Multimodal Combination of Text and Image Tweet for Disaster Response Assessment

Saideshwar Kotha, Smitha Haridasan, Ajita Rattani, Aaron Bowen, Glyn Rimmington, Atri Dutta,
Disaster Resilience Analysis Center
Wichita State University, USA.
Outline

Introduction
Problem Statement
Proposed Method
Dataset
Experimental Results
Conclusions and future work
Introduction

Information related to natural disaster is posted on social media.

Sharing such information is useful for humanitarian support.
Example keywords in tweets useful for Informative category

• Caution
• Help needed
• Warnings
• Rescue
• Donation request

Kathy @Katnow123 · Jul 1
Another earthquake in Dubai. Anyone feeling this?! It’s still shaking!

Show this thread

joshdoit8 @joshdoit8 · Jul 1
Another Saturday, another #earthquake in Dubai. Only this time, it was longer... almost a minute! Hopefully this quake has not caused any damage or casualty in the UAE and nearby countries. 🙏
Anyways, I drove through the Redwood Forest then next to the California wildfires
https://t.co/yiK5BQh7PF

Rampant sharing of crisis related posts has led to the need for categorizing data into Informative and Non-Informative categories
Humanitarian Aid

- Humanitarian aid workers help people during crisis events to save lives, reduce suffering and rebuild affected communities.
Prior Work

• Ofli et al. evaluated VGG16 on tweet texts and tweet images from CrisisMMD dataset for two categories
  • Informative vs Non-Informative
  • Humanitarian categories

<table>
<thead>
<tr>
<th>Modality</th>
<th>Informative vs Non-Informative categories (%)</th>
<th>Humanitarian categories (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text only</td>
<td>80.80</td>
<td>70.40</td>
</tr>
<tr>
<td>Image only</td>
<td>83.30</td>
<td>76.80</td>
</tr>
<tr>
<td>Text + Image</td>
<td>84.40</td>
<td>78.40</td>
</tr>
</tbody>
</table>

Problem Statement

Aim of this research is to process multimedia data and propose a **fusion-based decision-making technique** to classify social media data into

- Informative and
- Non-Informative category

**Informative** tweets are then classified into various humanitarian categories such as

- Rescue volunteering or donation efforts,
- Infrastructure and utility damage,
- Affected individuals,
- Other relevant information
- Not-Humanitarian.
Dataset - CrisisMMD

Multimodal Crisis (CrisisMMD) – consists of tweet texts and images collected during multiple major natural calamities including earthquakes, floods, hurricanes and wildfires.

The abundance of social media data clearly indicates the possibility of image processing research mainly by assisting humanitarian aid workers.
Proposed Method
Experiment #1: Text Modality for Disaster Response Assessment

Preprocessing

Text Tweets

DistilBert

Dense Layers

Informative Tweets
Non-Informative Tweets

Rescue volunteering or donation efforts
Infrastructure and utility damage
Affected individuals
Other relevant information
## Evaluation of DistilBERT on Text Modality

<table>
<thead>
<tr>
<th>Text Modality</th>
<th>Informative vs Non-Informative (%)</th>
<th>Humanitarian categories (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Accuracy</td>
<td>82.65</td>
<td>75.39</td>
</tr>
</tbody>
</table>
Experiment #2: Image Modality for Disaster Response Assessment

CNN models used:
- VGG16
- VGG19
- ResNet50
- DenseNet121
- RegNetY320
# Evaluation of CNN Models on Image Modality

<table>
<thead>
<tr>
<th>Image Modality</th>
<th>Informative vs Non-Informative (%)</th>
<th>Humanitarian categories (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VGG16</td>
<td>83.11</td>
<td>73.5</td>
</tr>
<tr>
<td>VGG19</td>
<td>83.18</td>
<td>76.23</td>
</tr>
<tr>
<td>ResNet50</td>
<td>83.11</td>
<td>77.17</td>
</tr>
<tr>
<td>DenseNet121</td>
<td>83.63</td>
<td>76.12</td>
</tr>
<tr>
<td>RegNetY320</td>
<td>85.13</td>
<td>80.20</td>
</tr>
</tbody>
</table>
Experiment #3: Fusion of Text and Image Modality for Disaster Response Assessment

- DistilBert: Text Tweets
- CNN Models: Image Tweets
- Concatenated Features
- Dense Layers
  - Informative Tweets
  - Non-Informative Tweets
  - Rescue volunteering or donation efforts
  - Infrastructure and utility damage
  - Affected individuals
  - Other relevant information
## Evaluation of DistilBERT on Text + Image Modality

<table>
<thead>
<tr>
<th>Image Modality</th>
<th>Informative vs Non-Informative (%)</th>
<th>Humanitarian categories (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bert + VGG16</td>
<td>86.50</td>
<td>82.4</td>
</tr>
<tr>
<td>Bert + VGG19</td>
<td>86.83</td>
<td>81.67</td>
</tr>
<tr>
<td>Bert + ResNet50</td>
<td>88.00</td>
<td>84.71</td>
</tr>
<tr>
<td>Bert + DenseNet121</td>
<td>89.50</td>
<td>85.96</td>
</tr>
<tr>
<td>Bert + RegNetY320</td>
<td>89.63</td>
<td>86.59</td>
</tr>
<tr>
<td>Modality</td>
<td>Informative vs Non-Informative categories (%)</td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Text only</td>
<td>82.65</td>
<td></td>
</tr>
<tr>
<td>Image only</td>
<td>85.13</td>
<td></td>
</tr>
<tr>
<td>Text + Image</td>
<td>89.63 (5% increase)</td>
<td></td>
</tr>
</tbody>
</table>
## Humanitarian Result Analysis

<table>
<thead>
<tr>
<th>Modality</th>
<th>Humanitarian categories (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text only</td>
<td>75.39</td>
</tr>
<tr>
<td>Image only</td>
<td>80.20</td>
</tr>
<tr>
<td>Text + Image</td>
<td><strong>86.59 (8.19% increase)</strong></td>
</tr>
</tbody>
</table>
Conclusions

Experimental results suggest that fusion of text and image tweets using Multimodal deep learning model on CrisisMMD dataset performs better than either the single text or image modality.
Future Work

• As part of future work, we will explore advanced fusion-techniques for combining text and image modalities and advanced deep learning models to improve the classification performance.
Thank you
Questions