EMS Annual Meeting Abstracts
Vol. 10, EMS2013-802-1, 2013
13th EMS / 11th ECAM


# Social media for meteorological prediction and information

K Stanisławska
Poland

The meteorological data nowadays is obtained thanks to millions of sensors located worldwide and constantly measuring temperature, pressure, wind speed and other parameters. This data is collected by entitled institutions and processed by physical models, giving the final product of a forecast. Developing such a complex network takes years and in a need of deployment of a new one, a lot of time will pass before it gets fully operational. One of the ideas that arose recently with the society getting more mobile and equipped with smart devices is the concept of treating humans as sensors and their interaction with those devices as sensory information. At the same time, development of various machine learning techniques led to better possibilities of deriving models from more free-form data. Therefore, what would happen if people were treated as weather conditions receptors and their online activity as a source for meteorological data?

Every minute the people around the world produce gigabytes of social media content. It appears as text, pictures or videos shared through such services as Facebook, Twitter, Flickr, YouTube or Instagram. A great amount of this content is publicly available, i.e. can be downloaded without privacy restrictions.

This big volume of people-produced data has already led to various studies on their usage and application. Such domains as political sciences, urban studies or tourism have taken advantage from the research based on social media sensors. Several years ago the Japanese have successfully built an earthquake prediction system based on the information from users' posts on Twitter. Now it is time for meteorology to try.

The subject of this project is to construct a meteorological system based on the information extracted exclusively from the social media content. The system would be a service providing information about current weather conditions, trends (understood as a certain phenomenon being seen in consequent locations over time) and a forecast. The steps leading towards the final product would include:

- Fetching a real-time content stream from social media services (especially Twitter), customizable by the available Application Programming Interfaces,
- extracting the relevant data from the stream using keyword parsing and text classifiers,
- building a knowledge base from the extracted data, taking into account possible fuzzy relationships between them,
- applying a machine learning technique (possibly genetic programming) to detect cause-effect relationships and using obtained rules to project the current state to the future to get the forecast,
- preparing an automatically updated map of weather conditions, to be presented to end users.

The creation of the system would be at the same time a broad research on the potential of social media for natural phenomena anticipation. The state-of-the art methods for data classifying and forecast building could be used in further applications. And as the social media services get more and more users every day and more of its content becomes available for input, there would be a potential for a continuous, unconscious enhancement.