Sentiment Analysis: A Hard Pattern Classification Task

The case of isMOOD

This work presents the business case study of *isMOOD,* a startup company emphasizing on the field of social media analytics in order to extract meaningful knowledge concerning consumers' attitude towards products, services or brands in general. isMOOD, in particular, has developed a highly sophisticated software platform that incorporates a wide range of text mining and machine learning tools in order to provide its customers the ability to monitor and predict the public opinion and market sentiment. However, transforming the unstructured textual information contained within the various social media streams into useful business knowledge is an extremely difficult computational task, mainly, due to the underlying hard pattern classification problem of sentiment analysis within the context of the Greek language. In this study we conduct a thorough investigation of the pattern classification problem of sentiment analysis focusing on the particular sub-problems associated with efficient corpus vectorization of unseen text, the class imbalance problem, irony detection and Support Vector Machines – based classification analyzing the particular methodology adapted by isMOOD. Finally, we present an experimental comparison with a set of state-of-the-art machine learning classifiers on a benchmark dataset originating from the bank sector revealing the classification superiority of radial basis support vector machines.