A Lexicon for Polarity: Affective content in financial news text

Ann Devitt   Khurshid Ahmad

In recent years, the area of sentiment analysis in text has become a focus of attention in the fields of theoretical and computational linguistics, investigating the production and processing of affective contours in text, the textual corollary of emotional prosody in speech. Research has drawn on text from many domains ranging from on-line film reviews [17] to newspaper editorials [19] to Dow Jones News Service headlines [12] and has tended to focus on supposedly unequivocal, cross-domain markers of affect polarity in text, such as the terms "good" and "bad". Financial markets appear to be particularly prone to the polarity of opinions, some based on rumours and others based on exaggerations. The study of actual or fictitious polarity on financial markets has another dimension: not all the players in a financial market have equal information - market traders know about the behaviour of a number of investors but each individual investor is aware of the behaviour of only very few. This notion of asymmetry of information has been studied mathematically by Robert Engle and his colleagues [8, 2] and within a social context by Donald Mackenzie and others [10, 6]. Implicit amongst all the discussions is the question of how much texts written in the special language of finance impact on the well-being or detriment of the markets. A number of authors who have analysed the effect of the affective content of the news rely largely on their intuition on prescribing a lexicon that is used in expressing polarity [4, 5, 11] or existing affective word lists [16].

The language of finance exhibits many characteristics, such as dominant grammatical constructions evident in local grammars [1] or a propensity for certain metaphors [15] which strongly differentiate its users from non-initiates. This paper sets out to examine how affective content is expressed in the language of finance and what spectrum of emotion this language favours from two complimentary perspectives. On the one hand, we examine the language of emotion and the terms that contribute to how contours of emotion are realised in text. On the other, we examine the language of finance and what characterises text, in particular biased or emotionally charged text, in this domain.

In order to understand how emotion can be realised in text, we must first have a notion of what emotion is and how people experience it. Current cognitive theories of what constitutes emotion underpin this investigation of the language of emotion. There are two primary approaches to a cognitive account of emotion: categorical and dimensional. The Darwinian categorical approach posits a finite set of basic emotions which are experienced universally across cultures, (e.g. anger, fear, disgust, sadness, happiness and surprise [7]). The second approach delineates emotions according to multiple dimensions rather than into discrete categories. The two primary dimensions in this account are a good-bad axis (termed the dimension of valence, evaluation or pleasantness) and a strong-weak axis (termed the dimension of arousal, activation or intensity) [13, 14]. We investigate to what extent existing lexica of emotion reflect these underlying cognitive structures and the degree of consensus between these resources on a structural and lexical level. The analysis focuses on the Dictionary of Affect [18], SentiWordNet [9] and WN-Affect [3] and other resources used in the Sentiment Analysis literature to define the cognitive and empirical basis and the emotional spectrum or bias of what to date have been considered the lexica of emotion.

Based on this evaluation of the language of emotion, we carry out a corpus-based study of financial news texts. Past analyses of emotion or sentiment in financial news have relied on a selection of key words to denote positive or negative aspect, often derived from the researcher's intuition or

Osgood's classification [13]. This analysis aims to identify what aspects of the lexica of emotion are used in financial texts, which categories or dimensions of emotion predominate and how these emotionally charged terms collocate within the financial domain. We use a corpus of financial news texts which has been tagged for positive/negative polarity and a normative corpus of general world news to evaluate the characteristic distributions of affective language in the financial domain.

References

[1] K. Ahmad, D. Cheng, and Y. Almas. Multi–lingual sentiment analysis in financial news streams. In S. Cozzini, S. d'Addona, and R. Mantegna, editors, *Proc. of the 1st Intl Conf on Grid in Finance*, Palermo, Italy, 2006.

[2] T. G. Andersen, T. Bollerslev, F. X. Diebold, and C. Vega. Micro effects of macro announcements: Real-time price discovery in foreign exchange. Technical Report 8959, NBER, 2002.

[3] L. Bentivogli, P. Forner, B. Magnini, and E. Pianta. Revising wordnet domains hierarchy: Semantics, coverage, and balancing. In *Proceedings of COLING 2004 Workshop on Multilingual Linguistic Resources*, pages 101–108, Geneva, Switzerland, 2004.

[4] W. S. Chan. Stock price reaction to news and no-news: Drift and reversal after headlines. Available at SSRN: http://ssrn.com/abstract=262452 or DOI: 10.2139/ssrn.262452, 2001.

[5] R. P. DeGennaro and R. E. Shrieves. Public information releases, private information arrival and volatility in the foreign exchange market. *Journal of Empirical Finance*, 4(4):295–315, dec 1997.

[6] M. deGoede. Resocialising and repoliticising financial markets: Contours of social studies of finance. *Economic Sociology European Electronic Newslettter*, 6(3), july 2005.

[7] P. Ekman and W. V. Friesen. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17:124–129, 1971.

[8] R. F. Engle and V. K. Ng. Measuring and testing the impact of news on volatility. *Journal of Finance*, 48(5): 1749–1778, 1993.

[9] A. Esuli and F. Sebastiani. Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of LREC 2006 - 5th Conference on Language Resources and Evaluation*, Genova, Italy, may 2006.

[10] I. Hardie and D. MacKenzie. An economy of calculation: Agencement and distributed cognition in a hedge fund,. *British Journal of Sociology*, 2006. forthcoming.

[11] M. Koppel and I. Shtrimberg. Good news or bad news? let the market decide. In *AAAI Spring Symposium on Exploring Attitude and Affect in Text*, pages 86–88, Palo Alto CA, 2004. AAAI Press.

[12] M. L. Mitchell and J. H. Mulherin. The impact of public information on the stock market. *Journal of Finance*, 49(3):923–950, 1994.

[13] C. E. Osgood, G. J. Suci, and P. H. Tannenbaum. *The Measurement of meaning*. University of Illinois Press, Chicago, Ill, 1957.

[14] J. A. Russell. Affective space is bipolar. *Journal of Personality and Social Psychology*, 37:345–356, 1979.

[15] G. P. Smith. How high can a dead cat bounce?: Metaphor and the hong kong stock market. *Hong Kong Papers in Linguistics and Language Teaching*, 18, 1995.

[16] P. C. Tetlock. Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance*, 2007. forthcoming.

[17] P. D. Turney. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proc. of ACL'02*, pages 417–424, Philadelphia, Pennsylvania, 2002.

[18] C. Whissell. The dictionary of affect in language. In R. Plutchik and H. Kellerman, editors, *Emotion: theory research and experience*, volume 4, The measurement of emotions. Academic Press, London, 1989.

[19] J. Wiebe, E. Breck, C. Buckley, C. Cardie, P. Davis, B. Fraser, D. Litman, D. Pierce, E. Riloff, T. Wilson, D. Day, and M. Maybury. Recognizing and organizing opinions expressed in the world press. In *AAAI Spring Symposium on New Directions in Question Answering*, 2003.