

University : **Istanbul Kültür University**
Institute : **Institute of Science**
Department : **Computer Engineering**
Programme : **Computer Engineering**
Supervisor : **Prof. Dr. Coşkun Bayrak**
Degree Awarded and Date : **MS – January 2013**

ABSTRACT

SENTIMENT ANALYSIS FOR TURKISH LANGUAGE

Hakan Çelik

Sentiment analysis a.k.a. opinion mining is an application of natural language processing to extract subjective information from given texts. Consumer comments, evaluation of films, stock exchange predictions and political researches are some examples of subjective declarations that sentiment analysis involves.

Subjective data have hugely increased parallel to web 2.0 and social media growth in recent years. People generate petabytes of data every day. Processing of this amount of raw data became more and more important for both companies and individuals.

Many sentiment analysis studies conducted in the past are focused on English Language only. Therefore, for Turkish Language, sentiment analysis is a blue ocean yet. There is a single study focused on only one input domain and tested with only one machine learning algorithm.

Within the focus of this thesis new datasets were constructed in various domains, a Turkish specific input preparation algorithm was introduced, and different machine learning algorithms were applied over prepared data. ~85% accuracy was achieved with Naïve Bayes Classifier.

Keywords: Sentiment Analysis, Machine Learning, Natural Language Processing, Opinion Mining, Data Mining, Turkish, Text polarity, Naïve Bayes, Support Vector Machines, Logistic Regression, Decision Tree

Üniversite : İstanbul Kültür Üniversitesi
Enstitü : Fen Bilimleri Enstitüsü
Dal : Bilgisayar Mühendisliği
Program : Bilgisayar Mühendisliği
Tez Danışmanı : Prof. Dr. Coşkun Bayrak
Tez Türü ve Tarihi : Yüksek Lisans – Ocak 2013

KISA ÖZET

TÜRK DİLİ İÇİN DUYGU ANALİZİ

Hakan Çelik

Duygu yada düşünce analizi metinlerden kişisel değerlendirmelerin çıkarımını sağlayan doğal dil işleme branşıdır. Duygu analizinin ilgi alanına giren değerlendirmelere örnek olarak tüketici yorumları, film değerlendirmeleri, borsa tahminleri ve siyasi araştırmalar gibi geniş ölçekte örnekler verilebilir.

Son yıllarda web 2.0 ve sosyal medyanın artış hızına bağlı olarak öznel değerlendirmeler müthiş derecede arttı. Her gün insanlar petabaytlarca veri girişi yapıyor. Bu boyuttaki ham verinin analizi gerek bireyler gerekse de şirketler için gittikçe daha fazla önem arz etmeye başladı.

İngilizce için birçok duygu analizi çalışmaları mevcut. Ancak bu çalışma alanı Türkçe için hala bakir. Gerçekleştirilen tek çalışma tek bir alan üzerine yoğunlaşmış tek tipte veri üzerinde tek tip makina öğrenmesi algoritması kullanılmış.

Bu tez kapsamında farklı alanlara ait yeni veri grupları oluşturulmuş, Türkçe'ye özel veri hazırlama algoritması tanıtılmış ve oluşturulan veri üzerinde farklı makina öğrenmesi algoritmaları uygulanmıştır. Naïve Bayes sınıflandırıcı kullanılarak başarılı sayılabilecek %85'lik doğruluk oranı yakalanmıştır.

Anahtar Kelimeler: Duygu Analizi, Makine Öğrenmesi, Doğal Dil İşleme, Fikir Madenciliği, Veri Madenciliği, Türkçe, Metin Kutbu, Naïve Bayes, Destek Vektör Makineleri, Mantıksal Regresyon, Karar Ağaçları