

NLP İLE APRIÖRİ ALGORİTMASINI İYİLEŞTİRME

Taylan ÖNDER
Gebze Teknik Üniversitesi
taydu34@gmail.com

ÖZET

Veri kümelerindeki ilişkilendirme algoritması olan apriori algoritmasını, daha kesin sonuçlar verebilmesi ve veriyi daha saf haliyle ele alabilmesi için NLP algoritması ile veri kümelerindeki elemanların, syntatic bütünlüğüne göre tekrar düzenleyip, düzenlenmiş veriler üzerinden ilişki kurabilmesi için çalışma yapıldı. Veriler tekrar düzenlenip, anlamsal olarak bütünlük sağlandığında ve küçük hatalar giderildiğinde, sonuçların daha beklendiği gibi çıktığı saptandı.

Anahtar Kelimeler: NLP, Apriori, Veri Kümesi, Veriler Arası İlişki, Kelime Kökü

1. GİRİŞ

NLP İLE APRIÖRİ ALGORİTMASINI İYİLEŞTİRME isimli bu projede apriori algoritmasının düzgün uyarlanabilmesi için NLP yaklaşımının ön işlemenin Tokenization (kelimeleri ayırma) ve Stemming (kelimelerin kökünü bulma) adımlarından faydalanılmıştır. Verilen veri kümesindeki kelimeler içerisinde Türkçe karakter yanlışlıkları ve farklı yapım ve çekim eki almış kelimeler için kelime kökleri üzerinden ilişkilerine bakılarak apriori algoritmasıyla daha kesin ve doğru sonuçlar ortaya çıkarılmaya çalışılmıştır. Normal bir apriori yaklaşımını aslında NLP algoritması ile uyumlu şekilde geliştirilmiştir. Veriler okunurken öncelikle sözlükte dağarcığımızdaki kelimeler ile karşılaştırıp Türkçe karakter değişikliğine gerek olup olmadığına bakarak başlayıp, sonrasında kelimenin kökleri kontrol edilerek, kelimenin kök halinde olup olmadığını kontrol edip büyük ya da küçük olmaksızın aralarında bir ilişki kurabilmek için ortak format olarak verilerin hepsinde harfleri büyük hale çevrilmiştir. Apriori uygulaması ile büyük harflerle kökleri alınmış bu veriler daha sonra verilerin kendi aralarındaki sık kullanım miktarlarına bakılarak, ilişkilendirilmesi istenen veri ile aralarındaki ilişki çözümlenmiştir. Çalışma şu şekilde devam etmektedir. İkinci bölümde örnek verilmiş veri kümesi ve çıktı sonuçları ele alınmıştır. Son olarak sonuç kısmında önemli sonuçlar belirtilmiş, ayrıca konu ile ilgili muhtemel araştırma konuları belirtilmiştir.

2. İKİNCİ BÖLÜM

Bu bölümde örnek olarak verilen veri kümesi için alınan sonuçlardan bahsedilecektir. Karışık olarak verilmiş beklenen veri kümesinden hatalı veya değiştirilmiş olduğunu düşündüğümüz veri kümesindeki ilişki kurma algoritmamızın çalışma mantığını gözlemleyeceğiz.

İlişki çıkarılacak ürünler aşağıdaki tablodaki gibidir.

Satılan Ürünler
Ekmek
Bebek Bezi
Kola
Bira
Süt

Ekmekte,Sutle
 Ekmekte,Bebekle Bezi,Bira,Yumurta
 Sutle,Bebek Bezi,Bira,Kola
 Ekmekte,Sutle,Bebek Bezi,Bira
 Ekmekte,Sutle,Bebek Bezi,Köla
 Ekmekte,Sutle
 Ekmekte,Köla,Bira,Sütle
 Sutle,Ekmekte,Bira,Köla
 Ekmekte,Sutle,Bebek Bezi,Bira
 Ekmekte,Bira,Bebek Bezi,Bebek Bezi

Şekil 1

Yukarıdaki şekil 1’de verilen veri kümesi kökünden farklı, ek alarak farklılaşmış ve sesli harf değişikliği yapılmış verileri içermektedir. Bu bilgilerden doğru şekilde faydalanabilmek için NLP kullanılarak text üzerinde anlamlı veriler elde edildi ve aşağıdaki şekildeki gibi veriler arasındaki ilişki oluşmuş oldu.

<pre> frequent 1-term set: ----- ['EKMEK'] ['BİRA'] ['KOLA'] ['BEBEK BEZİ'] ['SÜT'] frequent 2-term set: ----- ['EKMEK', 'BİRA'] ['EKMEK', 'KOLA'] ['EKMEK', 'BEBEK BEZİ'] ['BEBEK BEZİ', 'SÜT'] ['KOLA', 'SÜT'] ['KOLA', 'BEBEK BEZİ'] ['BEBEK BEZİ', 'BİRA'] ['SÜT', 'BİRA'] ['EKMEK', 'SÜT'] ['KOLA', 'BİRA'] </pre>	<pre> frequent 3-term set: ----- ['KOLA', 'SÜT', 'BİRA'] ['EKMEK', 'SÜT', 'BİRA'] ['EKMEK', 'KOLA', 'SÜT'] ['EKMEK', 'BEBEK BEZİ', 'SÜT'] ['EKMEK', 'KOLA', 'BİRA'] ['KOLA', 'BEBEK BEZİ', 'SÜT'] ['EKMEK', 'BEBEK BEZİ', 'BİRA'] ['BEBEK BEZİ', 'SÜT', 'BİRA'] frequent 4-term set: ----- ['EKMEK', 'BİRA', 'SÜT', 'KOLA'] ['EKMEK', 'BEBEK BEZİ', 'BİRA', 'SÜT'] </pre>
<pre> ----- rules refer to ['EKMEK'] ['BİRA'] -> ['EKMEK']: 0.8571428571428572 ['KOLA'] -> ['EKMEK']: 0.7499999999999999 ['BEBEK BEZİ'] -> ['EKMEK']: 0.8 ['SÜT'] -> ['EKMEK']: 0.8749999999999999 ['SÜT', 'BİRA'] -> ['EKMEK']: 0.8 ['KOLA', 'SÜT'] -> ['EKMEK']: 0.7499999999999999 ['BEBEK BEZİ', 'SÜT'] -> ['EKMEK']: 0.7499999999999999 ['KOLA', 'BİRA'] -> ['EKMEK']: 0.6666666666666667 ['BEBEK BEZİ', 'BİRA'] -> ['EKMEK']: 0.7499999999999999 ['KOLA', 'SÜT', 'BİRA'] -> ['EKMEK']: 0.6666666666666667 ['BEBEK BEZİ', 'SÜT', 'BİRA'] -> ['EKMEK']: 0.6666666666666667 </pre>	

3. SONUÇ

Apriori algoritması NLP yöntemleri kullanılarak daha keskin sonuçlar verebileceği projeden beklenen sonucu vermiş oldu. Önemsiz görülebilecek küçük yanlışlık yapılmış veriler ya da beklenmedik hatalı kaydedilmiş büyük veriler üzerinde yapılan değişiklikler ile veri setleri arasındaki ilişki daha kanıtlanabilir bir sonuca ulaşmış oldu. İleriki aşamalarda geniş bir ürün havuzu verilerek isim benzerliğine sahip ürünlerde tercih edilme yoğunluklarına göre eğitilmiş bir program ile daha doğru sonuçlar veren hata oranını daha da aza indiren bir program geliştirilerek daha kesin sonuçlar alınması mümkündür.

KAYNAKLAR

Kitap:

Han j, Kamber M, Pei j. *Data Mining: Concepts and Techniques 3th ed.*, Morgan Kaufmann, 2000.

İnternet Sitesi

<http://bilgisayarkavramlari.sadievrenseker.com/2011/09/07/apriori-algoritmasi/>
<https://docplayer.biz.tr/14530243-Apriori-algoritmasi-ve-turkiye-deki-ornek-uygulamalari.html>
Gtü moodle / Ders Sunuları