

Münazaraların Twitter'da Etkisinin Araştırılması

Ahmet Yıldırım¹, Suzan Üsküdarlı¹

¹ Boğaziçi Üniversitesi, Bilgisayar Mühendisliği Bölümü, İstanbul
ahmet.yil@boun.edu.tr, suzan.uskudarli@boun.edu.tr

Özet: Sosyal medya ve sosyal medyada kısa mesajların yaygın olarak kullanılmaya başlamasının ardından, bu kısa mesajların içerdiği bilginin nasıl anlaşılabilceğinin araştırılması da önemli olmaya başladı. Artık araştırma kurumları ve şirketler bu bilgileri analiz ederek anlam çıkarmaya çalışıyorlar. Bu bildiri, münazaraları, örnek olarak da Amerika Birleşik Devletleri 2012 seçimlerinde yapılan başkan ve başkan yardımcısı münazaralarının, dünyadaki en büyük kısa mesaj paylaşım sistemi olan Twitter üzerinde nasıl etki bulduğunu göstermektedir. Bu bildiri kapsamında bu münazaralardan başkan yardımcılığı münazarası gösterilecektir. Bu amaçla Twitter Programcı Arayüzü (Twitter API) sorgulanmış, münazara boyunca ilgili veriler indirilmiş ve münazaranın yazılı hali web üzerinden toplanmıştır. Bu verilerin bazı istatistiksel özellikleri gösterilmekte, sonraki araştırma hedefleri hakkında bilgi verilmektedir.

Anahtar Sözcükler: Kısa mesajlar, Twitter, Sosyal Media, Seçimler, Münazaralar

Revealing Effect of Debates In Twitter

Abstract: Recently, after wide usage of social media and short messages in social media, analyzing the data in social media received attention. Research institutes and companies are trying to extract meaning from these data. This paper shows the effect of debates more specifically vice presidential debate of USA elections 2012 in Twitter which is the most widely used social media short messaging platform. For this purpose, Twitter Streaming API is queried. Datasets of tweets related to the debate and the transcript of the debate were collected from the Internet. Some statistical properties and future research direction are given.

Keywords: Short messages, Twitter, Social Media, Elections, Debates

1. Giriş

Sosyal medya ve sosyal medyadaki kısa mesajların kullanımının artması ile, bu sistemler üzerinde oluşan bilginin anlamının araştırılması önemli hale gelmiştir. Araştırma kurumları ve şirketler bu anlamda çalışmalar yapmaktadırlar. Bu bildiri de Amerika Birleşik Devletleri 2012 seçimlerinde yapılan münazaralar esnasında oluşan sosyal medya kısa mesaj verisi incelendi. Bu kısa mesajlar dünyanın en büyük kısa mesaj paylaşım platformu olan Twitter'dan, Twitter'ın sağladığı programcı arayüzü (API) kullanılarak indirildi.

Aynı zamanda, münazara esnasında konuşulanların neler olduğu hangi dakika ve saniyede ne konuşulduğu yazılı olarak [4] ten indirildi. Bu adreste NYTimes web sitesi, münazaranın hangi saniyesinde nelerin söylendiği bilgisini saklıyor. Bunu kullanıcılarına etkileşimli münazara videosu sunmak için yapıyor. Web sitesinde, herhangi bir kullanıcı münazaranın yazılı parçalarından herhangi birine tıklayıp, münazara video oynatıcısının tıklanılan yerden çalmaya devam etmesini sağlayabiliyor. Bu da açıkça olmasa bile bu web sayfasının kodları içinde münazarada saniye saniye nelerin söylendiği bilgisini bulduruyor. Bu web sayfasından yazılan özel bir parçalayıcı/çözümleyici (parser) ile bu bilgi çıkarılmıştır.

Bu bildirinin ikinci bölümü toplanan veri hakkında genel bilgileri, 3. bölümü münazaranın parçalarını ve neler konuşulduğunu, bunların Twitter üzerinde nasıl etkisinin olduğunu göstermektedir. 4. bölümde ise sonuçlar ve sonraki araştırma çalışmaları hakkında bilgiler yer almaktadır.

2. Toplanan veri hakkında

İlgili münazara verisi iki ayrı veri olarak toplandı. Birisi Twitter üzerindeki konuşulanların verisi, diğeri ise gerçekte münazarada konuşulanların verisi.

2.1 Twitter üzerinde konuşulanlar

Twitter üzerinde nelerin konuşulduğunu indirmek için Twitter Streaming API [1,2] kullanıldı. Bir programcı arayüzü olan bu yapıdan belirlenen bir kritere göre tweetler(kısa mesajlar) indirilebiliyor. Twitter ilgili kısa mesajların tamamını vermeyi garanti etmiyor. Fakat belirtilene göre genellemeye yeterli miktarda veriyor. Mümkün olduğunca farklı Twitter kullanıcısının tweetlerini vermeye çalışıyor. Twitter'dan 17 Ekim 2012 01:00:00 (GMT) zamanından başlayarak 17 Ekim 2012 02:34:59 (GMT) zamanına kadar, 90 dakikalık münazara için 95 dakikalık tweetler indirildi. İndirme işlemi, "obama", "romney", "biden", "ryan", "joe", "paul", "republican", "democrat", ve "VPDebate" kelimelerini içeren

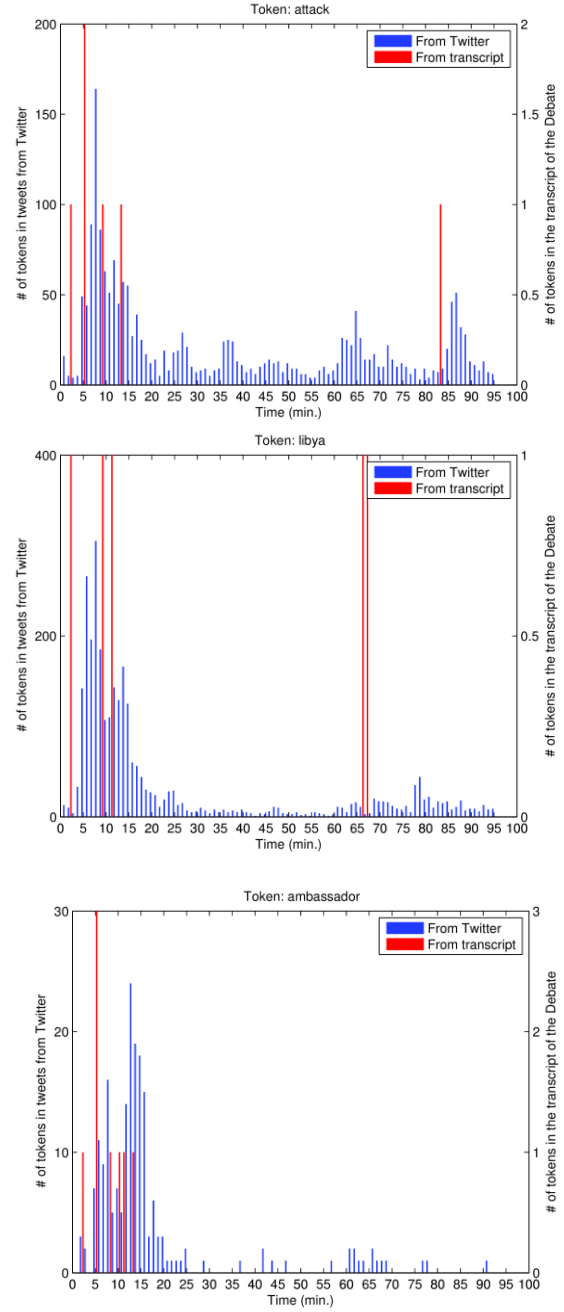
tweetler için yapıldı. Bu kelimeler Twitter programcı arayüzüne indirilmek istenilen tweetlerin içermesi gereken kelimeler bunlardır denilerek parametre olarak veriliyor. Sonuç olarak elde edilen tweet kümesinin özellikleri şu şekildedir: 285.003 tweet toplanmıştır. Bunlar 190.345 farklı kullanıcı tarafından yazılmıştır. Bu tweetlerin kısa mesaj kısımları parçalara ayrılmıştır. Her bir parça noktalamadan arındırılmış kelimeler, web kaynaklarına bağlantılar(linkler), hashtagler¹, mentionlar² veya noktalama işaretlerinin kendileri olabilir. Toplam 2.256.465 parça çıkarılmıştır. Bunlardan 118.788 tanesi biriciktir. Bu parçalar, anlam ifade etmeyen veya İngilizce dili içinde çok fazla kullanılan kelimeleri içermemektedir. Bunu yapabilmek için, [3,5] de bulunan liste stopword(yasaklı kelimeler) listesi kullanılmıştır. Bu listedeki kelimeler bu analizde elde edilen parçalardan çıkarılmıştır. Çıkarılan bütün parçalar, konuşulduğu an ile ilişkili bir şekilde tutulmaktadır.

2.1 Münazarada konuşulanlar

The Newyor Times websitesi, münazaranın etkileşimli bir yazısını[4] içermektedir. Hangi saniye kim ne söylemiş, hepsi bu yazı içerisinde vardır. Tweetler için uygulanan işlemlerin aynısı münazaranın yazısı için de uygulandı. Hangi zamanda ne konuşulduğu bilindiği için münazaradan elde edilen parçalar da konuşulduğu zamanla ilişkili olarak tutuldu. Münazaradaki ilk kelimenin 17 Ekim 2012 01:00:00 (GMT) zamanında söylendiği varsayılmıştır.

3. Parçaların kullanım sıklığına göre değerlendirmeler

Konuşulanlar kelime kelime parçalara ayrılıp konuşulduğu zamanla ilişkili olacak şekilde saklanmıştır. Bu kelimeler ve zamanlar, daha sonra dakika bazına çevrilmiştir. Bu şekilde hangi kelimenin hangi dakika ne kadar konuşulduğu bilgisi tutulmuş oldu. Bu veriyi ayırıştırma yöntemi, her iki veri kümesi(hem Twitter'dan gelen, hem de konuşulanların yazısından gelen) için de yapıldı. Bundan sonra görülecek çizimlerde, sol y-ekseni Twitter'dan gelen, sağ y-ekseni ise münazara metninden gelen kullanımların miktarını göstermektedir. x-ekseni ise dakikayı göstermektedir. Kırmızı renkli çizgiler münazaradaki konuşma sıklığını, mavi renkli çizgiler ise Twitter'daki konuşma sıklığını göstermektedir.

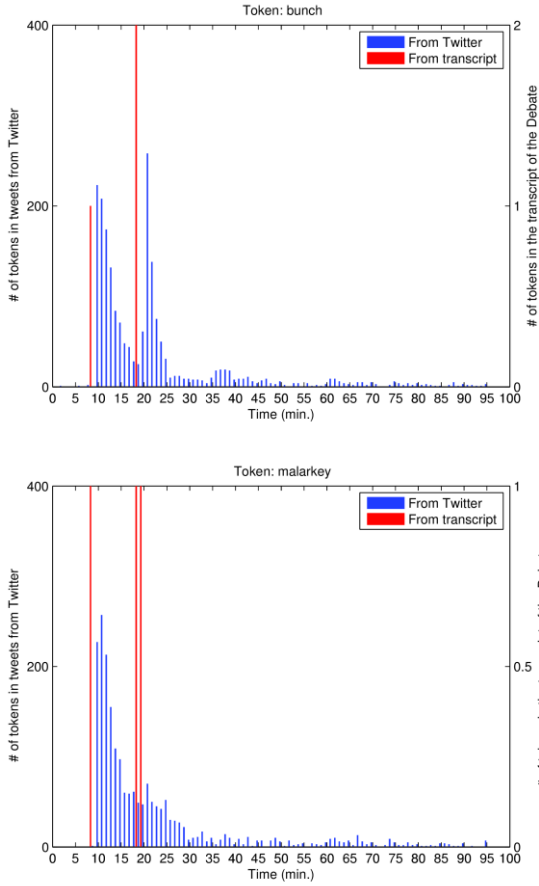


Şekil.1 "attack", "libya" ve "ambassador" kelimelerinin dakikada kullanım sıklıkları

¹ Hashtagler kullanıcılar tarafından kısa mesajlarda kullanılan ve üzerinde önem atfedilen kelime veya harf/sayı dizileridir. Boşluk içermezler. Birden fazla tweet, eğer aynı hashtagi kullanıyorsa, birbiri ile ilgilidir bilgisini verirler. Hashtagler # ile başlar.

² Mentionlar Twitter kullanıcıları tarafından başka Twitter kullanıcılarını belirtmek amacıyla kullanılan özel yapılardır. Mentionlar @ ile başlar.

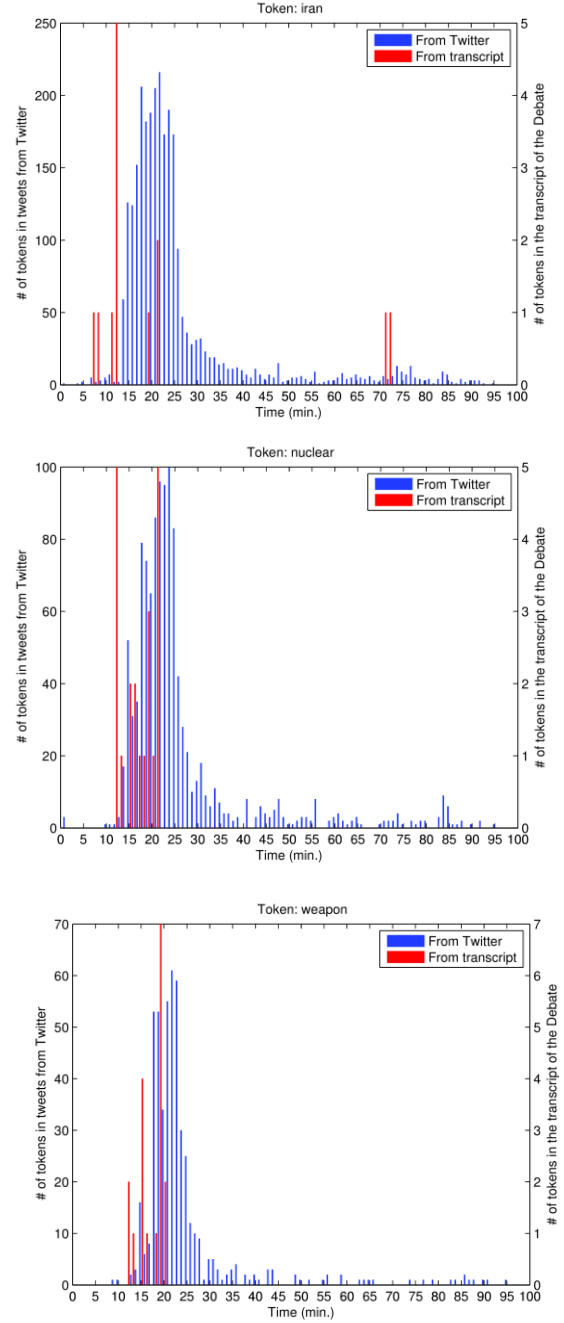
Münazarada konuşulan ilk konu, Libya'da bir terörist saldırıda öldürülen ABD büyükelçisi hakkında idi. Şekil.1 de "attack", "libya" ve "ambassador" kelimelerinin kullanım sıklıkları verilmiştir. Buradan, münazaranın başında hem münazaranın metninde, hem de Twitter'da kullanımın yoğun olduğu görülmektedir. Bu konuşmalar esnasında Joe Biden, "bunch of malarkey" tabirini rakibi Paul Ryan'ın söylediklerine karşılık olarak söyledi. Bu sosyal medyadan büyük tepki aldı. Şekil.2 de bu etki görülmektedir.



Şekil.2 “bunch” ve “malarkey” kelimelerinin dakikada kullanım sıklıkları

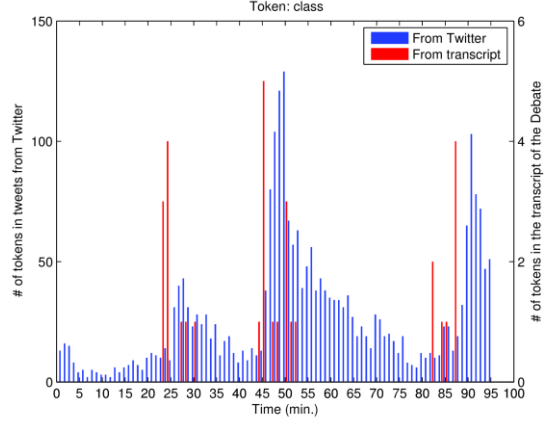
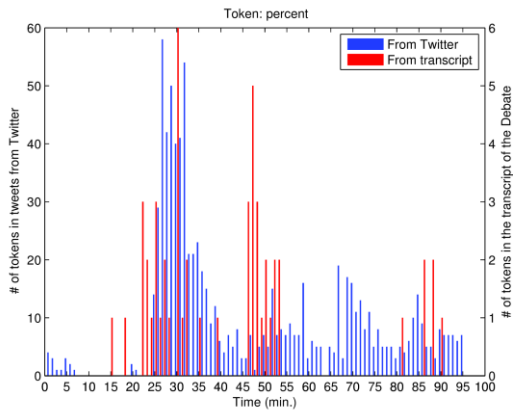
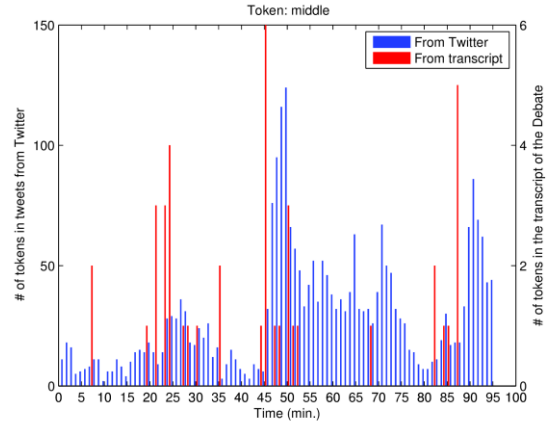
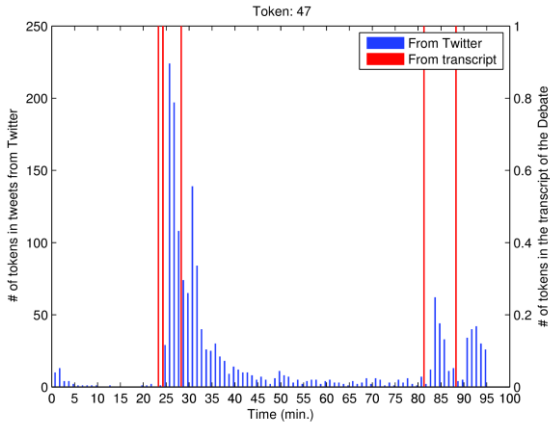
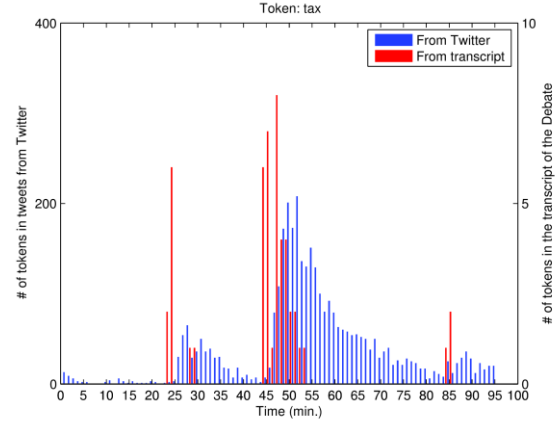
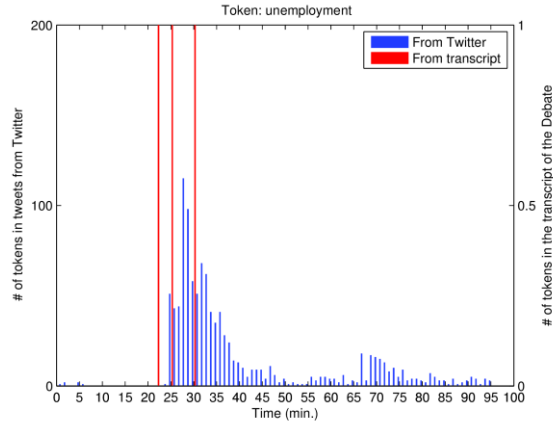
Bir sonraki konuşulan konu İran'ın nükleer silahları idi. Şekil.3'ten de görüldüğü gibi hem münazara metninde, hem de Twitter kullanımında "iran", "nuclear", ve "weapon" kelimeleri birlikte ve diğer zamanlara göre daha fazla kullanılıyor.

Bir sonraki konu "işsizlik" ve "%47" olayı. İşsizlikle ilgili konuşulurken, Joe Biden, karşı Başkan adayı Mitt Romney'nin İnternet'e yansıyan bir kaydını hatırlattı. Bu kayıttta Mitt Romney'nin, "bu ülkenin %47'si benim hedef kitlem değil" dediği iddia edilmişti. Bütün bu konuşulanlar o zaman aralığında Twitter'a yansıdı. Şekil.4 te “unemployment”, “47” ve “percent” kullanımı görülmektedir. “percent” (yüzde) kullanımı



başka konulardaki yüzdeleri ifade etmek için de kullanıldığı için daha yaygın görülmektedir.

Şekil.3 "iran", "nuclear" ve "weapon" kelimelerinin dakikada kullanım sıklıkları

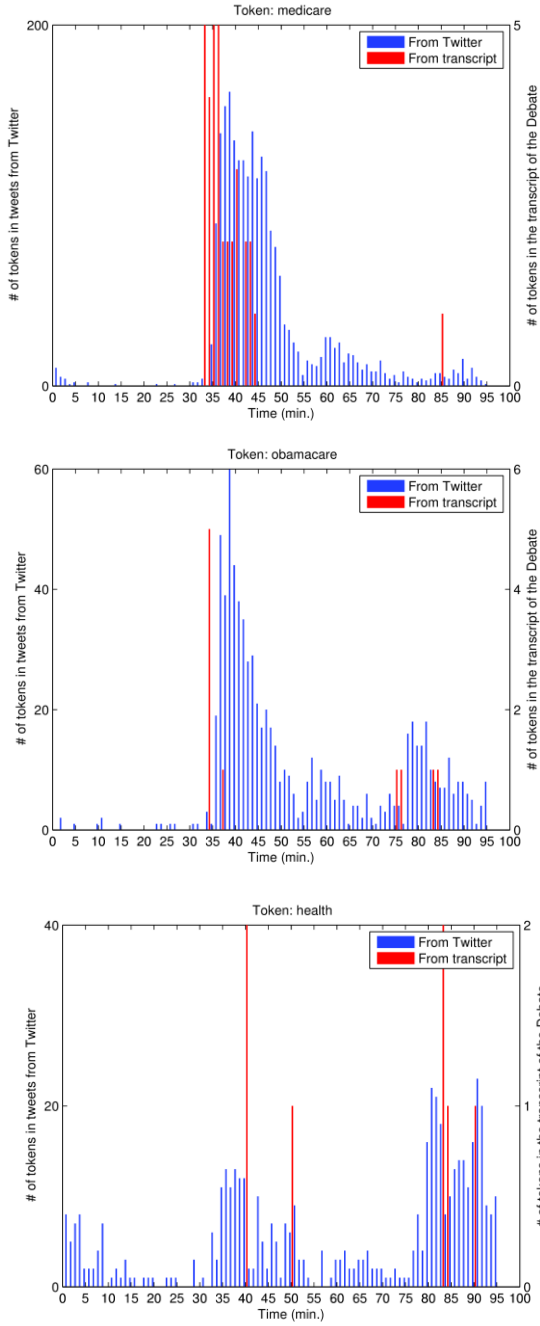


Şekil.4 "unemployment", "47" "percent" kelimelerinin dakikada kullanım sıklıkları

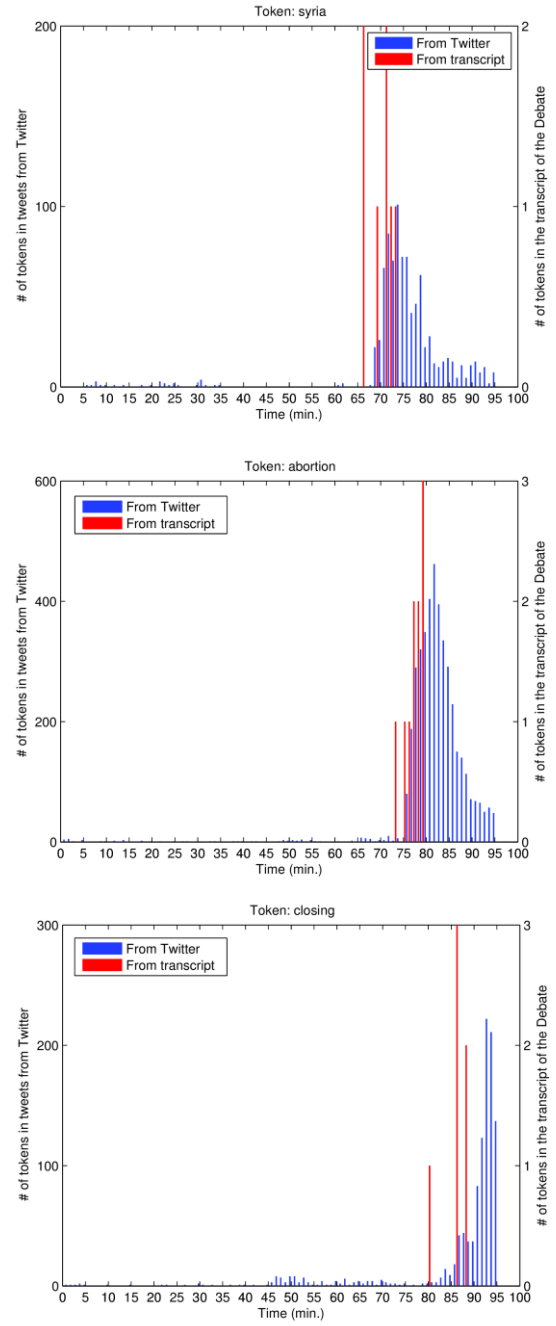
İşsizlik ve %47 meselesi konuşulurken, açılan bir diğer konu ise "vergi" ve "orta sınıf" tı. Şekil.5'te "tax", "middle" ve "class" kelimelerinin kullanımları gösterilmektedir. Bu konulara, yaklaşık 45. dakikadan sonra tekrardan dönmüştür

Şekil.5 "tax", "middle" ve "class" kelimelerinin dakikada kullanım sıklıkları

Bir sonraki konu Obama yönetiminde yapılan sağlık reformu ve sağlık sigortası idi. Şekil.6 da "medicare", "obamacare" ve "health" kelimelerinin kullanımı görülmektedir. En son konuşulan konulardan birisi Suriye'nin durumu idi. Ardından münazara moderatörü, başkan yardımcısı adaylarının küretaj konusundaki fikirlerini sordu. Ve son olarak da kapanış konuşmalarını yapmasını istedi.



Şekil.6 "medicare", "obamacare" ve "health" kelimelerinin dakikada kullanım sıklıkları



Şekil 7. "syria", "abortion" ve "closing" kelimelerinin dakikada kullanım sıklıkları

Bununla ilgili olarak Şekil.7 de, "syria", "abortion" ve "closing" kelimelerinin kullanımı görülmektedir.

4. Sonuçlar ve sonraki araştırma çalışmaları

Bu çalışmada, ABD seçimlerinden önce yapılan münazaralardan başkan yardımcılığı

münazarası ile ilgili bir kısa mesaj paylaşım platformu olan Twitter'daki kelime kullanım sıklıkları ve bunların zamana göre yayılımları verilmiştir. Gerçekteki münazara metni ile Twitter'dan çıkartılan veri karşılaştırılmıştır. Görülmüştür ki, gerçek münazara metni ile Twitter'daki kullanımlar arasında bir benzerlik bulunmaktadır. En önemli fark, kullanım sıklıklarının ölçekleridir. Münazara metni 3 kişi (2 karşı karşıya gelen kişi ile moderatör) arasında geçmekte iken Twitter verisi yaklaşık 200 bin kişi tarafından oluşturulmuştur.

Bundan sonraki araştırmalar ilk, bu konuların başlangıçlarının bulunması, ikinci olarak konuların konuşulmasının bitişinin algılanması üzerine olacaktır. Sonraki aşamada bulunan zaman dilimleri arasında kelime-kelime birlikte kullanım grafi oluşturularak bu graflar incelenecek ve bu graflar kullanılarak, aktif konu başlıklarının zamanla değişiminin algılanabilip algılanamadığı araştırılacaktır.

Hali hazırda basit bir fonksiyon ile konuların konuşulmaya başlandığı anların yakalanabileceği düşünülmektedir. Bu fonksiyon aşağıdaki gibidir:

$$konuB(t) = \begin{cases} 1 & , A(t) > \alpha A(t-1) \wedge A(t) > \beta \wedge konuB(t-1) \neq 1 \\ 0 & , \text{diğer durumlarda} \end{cases}$$

A dizisi, bu yapıya girilen zaman serisidir. konuB(t) t anında ilgili kelimenin konuşulmaya başlanıp başlanmadığı bilgisini içermektedir. a ve B, fonksiyonun diğer iki parametresidir. a ne kadar büyük olursa, zaman serisinde o kadar dik çıkışlar konu başlangıcı olarak algılanır. B ise belli bir değerden altının hesaba katılmamasını sağlamak için konulmuştur. Bu değerler hesaplama esnasında datanın durumuna göre ayarlanabilir. A dizisinin sıfır olan elemanlarına küçük önemsenmeyecek değerler atanmalıdır.

5. Kaynaklar

[1] Twitter, Twitter api wiki/twitter api documentation. <http://apiwiki.twitter.com/w/page/22554679/> Twitter-API-Documentation, Erişim: 07.05.2011.

[2] Twitter, The streaming apis. <https://dev.twitter.com/docs/streaming-apis>, Erişim: 06.12.2012.

[3] G. Salton, The SMART Retrieval System Experiments in Automatic Document Processing. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1971.

[4] NYTimes.com, Video of the Biden-Ryan vice-presidential debate - interactive- feature., <http://www.nytimes.com/interactive/2012/10/11/us/politics/20121011-vice-presidential-debate-biden-ryan.html>, Erişim: 06.12.2012.

[5] S. W. List, A stop words list for english. <ftp://ftp.cs.cornell.edu/pub/smart/> , Erişim: Mayıs 2010.