Extract opinions to predict the future!
Semester or Master project

Keywords: *big data, machine learning, opinion mining, natural language processing*
Supervisors: Florent Garcin, {*firstname.lastname*}*@epfl.ch*

Description

We have developed a prediction platform called swissnoise (http://go.swissnoise.ch), which uses the wisdom of the crowds to predict outcomes of future events. One prediction technique is based on a prediction market. Users trade shares of the outcomes for an event with virtual money. The current market price of an outcome is interpreted as an estimate of the probability of the outcome.

For instance, the event "Will it snow during Christmas Eve?" has outcomes "yes" and "no". If a user is confident that it will snow, he will buy shares of "yes". If he changes his mind because he saw a new weather report, he can sell his shares of "yes", and make a potential profit. The user gets $10 per share if he has the correct outcome. The current market price of one share of a given outcome is interpreted as the predicted probability of that outcome.

The goal of this project is to implement a polarized news aggregator, and study how news articles affect users' decision making and the overall prediction.

The first step will be to get news content from different sources such as BBC and techcrunch. Then, we will extract relevant information (keywords and topics) in order to cluster the news and match them to the events we want to predict on swissnoise. With this, we will have a set of relevant news articles per event we are trying to predict.

The next step is about mining the opinion of each news articles in order to tell whether the article is positive, negative or neutral about the current event. For instance, a news article talking about an upcoming snow storm would indicate a positive polarity towards the event we are trying to predict, i.e. "Will it snow during Christmas Eve?".

After that, we will display this filtered and polarized information to swissnoise's users and see how this information impact their decision making and predictions.

If time allows, we will also implement a trading agent which leverages this polarized information as a trading strategy.

Tasks

1. implement a news aggregator from different news sources.
2. extract relevant information from these news articles, such as topics and important keywords.
3. tag the news articles (or part of the article) with a polarity (positive, negative, neutral).
4. integrate your aggregator into our prediction platform.
5. optional: implement a trading agent which takes advantage of this information.