Financial Event Detection Using Semantics

Frederik Hogenboom[1], Alexander Hogenboom[1], Flavius Frasincar[1], Franciska de Jong[2], and Uzay Kaymak[3]

{fhogenboom,hogenboom,frasincar}@ese.eur.nl, f.m.g.dejong@utwente.nl, u.kaymak@tue.nl

[1] Erasmus University Rotterdam	[2] University of Twente	[3] Eindhoven University of Technology
P.O. Box 1738, NL-3000 DR	P.O. Box 217, NL-7500 AE	P.O. Box 513, NL-5600 MB
Rotterdam, The Netherlands	Enschede, The Netherlands	Eindhoven, The Netherlands

Abstract

Nowadays, emerging news on economic events like acquisitions substantially impacts financial markets. Hence, it is important to automatically and accurately identify economic events in news in a timely manner. For this, one has to be able to process a large amount of heterogeneous sources of unstructured data to extract knowledge useful for guiding decision making processes. We propose a Semantics-based Pipeline for Economic Event Detection (SPEED), with which we aim to extract financial events from emerging news gathered from RSS feeds, and to annotate these with machine-understandable meta-data, while retaining a speed that is high enough to make real-time use possible.

Our framework is modeled as a pipeline which takes news messages as input and is driven by a financial ontology developed by domain experts, containing information extracted from Yahoo! Finance on NASDAQ-100 companies. In our implementation, we have reused some of the components of the existing General Architecture for Text Engineering (GATE) framework and additionally, developed new ones, e.g., an Ontology Gazetteer and a Word Sense Disambiguator.

Experiments on 200 news messages fetched from the Yahoo! Business and Technology RSS feeds show fast, sub-second gazetteering and a precision and recall for concept identification of 86% and 81%, respectively. Precision and recall of fully decorated events result in lower values of approximately 62% and 53%, as they rely on multiple concepts that have to be identified correctly. Our Word Sense Disambiguator with an adapted SSI algorithm shows a precision and recall of 59%, compared to 53% and 31% for the original algorithm.