

Analiza opiniilor implicite din texte economice și a relațiilor dintre ele

Această lucrare are ca scop analiza opiniilor exprimate în mediul economic, metodele fiind centrate pe extragerea opiniilor implicite și pe evaluarea relațiilor existente între acestea. Absența certitudinii și dificultatea alegerii unor metrici adecvate situației economice date duc la apariția unor elemente subiective ce folosesc exprimări aparent obiective ce folosesc indicatori economici și predicții ale evoluției ulterioare a acestora. Lucrarea analizează importanța opiniilor implicite astfel definite și compară acuratețea detecției subiectivității prin această metodă cu metode bine cunoscute bazate pe apariția opiniilor explicite. Sunt investigate totodată modalități de îmbunătățire a performanțelor metodei de extragere a opiniilor implicite, prin includerea influenței modificatorilor de valență, și prin studierea dependenței sistemului de extragere de caracteristicile textelor pe care se realizează analiza. Experimentele au demonstrat existența unei particularități a textelor cu un caracter economic, privind diminuatorii de intensitate, precum „mai puțin”, „slab” sau „mic”. Dacă în mod normal aceștia diminuează intensitatea polarității opiniei conținute într-o construcție, în această situație au mai degrabă rolul de a o inversa, comportându-se ca negații. Ulterior am examinat modul în care sistemul de extragere a opiniilor implicite poate fi utilizat pe texte obținute direct din internet, din webloguri sau publicații financiare. Obținerea unei acurateți similare celei din experimentele efectuate pe un corpus static de texte duce la concluzia ca în general în mediul economic predicțiile sunt realizate folosind un vocabular restrâns, ușor identificabil. A doua parte a tezei a fost dedicată unei abordări conceptuale pentru determinarea celor mai importante idei ce pot fi identificate în textele economice. Folosind *topice* statistic relevante și proiecția lor pe ontologia folosită, WordNet, am creat o modalitate nouă de a evalua din punct de vedere al coeziunii semantice atât rezultatul modelelor de *topice* cât și importanța individuală a celor mai relevante concepte din texte. Pentru validarea acestui sistem am utilizat răspunsurile a 37 de evaluatori umani, a căror părere a fost constant corelată cu rezultatele obținute automat. Conceptele asociate *topicelor* au fost folosite pentru etichetarea celor din urmă, definirea contextului și asocierea cu opiniile implicite extrase anterior. Corelațiile obținute prin asocierea conceptelor relevante cu polaritatea opiniilor implicite sunt coerente cu realitățile economice.

The Analysis of Implicit Opinions in Economic Texts and Relationships between Them

This work analyses opinions within the economic environment, centered on extracting implicit opinions and evaluating relations between them. The absence of certainty and the difficulty of choosing adequate metrics for a given situation lead to the emergence of subjectivity in the economic discourse, involving economic indicators and future state predictions. This work analyses the importance of these implicit opinions and compares the accuracy of this way of detecting subjectivity with a well-known standard one, based on explicit opinions. We also investigate ways of improving the retrieval accuracy by including the influence of valence shifters and by studying the system's dependence on the characteristics of the analyzed economic texts. Experiments show the existence of a particularity of economic texts, regarding diminishers, such as “less”, “weak” or “small”. While their regular role is to diminish the intensity of the polarity of an opinion within a construct, in this case they have the role of reversing the polarity altogether, acting as negations. Afterwards we examine the way the implicit opinion extraction works on texts extracted from the Internet, from weblogs and financial publications. By obtaining an accuracy similar to the one we observed on the initial static corpus we concluded that generally economic predictions are uttered using a small and predictable vocabulary. The second part of the thesis is dedicated to a conceptual approach for determining the most important ideas that permeate economic texts. By using statistically relevant *topics* and their projection unto an ontology, in this case WordNet, we created a novel method of evaluating, from a semantical coherence perspective, both the *topics* themselves and the most important concepts within the documents. We used the answers of 37 different human evaluators to assess the validity of our automatically obtained results. The concepts associated to the relevant topics were then used for topic labeling, for defining the idea of context within this framework and they were associated with previously extracted implicit opinions. The correlations obtained from this association are coherent with economic realities.