different features extraction methods

References

Ashwin Devaraj, Dhiraj Murthy, and Aman Dontula. 2020. Machine-learning methods for identifying social media-based requests for urgent help during hurricanes. *International Journal of Disaster Risk Reduction*, 51:101757.

Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2016. Bag of tricks for efficient text classification. *arXiv preprint arXiv:1607.01759*.

Md Yasin Kabir and Sanjay Madria. 2019. A deep learning approach for tweet classification and rescue scheduling for effective disaster management. In *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 269–278.

Deden Ade Nurdeni, Indra Budi, and Ariana Yunita. 2022. Extracting information from twitter data to identify types of assistance for victims of natural disasters: An indonesian case study. *J. Manag. Inf. & Decis. Sci*, 25.

Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

Hemant Purohit, Carlos Castillo, Fernando Diaz, Amit Sheth, and Patrick Meier. 2014. Emergency-relief coordination on social media: Automatically matching resource requests and offers. *First Monday*.

Irfan Ullah, Sharifullah Khan, Muhammad Imran, and Young-Koo Lee. 2021. Rweetminer: Automatic identification and categorization of help requests on twitter during disasters. *Expert Systems with Applications*, 176:114787.

Carlos Villegas, Matthew Martinez, and Matthew Krause. 2018. Lessons from harvey: Crisis informatics for urban resilience.

Zheye Wang, Nina SN Lam, Mingxuan Sun, Xiao Huang, Jin Shang, Lei Zou, Yue Wu, and Volodymyr V Mihunov. 2022. A machine learning approach for detecting rescue requests from social media. *ISPRS International Journal of Geo-Information*, 11(11):570.