

ÖZET

TÜRKÇE DOKÜMANLAR İÇİN YAZAR TANIMA

Özcan KOLYİĞİT

Yüksek Lisans Tezi, Matematik Anabilim Dalı
Tez Danışmanı: Yrd. Doç. Dr. Rıfat AŞLIYAN
2013, 57 sayfa

Günümüzde, yazar tanıma çalışmaları, teknolojinin gelişmesi ve bilginin yaygınlaşması ile ortaya çıkan bir takım sorunlara çözüm üretmek için yapılmaktadır. Bu sorunlardan bazıları yazarı belli olmayan dokümanların yazarlarının belirlenmesi ve yazarının kim olduğundan tam olarak emin olunamayan metinlerin yazarlarının belirlenmesidir. Bu çalışmada, Türkçe dokümanlar için yazar tanıma sistemleri geliştirilmiştir. Sistemlerin eğitilmesinde ve test edilmesinde kullanılmak üzere, gazetelerden seçilen 6 yazara ait köşe yazıları kullanılmıştır. Yazarların 70'er makalesinden oluşan 420 dokümandan oluşan bir derlem hazırlanmıştır. Bu dokümanlardan 20'şer tanesi eğitim için, 50'şer tanesi test için kullanılmıştır. İlk olarak, 6 yazara ait dokümanlar toplanmış, daha sonra her yazara ait 20 doküman birleştirilerek tek bir doküman haline getirilmiştir. Bu şekilde elde edilen 6 doküman için sözcük, gövde, hece ve karakter n -gramlarının öznitelik vektörleri belirlenmiştir. K-En Yakın Komşu algoritması için öznitelik vektörleri belirlenirken her yazar için vektör uzunlukları 120, 180 ve 240 olarak seçilmiş, oluşan öznitelik vektörleri için K-En Yakın Komşu algoritmasıyla test edilmiştir. En başarılı sonuçlar, vektör boyu 120 olduğunda elde edildiğinden diğer metotlar için de vektör boyu 120 olarak kullanılmıştır. Geliştirilen sistemler eğitildikten sonra test edilerek doğruluk ve F-ölçüsü değerlerine göre birbirleriyle karşılaştırılmıştır.

Anahtar sözcükler: Yazar Tanıma, K-En Yakın Komşu, Çok Katmanlı Algılayıcı, Destek Vektör Makinesi, LVQ, n -gram.