A Prospect on How to Find the Polarity of a Financial News by Keeping an Objective Standpoint
Position Paper

Roxana Bersan, Dimitrios Kampas and Christoph Schommer
Dept. of Computer Science and Communication, University of Luxembourg, Luxembourg City, Luxembourg

Keywords: Artificial Agents, Financial News, Document Polarity, Wisdom of Crowd, Self-organisation, Social Systems.

Abstract: This position paper raises the question on how we can keep an independent standpoint regarding the finding of a polarity in a news document. As we know, an usefulness and relevance of a text news may be seen differently by a group of evaluators. The differences are depending on their interests, their knowledge, and/or their ability to understand. Recent research in literature mostly follow a top-down approach, which is either a context-based solution or a dictionary-based approach. With respect to the perspective (standpoint) of an evaluator, we therefore come up with an alternative approach, which is bottom-up and which tends to overcome the power of a single evaluator. The idea is to introduce a collection of theme-related artificial agents (financial, economic, or political, ...), which are able to vote. A decision regarding the polarity of a financial news bases on the interplay of a social collection of agents (a swarm), which serve and assist the artificial agents while fulfilling simple (linguistic, statistical) tasks.

1 MOTIVATION

The European Financial Crisis has emerged within the last years, with many ups and downs, with many consequences and decisions for politics and economy. For example, Eurobonds have been suggested, attracting a great deal of attention while financial news appeared in a Tsunami-style of eruptively flowing pace. Besides, financial and political activities have taken place, political communities have emerged, and coalitions established. Also, a certain number of states have been down-rated, Greece (and potentially other European states) are close to insolvency. All these information has been well-noted in financial news.

We concern ourselves with such *financial text news* (Thomson Reuters), which represent a reflection of momentary political, economical, and financial incidents. *Financial text news* can influence decisions, expose realistically and unaltered current events, and/or contribute to the formation of an opinion. Our concern therefore is: assuming that we have financial texts with an exclusive concentration on facts and objectivity, can we then find indications regarding financial, political, or economic decisions, for example with respect to the European crisis? Can we identify a relationship to the financial market (stock market and others)? Can we observe a composition of similar thinking people? Can we be proactive and illustrate the emergence of the crisis as well as a future recurrence?

In Computer Science research, several directions regarding the analysis of texts have evolved. One of them is *sentiment analysis* of texts and with it the finding of an inherent polarity of the document. A sentiment classification refers to identifying and extracting subjective information that appears in source materials and to determining the attitude of a person concerning an overall contextual polarity of a document. The sentiment may be a judgement or an evaluation, an affective state, or the intended emotional communication. Following this, a finding of answers to the questions above might be rather simple in a way that a certain number of existing techniques may be applied. More easily, we could argue that we only have to analyse the documents linguistically, statistically, and from a Machine Learning point of view, and that we then may come up with a sentiment decision. However, this is not as easy as it seems.

A crucial argument is that we must guarantee a neutral perspective (or standpoint), with almost no apriori expectations. The reason is that a financial news may be interpreted in a different way, depending on what a evaluator thinks, believes, and/or knows. As an example, let us consider the following financial

Bersan R., Kampas D. and Schommer C..
A Prospect on How to Find the Polarity of a Financial News by Keeping an Objective Standpoint - Position Paper.
DOI: 10.5220/0004191601720177
In *Proceedings of the 5th International Conference on Agents and Artificial Intelligence* (ICAART-2013), pages 172-177
ISBN: 978-989-8565-38-9

text news:

> "Juncker suggested to introduce Eurobonds. This is a good sign for the new Europe."

The interpretation of the above sentences is ambiguous and may cause - because of the evaluator's position - different conclusions as well as a misunderstanding. If the apriori perspective of the evaluator is somehow positive with respect to Eurobonds, then the content is very appreciated and the document classified as to be positive. If it is not, then the document becomes a negative polarity and with it, possibly a negative signal for financial markets. Therefore, the argument of having a neutral perspective is needed with respect to a classification of the polarity.

In the following, we will concern the polarity/sentiment of a document and present a collection of research works that has been made recently. We will target the problem of having independent perspectives (standpoints) and claim that a fair, stable, and reliable decision can only be made by a voting of emancipated artificial agents.

2 SELECTED RESEARCH

2.1 Feature Spaces

A first idea on how to discover a polarity of a news texts are the geometric models. Given a set of (predefined) features F=$\{f_1, \ldots, f_n\}$, where for example the features represent financial terms, locations, interests, et cetera, then each financial news can be represented itself as a vector in the space. We take the frequency of a feature f_i in a financial news document as a coordinate of the normalized vector F.

Regarding the polarity of financial news documents, we may start with a set of training documents, whose polarity is already known (supervised approach). Their position in the space then gives a first hint on whether new documents, which are close enough, are becoming even more positive-polar or negative-polar. However, the assignment of a new financial news to the polarity feature space is problematic. The features (dimensions) may be too weak or less appropriate or their relevance has changed over time. There is also a big uncertainty regarding financial news with regards to their contents: two such documents can be similar, but the presence of a negation or an antonym may force a different polarity. Moreover, the perspective (standpoint) of an individual is not sufficiently respected, since the fixing of the dimension and/or the supervised polarity assignment of the training documents are subjective.

The reason why we mention this is caused (among others) by a work of (C. Scheible, 2012), who present a novel graph-theoretic method for the initial annotation of high-confidence training data for bootstrapping sentiment classification. Here, the polarity is estimated here by a theme-specific 'PageRank' algorithm. The authors argue that basically sentiment information is propagated from an initial seed lexicon through a joint graph representation of words and documents. They show that their approach outperforms a baseline classifier and that its performance can be further improved by a bootstrapping method that can take advantage of the entire feature space available.

2.2 Polarity in Text Documents

In literature, a conscious discussion on perspectives (standpoints) is rarely made. Almost any research work concerns a concrete application or a technical how-to-do, accomplished with arguments describing its need. Some applications use a dictionary-based solution, others a context-based solution. As some examples, (Hassan and Radev, 2010) propose a method to automatically identify the polarity of words by taking advantage of a Markov random walk model to a large word relatedness graph and producing a polarity estimate for any given word. The authors say that a key advantage of their model is its ability to accurately and quickly assign a polarity sign and magnitude to any word. (Richter et al., 2010) describe a new method for extracting negative polarity item candidates (called NPI candidates) from dependency-parsed German text corpora focusing on target multiword expressions. (Schumaker et al., 2012) raise the question whether the choice of words and tone used by the authors of financial news articles can correlate to measurable stock price movements. If yes, so the authors, can then the magnitude of price movement be predicted using these same variables? The authors answer these questions by using the Arizona financial Text (AZfinText) system, a financial news article prediction system, and pair it with a sentiment analysis tool.

(Devitt and Ahmad, 2007) aim to explore a computable metric of positive or negative polarity in financial news text, which is consistent with human judgments. The authors say that this can be used in a quantitative analysis of news sentiment impact on financial markets. (Sakai and Masuyama, 2009) propose a method of assigning polarity to causal information extracted from Japanese financial articles concerning business performance of companies. The authors assign a polarity (positive or negative) to causal information in accordance with business per-

formance. (Drury et al., 2011) propose a strategy to segment quotations inside a text by an inferred "opinion maker" role and then apply individual polarity classification strategies to each group of the segmented quotations. They have modelled a contextual information with Random Forests based on a vector of unigrams. (Heerschop et al., 2011) propose a system called *Pathos*, which is a framework to perform document sentiment analysis. *Pathos* is partially based on a discourse structure of the document. The authors hypothesize that - by splitting a text into important and less important text spans and by subsequently making use of this information by weighting the sentiment conveyed by distinct text spans in accordance with their importance - they improve the performance of a sentiment classifier. A document's discourse structure is obtained by applying Rhetorical Structure Theory on sentence level. (Kaji and Kitsuregawa, 2006) propose a novel fully-automated method of building polarity-tagged corpus from HTML documents to utilize certain layout structures and linguistic pattern. In general, Polarity Dictionaries are not less prone to being subject to subjectiveness. (Paulo-Santos et al., 2011), for example, argue that most approaches in finding polarity dictionaries rely on linguistic works concerning part-of-speech tagging or rich lexical resources such as WordNet. The authors show and examine the viability to create a polarity lexicon using only a common online dictionary with five positive and five negative words, a set of highly accurate extraction rules, and a simple yet effective polarity propagation algorithm. The algorithm evaluation results show an accuracy of 84.86% for a lexicon of 3034 words.

3 A PROSPECTIVE APPROACH

The term rational agent is described by (Russell and Norvig, 2010) as an entity that given the built-in knowledge representation, the actual state of the environment and the set of possible actions he can take, he selects an action so as to maximize its utility measure. The most important agent categories

- *Model-based Reflex Agents*. These take a decision about its next action based on a set of conditional action rules about the effect of the actions on the environment state
- *Goal-based Agents*. These types of agents are dedicated to find action sequences for achieving a higher goal
- *Utility-based Agents*. These agents worke more on representation, modelling and learning.

(Michael Brenner, 2009) consider the problematic of intermingling planing with acting in dynamic and partially observable multi-agent environments. In (Trevor Bench-Capon, 2012), the authors utilize artificial agents in order to foster on a decision regarding economic experiments with games implementations. In (Jérôme Lang, 2012), binary tree rules are used for deciding the winner as elected by a majority vote. (Davide Grossi, 2012) look into the possibility of incorporating a form of dependence relation in the field of game theory for an agent interaction. Cooperative games are analyzed, where coalitions undertake agreements based on dependency relations. (AndreasWitzel, 2012) analyze the epistemic reasoning of communicating rational agents with focus on the distributed form of iterative elimination of strictly dominated strategies.

3.1 Demands

Regarding the polarity finding, we demand for adaptive artificial agents (Russell and Norvig, 2010), which are able to represent and acquire knowledge, to learn from internal and external information ((Clark, 2001)), and to take advandage of the Wisdom of Crowds ((Surowiecki, 2004)). An artificial agent is able to classify documents based on its internal *knowledge base*. With this, it owns an *aptitude*, which is a competency regarding a field of application, for example *finance*, *politics*, *economics*, and others. An artificial agent can either represent a natural person (Merkel, Sarkozy, ...), a country (Greece, Germany, ...), or another individual. Each agent owns a standpoint and is able to vote, classifying a financial news individually to a polarity, which is either positive, negative, or neutral. Using a set-based operation like the *intersection* can be applied to prove a stable, reliable, and plausible polarity decision, where a financial news is then positive (negative, neutral), if the majority of the agents vote for positive (negative, neutral), respectively. We assume that the fundamental characteristics of an agent are the following:

- Contradicting *opinions*: since there exist much more than one perspective (there are often many truths), an agent must have have an adaptive knowledge base.
- Able to take a *decision*: an agent must be able to operate on text, to find associative structures, co-occurrences, or other forms of patterns.
- *Independence* of voting: An agent must state his/her polarity decision independently and has only then the right to vote.
- Presence of a *theme*: An agent can only decide a

document polarity, if an evaluation theme exists. Otherwise, a voting may become directionless.

Then,

$$polarity(d, r_n) = \begin{cases} +1, & \Delta(r_{i+}, r_{j-}) > \varepsilon \\ -1, & \Delta(r_{i+}, r_{j-}) < -\varepsilon \\ 0, & -\varepsilon \leq \Delta(r_{i+}, r_{j-}) \leq \varepsilon \end{cases} \quad (1)$$

with $|d| = n = i + j$ and where r_{i+} (r_{j-}) refer to a positive (negative) vote regarding the document. ε is a threshold, which only classifies a financial news to -1 or +1 if Δ (x,y) is the absolute difference of (x,y). Only in case that the voting is equal, the document is seen to be neutral. This follows the concept of (Surowiecki, 2004), who argues that decisions are taken by a large group, even if the individuals within the group are not smart; but these decisions are always better than decisions made by small numbers of 'experts'.

3.2 Artificial Agents

An alternative idea is to understand the polarity as a decision, which is taken by a majority (or a weighted sum of) of *slave agents*, which serve a given *master agent*. These *slave agents* share a small capacity, are assigned a simple task, and collaborate as a part of a social system or swarm. One consequence of such an architectural framework is a small amount of apriori knowledge, because the participating entities have to do a little task requiring less of it. Also, a plausibility of the polarity decision will be inherently given. All entities' decision, being either 'positive', 'negative', or 'neutral' can be identified and arguments for the final decision retrieved. As the decision is made by many collaborating entities, we assume the decision to be more fault-tolerant, more resistant against temporal changes, and less vulnerable to a wrong document classification. A single change of the knowledge landscape (for example 'Sarkozy' is no longer president but 'Hollande' is now) will not have such a big effect. Moreover, the social system might work autonomously and organises itself, reducing the number of investigated efforts. And finally, an independent perspective is maintained. With that, we may understand an agent as an artificial entity, which knows its user (reviewer of the financial news), but which is served by even small entities, i.e., slave agents.

Assume that we have a certain number of European key players, country names, locations, and other *facts*. In an intelligent environment, the social system could detect such *facts* by itself and neglect such *facts* in case of inactivity over a certain period of time, but we keep it more simple here and assume that a certain number of *fact slave agents* (for example, focusing on the key players in Europe, countries, locations, etc.), whose task is to serve the artificial agent and to check a document for occurrences and frequencies of assigned terms. If the frequency is sufficient, possibly above a given threshold, then each *fact slave agent* contributes to the polarity decision.

We also may consider k-ary operations like *agrees(X)*, *brings(X,Y)*, *has(X,Y)*, or *gives(X,Y,Z)*, which are addressed by *action slave agents* aiming at instantiating the arguments or even word polarities like *war (-)*, *Eurobond (+/-)*, or *good (+)*, where we assign an individual *polarity agent*, whose task is to control the presence of predefined words. Of course, many other types of agents may be used, for example a *negation slave agent*, whose task could be to convert an *action slave agent*'s decision; or an *uncertainty agent*, whose task is to reduce a certainty of the agent's decision, for example by a multiplicative compensation. *Warehouse slave agents* may have the task to put all these information together, bringing the whole information landscape to a consistent and reasonable decision. Finally, *statistical slave agents* and *linguistic slave agents* may be taken as well, for example to deliver statistical and linguistic numbers/values.

But which role do the *warehouse slave agents* play? Do they just compute weights and relation between action slave agents and polarity slave agents or should they perform more than that? Moreover, which are the role of the linguistic agents? Is a linguistic analysis not already incorporated by the *action slave agents*? Which is the role of the *statistical slave agents*, especially, are they not yet incorporated by the *polarity slave agents*? To give a more precise answer, we suggest the following:

- A *Fact slave agent* can be either a *subject slave agent* or an *object slave agent*. Each of these agents can have sub-hierarchies of their own, for example a *subject slave agent* may have subcategories like 'politician', 'company', et cetera, an *object slave agent* categories 'location', 'event', et cetera. As an example, 'Merkel is the chancellor' is a subcategory of 'politician', and with that, a subcategory of the *subject agent*. 'Madrid is a city' is a subcategory of a country, and with it, a subcategory of a 'location'. 'Summit G-20' is subcategory of 'event', which is a subcategory of an *object agent*.
- *Action slave agents* can be, as mentioned above, k-ary. But we think that using *action slave agents* with ≥ 3 parameters burdens the relation extraction too intensively. Assuming to determine one term would be as good as 90%, then we probably get an accuracy of almost 72% for three terms.

- *Polarity slave agents* can be applied for the subcategories 'verb', 'noun', and 'adjective'. Some verbs and adjective may have a given standard polarity (Example 'good' polarity is positive, 'hates' polarity is negative), whereas a polarity of nouns may differ over time (Example: 'war' is constantly, but the polarity of 'Eurobond' is probably not).
- *Linguistic slave agents* would perform some linguistic analysis deciding on the polarity of the sentence based on some predefined policy. E.g., when the verb has a negative or positive polarity the sentence takes the polarity of the verb. If the verb has an objective polarity, then the polarity of the sentence is the polarity of the nouns, the adjectives, or the adverbs (Example: "Merkel supports Eurobonds."). Thus, the polarity of the sentence is taken by the polarity of Eurobonds. Linguistic agents capture negations as well reversing the polarity of a sentence as well. E.g."Merkel does not agree on Eurobonds".
- *Uncertainty slave agents* are responsible for decreasing the polarity volume of the sentence by capturing the uncertainty word.
- *Warehouse slave agents* are basically responsible for the decision, because they integrate the information coming from the other agents. However, they are not allowed to vote.

3.3 How to Find a Decision?

An agent may be composed of many slave agents, which perform a simple task. The working together of these agents will then become the fundament with respect to the polarity. In the simplest case, a voting of all agents with equal rights can be taken into account. But before the decision on the polarity of a document is taken, it must be considered whether alternative types of voting can be applied (or not), especially plurality voting systems, single-winner voting systems, or multiple-winner voting systems. For example, whether there exist the word 'Eurobond' or not can be subject to a plurality voting system. But, which countries are pro 'Eurobond' or against 'Eurobond' it is a multiple-winner one.

Regarding the voting decision, there are numerous paths in theory and application, which can be applied. These theories have their roots in the fields of Game Theory, Auction theory, and multi-agent systems. Examples for the field of Game theory are *Nash equilibrium* and the *revelation principle of economics*; examples for the field of *Auction Theory* are *English auction*, *Dutch auction*, *Vickrey auction*, and *sealed first-prive auction*. It is important to keep this in mind, because such a system takes into consideration an application, such as the increase of utility functions, the prediction of some economical phenomena, et cetera. Without this, a system would sound nothing more than a data collection system with no direct application.

4 CONCLUSIONS

The given idea is a visionary and prototypically try to overcome the problem of having a subjective perspective (standpoint) regarding the polarity finding of a document. As mentioned, the interpretation of a financial news may depend on a reviewer's knowledge, interest, and much more. Having designed artificial agents of different thematic directions, being assisted by many self-organising and self-evaluating types of agents, then this may overcome the given problem. Regarding the voting, which is only allowed for the agents, an independent voting is recommended. With that, we believe to fulfill the given demands of holding a perspective, an ability to take a decision, an independence of voting, and a presence of a theme. Our aim is now to follow this idea and to come up with a more detailed concept. An implementation is planned.

ACKNOWLEDGEMENTS

This work is supported by the Fonds National de la Recherche (FNR) Luxembourg within the research project *ESCAPE*, which is an interdisciplinary project between the Department of Computer Science & Communication and the Luxembourg School of Finance (LSF). The authors thank Mihail Minev as well as Prof. Grammatikos and Andreas Chouliaris for internal discussions.

REFERENCES

AndreasWitzel, Krzysztof R. Apt, J. A. Z. (2012). Distributed iterated elimination of strictly dominated strategies. *Autonomous Agent Multi-Agent Systems*.

C. Scheible, H. S. (2012). Bootstrapping sentiment labels for unannotated documents with polarity pagerank. In *LREC 2012*.

Clark, A. (2001). Reasons, robots, and the extended mind. In *Mind and Language*, volume 16:2, pages 121–145.

Davide Grossi, P. T. (2012). Dependence in games and dependence games. *Autonomous Agent Multi-Agent Systems*.

Devitt, A. and Ahmad, K. (2007). Sentiment polarity identification in financial news: A cohesion-based approach. In *ACL 2007*.

Drury, B., Dias, G., and Torgo, L. (2011). A contextual classification strategy for polarity analysis of direct quotations from financial news. In *RANLP 2011*, pages 434–440.

Hassan, A. and Radev, D. (2010). Identifying text polarity using random walks. In *ACL 2010*, pages 395–403.

Heerschop, B., Goossen, F., Hogenboom, A., Frasincar, F., Kaymak, U., and de Jong, F. (2011). Polarity analysis of texts using discourse structure. In *CIKM 2011*, pages 1061–1070.

Jérôme Lang, Maria Silvia Pini, F. R. D. S. K. B. V. T. W. (2012). Winner determination in voting trees with incomplete preferences and weighted votes. *Autonomous Agent Multi-Agent Systems Agent Multi-Agent Systems*.

Kaji, N. and Kitsuregawa, M. (2006). Automatic construction of polarity-tagged corpus from html documents. In *ACL 2006*.

Michael Brenner, B. N. (2009). Continual planning and acting in dynamic multiagent environments. *Autonomous Agent Multi-Agent Systems*.

Paulo-Santos, A., Ramos, C., and Marques, N. (2011). Determining the polarity of words through a common online dictionary. In *EPIA*, pages 649–663.

Richter, F., Fritzinger, F., and Weller, M. (2010). Who can see the forest for the trees? extracting multiword negative polarity items from dependency-parsed text. *JLCL 2010*, 25(1):83–110.

Russell, S. and Norvig, P. (2010). *Artificial Intelligence - A Modern Approach (3. internat. Ed.)*. Pearson Education.

Sakai, H. and Masuyama, S. (2009). Assigning polarity to causal information in financial articles on business performance of companies. *IEICE Transactions*, 92-D(12):2341–2350.

Schumaker, R., Zhang, Y., Huang, C., and Chen, H. (2012). Evaluating sentiment in financial news articles. *Decision Support Systems*, 53(3):458–464.

Surowiecki, J. (2004). *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. Little Brown & Co.

Trevor Bench-Capon, Katie Atkinson, P. M. (2012). Using argumentation to model agent decision making in economic experiments. *Autonomous Agent Multi-Agent Systems*.