

Trafik Yoğunluğu Twitter Verilerinin Semantik Yöntemler İle Çözülmesi

¹Çağrı Kaçmaz

¹Galatasaray University, Engineering Management Department MSc, Software Development Consultant at KocSistem Inc., TURKEY

Özet

Konu alınan çalışmanın amacı; Twitter'dan sistematik bir düzen ile alınan verilerin, oluşturulan ontoloji ile trafik yoğunluğunu analiz etmek için anlamlandırılmasıdır. Dinamik olarak ulaşılan ve gerekli zaman dilimlerinde güncellenen Twitter trafik verilerinin, önceden tasarlanmış metodoloji ile Türkçe dil yapısı göz önünde bulundurularak anlamlandırılmasını ve analiz edilmesini sağlayan bir yöntem ortaya konulmaktadır.

Bu çalışmada; "Akıllı Şehirler" konusu bazında uluslararası bir Ar-Ge projesi olan INSIST için geliştirilen analiz yöntemi kullanılmıştır. Sosyal medya yorumlarından trafik yoğunluk bilgisinin elde edilmesini sağlayan bir yöntem sunulmuştur. Trafik yoğunluk bilgileri, kullanıcıların bilgilendirilmesi amacıyla, kurumsal hesaplar aracılığı ile sosyal medya ortamlarında paylaşılmaktadır. Sunulan yöntem sayesinde bahsi geçen yorumların semantik anlamlandırma yöntemleri ile çözümlenmesi, konum, yoğunluk durumu ve olay bilgilerin elde edilmesi sağlanmış ve elde edilen bilgiler sayısal harita uygulamasına aktarılmıştır. [2], Konuma ilişkin verilerin elde edilmesi için standart yöntemler kullanılmıştır. Yoğunluk ve olay bilgilerinin elde edilmesi için ise trafik özelinde oluşturulan ontolojiler kullanılmıştır. Açık olarak ifade edilmiş ve ima özelliği taşımayan cümleler için yüksek doğruluk elde edilmesi proje bazında sağlanmıştır. İma özelliği taşıyan kelimeleri içeren yorumlar için yapay sinir ağları kullanan yöntemler önerilmiştir.

Anahtar Kelimeler: twitter içerik analizi, trafik yoğunluğu analiz problemi, nesne tanımlama, semantik algoritma ile dil anlamlandırması

Abstract

In this work, we presented the ontology is created to make sense of the data taken in a systematic order to analyze the traffic from Twitter. A procedure providing to interpret and analyze the data which is acquired dynamically and updated on the essential periods of time with a predesigned methodology and by considering in the structure of Turkish language has been presented.

In this study, "Smart Cities" focused on the subject has been developed in international R & D projects, which INSIST the project. A method of obtaining the traffic information, comments, social media is presented. Traffic density information, in order to inform the citizens, it is shared on social media and corporate accounts. Traffic density information, in order to inform the citizens, it is shared on social media and

corporate accounts. The proposed method to solve this interpretation of semantic interpretation methods, location, used to obtain the intensity status and event information, and the information obtained has been transferred to digital map applications. Standard methods were used to obtain data on the subject; and event information to obtain the density of traffic generated ontologies are used in special. High accuracy is obtained to open the expressed and implied feature carry sentences. A method that uses neural networks for comments that contain the word carries the implication feature are recommended.

Key words: twitter content analysis, traffic jam analysis problem, object tracking, language definition with semantic algorithm

1. Giriş

21. yüzyıla teknoloji ve insan perspektifinde bakıldığında, yaşamın birçok alanında içiçe geçilen bir çevre yapısı sunulmaktadır. Bu süreçte insan sağlığını, güvenliğini ve konforunu artıran en önemli unsur yeni teknolojilerle insan yaşamına giren uygulama ve cihazlardır.

Bu çalışmada sunulan yöntem, ‘Akıllı Şehirler’ konusuna odaklı olarak geliştirilen uluslararası bir Ar-Ge projesi olan INSIST projesi kapsamında geliştirilmiştir. Modern şehir yaşamının en büyük problemlerinden biri olan araç trafiği yoğunluğu üzerine geliştirilen çalışmada, sosyal medyanın bilgi paylaşım gücünden yararlanılarak trafik yoğunluğunun ölçülmesi ve anlamlandırılması üzerine odaklanılmaktadır.

Pilot bölge olarak İstanbul üzerinde gerçekleştirilen çalışmada, Twitter’da yer alan kurumsal trafik hesaplarının gönderdikleri yorumlar (tweet), belirli zaman aralıklarıyla veritabanına kaydedilmektedir. Bu yorumlar üzerinden bölge, yoğunluk, durum belirten kelimeler analiz edilerek trafik yoğunluğunun çekilmesi ve trafik konusu özelinde oluşturulan ontolojiler ile anlamlandırma süreçleri uygulanmaktadır. [4]

2. Materyal ve Yöntemler

Çalışma kapsamında geliştirilen yöntem şu adımlardan oluşmaktadır:

1. İçeriğin sosyal medya ortamından (Twitter) alınması
2. İçeriğin veritabanı ortamında (Cassandra) saklanması [1],
3. Anlamlandırma (semantik) algoritmaları ile içeriklerin değerlendirilmesi [10],
 - a. Her yorum cümlelerine ayrılır, her bir cümle ayrı ayrı ele alınır.
 - b. Her bir cümle kelimelerine ayrılır
 - c. Her bir kelime için trafik özelinde oluşturulan ontolojiler kullanılarak semantik değerlendirme yapılır. Bu çalışmada yapılan semantik değerlendirme 3 başlık altında toplanabilir:
 - i. Yorumdaki coğrafi konum verisinin bulunması (hangi ilçe, hangi yol)
 - ii. Yorumdaki trafik yoğunluk bilgisinin alınması (yoğun, akıcı, orta derecede yoğun)

- iii. Yorumdaki trafik yoğunluğuna neden olan olay bilgisinin alınması (kaza, buzlanma gibi olaylar)
- d. Üç kırılımda gerçekleştirilen bu değerlendirmeye ek olarak bazı yorumlar için genel değerlendirme yapılması da gerekli görülmüştür. Örnek olarak “Şişli Bomonti yolu üzerinde kaza yok” şeklinde bir yorumda yer alan kaza ifadesinden trafiğin yoğun olduğu sonucu çıkartılmamalı; cümlenin geneli anlamlandırılarak trafiğin akıcı olduğu sonucuna varılmalıdır. Ancak “Şişli Bomonti yolunda kaza” şeklinde bir yorumdan, trafiğin yoğun olduğu sonucu üretilmelidir. Bu özel durum için, semantik değerlendirme yeterli olmamaktadır, yapay sinir ağları kullanılarak anlamlandırma yapılmalıdır. [9]

A. Trafik Analizi İçin Kullanılacak Verilerin Twitter Üzerinde Yapılandırılması

Twitter’da trafik analizini güncel ve güvenilir veriler ile beslemek için kurumsal hesaplardan oluşan ‘Trafik Bilgilendirme Listesi’ oluşturulmuştur:



Şekil 1 Twitter Kurumsal Veri Akışı

Kurumsal listeyi verisel yönden beslemek, aynı zamanda tek bir kaynak üzerinde bağımlılık yaratmamak için, Twitter’da İstanbul trafiği ile ilgili yoğunluk yorumları paylaşan önemli hesapların bulunduğu ikinci bir liste oluşturulmuştur.

GenelTrafikVerileri
A public list by Insist

MEMBERS: 3 SUBSCRIBERS: 0

Edit Delete

Tweets >

List members >

List subscribers >

More lists by @InsistKoc · View all

GenelTrafikVerileri
KurumsalTrafikVerileri
Create new list

Tweets

Cezayir Doğan Retweeted
Murat Babür @MuroBabur · 4h
@cezayirdogan nuhkuyusundaki sıkıntı sadece çalışmadan kaynaklı avm önünden köprüye yol açık

Murat Kazanasmaz Retweeted
mesudguven @mesudguven · 4h
Konut-2'den Eskişehir Yolu'na cikista kaza... yol kapalı.. @maxfmmusic @ozguraksuna

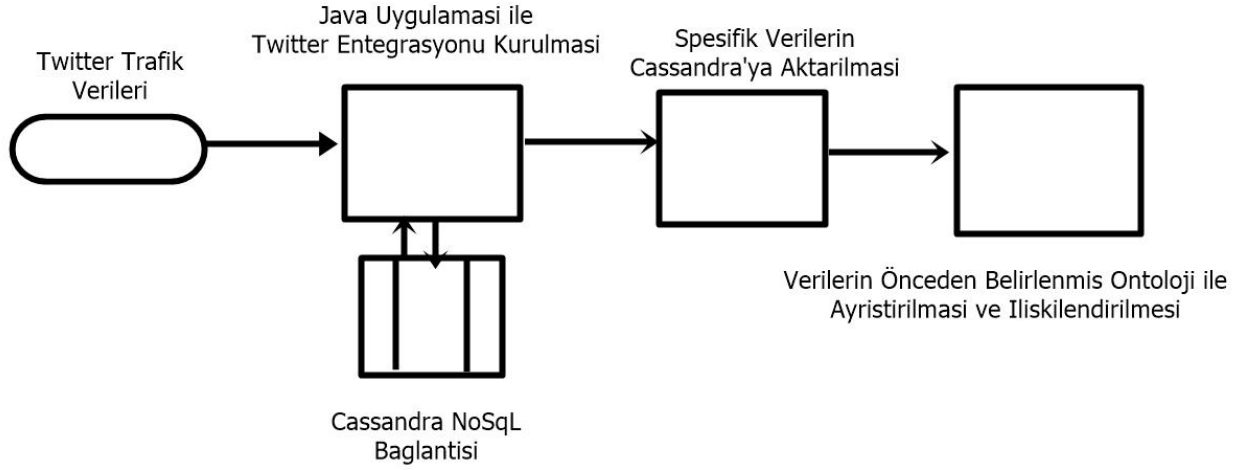
Murat Kazanasmaz Retweeted
Gökhan Topkaraoğlu @GokhanFB1984 · 4h
Fındığı emniyet şeridinde mi topluyorsunuz @tadelle @MuratKazanasmaz @uzayligul @sekeremurat @nihatsirdar

Şekil 2 Twitter Kullanıcı Veri Akışı

B. Twitter Üzerinde Toplanan Verinin NoSQL Database İle İlişkilendirilmesi

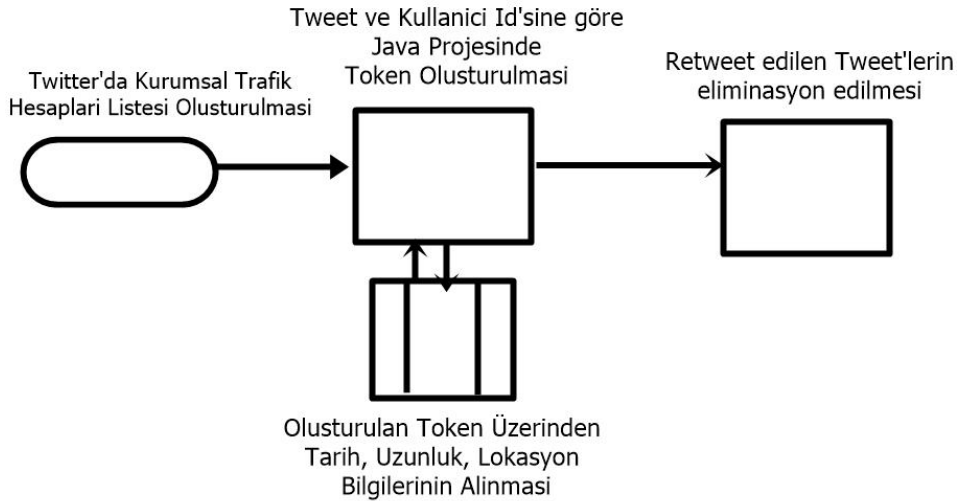
Twitter platformunda oluşturduğumuz ve sürekli güncellenen trafik bilgi akışı listelerini, metodolojimize uygun şekilde anlamlandırma çalışmalarına hazırlamak için altyapı NoSQL veritabanı ile ilişkilendirilmiştir. Bunun için Maven Dependencies kütüphanesi kullanarak Java platformunda, Cassandra ile Twitter'ı projeye uygun şekilde entegre edecek uygulama geliştirilmiştir. Oluşturulan yöntemde NoSQL kullanımının sebebi, ileride ekleyebileceğimiz metodolojileri, esnek bir yapıda barındırmak için klasik ilişkisel veritabanı sistemleri yerine yatay olarak ölçeklendirilen bir veri depolama sistemi kullanmaktır. Ayrıca sosyal medya ortamından toplanacak olan trafik yoğunluk yorumlarının özellikle İstanbul gibi megakent olan şehirler için çok fazla sayıda olma ihtimali ve bu nedenle verinin büyük veri niteliği taşıyabilecek nitelikte olmasıdır.

Şekil 3'te Cassandra ortamında veritabanı tablosunu oluşturan akış diyagramı görülmektedir.



Şekil 3 Verilerin İlişkilendirilmesi Sürecini Gösteren Akış Diyagramı

Şekil 4'te Twitter ortamından yorumları alarak Cassandra ortamına ekleyen kod bloğu görülmektedir.



Şekil 4 Twitter Spesifik Veri Çekimi Akış Diyagramı

Bu uygulama ile oluşturulan listelerde bulunan yorumlara ilişkin aşağıda verilen bilgiler platforma çekilerek anlamlandırma için gerekli ortam sağlanmaktadır:

- Gönderilen yorum içeriği
- Yorumun gönderildiği zaman
- Uygulamanın yorum veritabanına kaydedildiği zaman
- Gönderilen yorumun Twitter'daki tekil numarası (Unique Id)
- Yorumu gönderen hesabın kullanıcı ismi

C. Semantik Anlamlandırma Algoritmaları ile Trafik Yoğunluğunu Çözümleyen Metodoloji

Oluşturulan yapılar ile sosyal medya ortamından alınan trafik bilgi akışı, A ve B maddelerinde açıklanan yöntemler ile elde edildikten sonra aşağıda belirtilen akıştan geçirilirler. [5],

- Her yorum önce cümlelerine ayrılır, her bir cümle ayrı ayrı ele alınır. Bu sayede farklı verilerin birbirinden sağlıklı bir yapıyla ayrışması ve bilgi karmaşasının yaşanması engellenir.
- Her bir cümle kelimelerine ayrılarak oluşturulan ontoloji ile anlamlandırma aşamalarında düzenlenir. Şekil 5'te bu ayrıştırmadan oluşan kelime yapısı ağacı görülmektedir.

- c. Her bir kelime için semantik değerlendirme yapılır. Bu çalışmada yapılan semantik değerlendirme 3 başlık altında değerlendirilebilir:
- i. Yorumdaki coğrafi konum verisinin bulunması (ilçe yol bilgileri): Çalışma kapsamında pilot bölge olarak seçilen İstanbul ili için semt, ilçe, mahalle, yol adı verileri elde edilmiş ve bir tabloda saklanmıştır. Yorum çözümlenirken yorumda yer alan kelimeler bu tablo ile karşılaştırılarak yorumda bahsi geçen konumun elde edilmesine çalışılmıştır.
Örnek: “Şişli Bomonti yolu üzerinde ölümlü kaza, trafik yoğun.” yorumu çözümlenerek Şişli ve Bomonti kelimeleri elde edilmiştir.
 - ii. Yorumdaki trafik yoğunluk bilgisinin alınması (yoğun, akıcı, orta derecede yoğun): Bu bilginin elde edilmesi yoğunluk belirten kelimelerden oluşan bir ontoloji oluşturulmuştur. Yorumda yer alan kelimeler bu ontoloji ile taranarak bahsi geçen trafik yoğunluğu bilgisinin elde edilmesine çalışılmıştır. [6],
Örnek: “Şişli Bomonti yolu üzerinde ölümlü kaza, trafik yoğun.” yorumu çözümlenerek yoğun kelimesi elde edilmiştir.
 - iii. Yorumdaki trafik yoğunluğuna neden olan olay bilgisinin alınması (kaza, buzlanma gibi olaylar): Bu bilginin elde edilmesi olay belirten kelimelerden oluşan bir ontoloji oluşturulmuştur. Yorumda yer alan kelimeler bu ontoloji ile taranarak bahsi geçen olay bilgisinin elde edilmesine çalışılmıştır. [7]
Örnek: “Şişli Bomonti yolu üzerinde ölümlü kaza, trafik yoğun.” yorumu çözümlenerek ölümlü kaza kelimesi elde edilmiştir.

Test ve Uygulama

Oluşturulan yöntemler ve elde edilen çalışmalar sonrasında yöntemin test edilmesi için Twitter’da oluşturulan havuzdan 45 adet trafik yorumu seçilmiştir. Bu yorumların 17’si trafik kazası, 8’i yaralanmalı kaza, 14’ü yoğunluk durumu, 6’sı trafik yoğunluğu öngörüsü içermektedir. Trafik yoğunluğu analizi için oluşturduğumuz yapının testleri bu 45 yorum ile yapılmıştır. Aşağıdaki listede, yorum havuzundan örnekler görülmektedir [3],.

- IBB Trafik Kontrol Merkezi hesabının: “Tem Kurtköy Kavşak-Sultanbeyli Yönü orta şerit Trafik kazası (hasarlı). tüm şeritler trafiğe kapalı. Ekip Sevk Edildi.”
- IBB Trafik Radyosu hesabının: “D100 Boğaziçi-Zincirlikuyu yönü sağ şeritte #araçarızası nedeniyle bölge trafiği yoğun.”
- IBB Trafik Kontrol Merkezi Hesabının “D100 Boğaziçi-Zincirlikuyu Yönü sağ şerit araç arızası bölgedeki trafiği yoğunlaştırmaktadır.”
- Radyo Trafik hesabının “Başbüyük orman yolu Maltepe Üniversitesi önünde, Sancaktepe yönünde kaza!”
- IBB Trafik Kontrol Merkezi hesabının: “Tem Bahçeşehir(Yan Yol)-Hadımköy Yönü sol şerit Trafik kazası (yaralanmalı). 1 şerit trafiğe kapalı. Kazaya Müdahale Ediliyor.

3. Bulgular

Sunulan çalışmada geliştirilen yöntemin yorum havuzunda yer alan 45 yorum üzerinde test edilmesi sonucunda 38 yorumda başarılı bir şekilde çözümleme yapılabilmektedir. Ancak 7 yorum hatalı olarak anlamlandırılmıştır. Bunun nedeni, açık olarak yazılmamış, ima içeren cümlelerdir. Örnek olarak “Şişli Bomonti yolu üzerinde kaza yok, akıcı” gibi bir yorum çözümlendiğinde, yorum içerisinde geçen ‘kaza’ kelimesi, yine yorum içerisinde geçen ‘akıcı’ kelimesi ile çelişerek yöntemin karar verememesine yol açmaktadır. [8]

4. Tartışma

Açık olarak yazılmamış, ima içeren cümlelerin yer aldığı yorumların doğru bir şekilde analiz edilebilmesi için yapay sinir ağı yöntemleri kullanılarak geliştirilecek uygulamalara gereksinim vardır. Bu tür yorumları içeren bir veri seti ile eğitilecek bir ağın kullanımı sonucunda uygulama sonuç doğruluğu artırılabilir.

Yöntemin ait olduğu INSIST projesi uluslararası bir proje olduğu için, proje kapsamında geliştirilecek uygulamalar sadece Türkiye’de değil tüm Avrupa ülkelerinde çalışacaktır. Ancak bu yöntemde oluşturulan ontolojiler, test aşamasında pilot bölgeye uygun olacak şekilde Türkçe dilinde oluşturulmuştur. Proje sürecinde bu ontolojilerin İngilizce ve diğer Avrupa dillerine uyarlanması da yapılacaktır.

Sonuçlar

Bu çalışmada görüldüğü üzere, günümüz akıllı şehirlerinin en büyük sorunlarının başında gelen trafik yoğunluğu üzerine yapılan analizleri genişletmek mümkündür. Günümüz teknolojisi ve iletişim dünyasında sosyal medya üzerinden yoğun kullanılan mecraların, sağladığı bilgi akışı çok büyük rakamlara ulaşmıştır. Bu bilgi akışını istenilen çalışma için anlamlandırmak ve sağlıklı bir yapı ortaya çıkarmak için kurumsal firmaların, belediyelerin, devlet kuruluşların sosyal medya hesapları baz alınarak verilen bilgiler işlenebilmektedir.

Trafik akış yoğunluğunu analiz etmek için dünyanın en büyük sosyal medya platformlarından biri olan Twitter'in kullanıldığı bu çalışmada, gerek dijital platformun en hızlı iletişim aracı olması, gerekse aktif bir düzlemde kullanılması sebebiyle anlık trafik yoğunluğu verilerine ulaşılabilmektedir. Elde edilen veriler bölge, semt, yol gibi spesifik kavramlarda ayrıştırıldığında, gelen bilgiler anlamlandırma aşamalarından geçerek gerekli alanlar için bilgi kutuları oluşturulmaktadır. Oluşturulan bilgi kutularında toplanan 'yoğun', 'trafik açık', 'kaza' gibi anahtar kelimeler semantic yöntemler ile çözümlenerek gerekli güncel trafik yoğunluk haritası çıkarılabilmektedir.

Alıntılama

Bu çalışmada KoçSistem'in paydaş olarak yer aldığı INSIST (Itea 2- TÜBİTAK Teydeb 1509) projesinin akıllı şehirler üzerine yaptığı çalışmalardan faydalanılmıştır.

Referanslar

- [1] Semantic Decoder which Exploits Syntactic – Semantic Parsin, for the Town Info Task – James Henderson – Feb 2009
- [2] Semantic Frame Identification with Distributed Word Representations – Karl Moritz Hermann, Dipanjan Das, Jason Weston, Kuzman Ganchev – 2010
- [3] Attirubtion in Turkish with Dependency Tree Analyses – İlker Aydın – 2013
- [4] Joint Learning of Words and Meaning Representations for Open-Text Semantic Parsing – Antoine Bordes, Xavier Glorot, Jason Weston, Yoshua Bengio – 2012
- [5] From Distributed Semantics to Discourse, and Back – Jacob Eisenstein – 2015
- [6] Grounded Unsupervised Semantic Parsing – Hoifung Poon – 2011
- [7] Natural Language Processing – Ann Copestake – 2004
- [8] Frame – Semantic Parsing – Nathan Schneider – 2015
- [9] Semantic Parsing – Horacio Rodriguez – 2009
- [10] Harun Reşit Zafer - <https://github.com/hrzafer/fatih-parser>