

University : Istanbul Kültür University
Institute : Institute of Sciences
Department : Computer Engineering
Programme : Computer Engineering
Supervisor : Assoc. Prof. Dr. Çağatay ÇATAL
Degree Awarded and Date: MSc – July 2014

ABSTRACT

DETECTING DECEPTIVE CUSTOMER REVIEWS USING MACHINE LEARNING METHODS

Suat GÜLDAN - 2014

The competition among companies has been considerably increased in the recent years due to the significant developments in online shopping of services and the widespread usage of e-commerce. The product reviews became a primary factor shaping the buyers' decisions. Due to this factor, product reviews created a marketing area for fake reviews about products and services. In this thesis, a model which uses a multiple classifier system has been proposed to identify the negative deceptive customer reviews and the validation has been performed on a dataset which consists of hotel reviews. The proposed model has a better performance than the best model reported in literature for this problem. In this model, five classifiers have been applied by using majority voting combination rule. These classifiers are libLinear, libSVM, Sequential Minimal Optimization (SMO), Random Forest and J48. LibSVM and libLinear are two different implementations of support vector machines.

Keywords: Deceptive Review Detection, Support Vector Machines, Decision Trees, Multiple Classifier Systems, Vote, Machine Learning

Üniversite : İstanbul Kültür Üniversitesi
Enstitüsü : Fen Bilimleri Enstitüsü
Dalı : Bilgisayar Mühendisliđi
Programı : Bilgisayar Mühendisliđi
Danışmanı : Doç. Dr. Çağatay ÇATAL
Tez Türü ve Tarihi : Yüksek Lisans – Temmuz 2014

ÖZET

MAKİNE ÖĞRENMESİ YÖNTEMLERİYLE GERÇEK OLMAYAN TÜKETİCİ YORUMLARININ TESPİTİ

Suat GÜLDAN – 2014

Çevrimiçi hizmetlerin gelişmesi ve e-ticaretin yaygınlaşması sonucu, firmalar arası rekabet giderek artmıştır. Bu bağlamda ürün yorumları, satın alıcıların kararlarını şekillendiren önemli bir faktör olarak ortaya çıkmıştır. Bu etkinin sonucu olarak ürün yorumları, ürün ve hizmetler hakkında aldatıcı yorumların yapılabileceđi bir pazarlama alanı oluşturmuştur. Bu tezde, olumsuz aldatıcı tüketici yorumlarını tespit edebilmek üzere, çoklu sınıflayıcı sistemler kullanılarak bir model önerilmiş ve önerilen modelin otel yorumları ile ilgili olarak hazırlanmış olan veri kümesinde geçerlenmesi sağlanmıştır. Önerilen model, üzerinde çalışılan problem için literatürdeki en iyi modelden daha yüksek performans sunmuştur. Bu modelde beş sınıflayıcı çoğunluk oylaması birleşim kuralına göre kullanılmıştır. Bu sınıflayıcılar libLinear, libSVM, ardışık minimal optimizasyon (SMO), Random Forest ve J48'dir. LibSVM ve libLinear, Destek Vektör Makinelerinin (DVM) iki farklı gerçekleştirilmesi olarak bilinmektedir.

Anahtar Kelimeler: Aldatıcı Yorum Tespiti, Destek Vektör Makineleri, Karar Ağaçları, Çoklu Sınıflandırıcı Sistemler, Oylama, Makine Öğrenmesi