Diyalog Tabanlı Metinlerde Konu Değişimi Tespiti

Topic Change Detection on Dialog Based Text

Lütfi Kerem Şenel^{1,2,3}, Veysel Yücesoy¹, Aykut Koç¹, Tolga Çukur^{2,3,4}

¹ASELSAN Araştırma Merkezi, Ankara, Turkey

²Elektrik ve Elektronik Mühendisliği Bölümü, Bilkent Üniversitesi, Ankara, Türkiye

³Sabuncu Beyin Araştırmaları Merkezi, UMRAM, Bilkent Üniversitesi, Ankara, Türkiye

⁴Sinirbilimi Programı, Bilkent Üniversitesi, Ankara, Türkiye

Email: {lksenel,vyucesoy,aykutkoc}@aselsan.com.tr, cukur@ee.bilkent.edu.tr

Özetçe —Son dönemde katlanarak gelişen haberleşme yöntemleri (internet, sosyal medya, akıllı telefon, vb.) sayesinde veriye ulaşmak ve paylaşmak kolaylaşmıştır. Özellikle son yıllarda sözlü ve yazılı paylaşım mecraları hızlı gelişim göstermiştir. Yazılı paylaşımın en hızlı yaşandığı alanlar arasında sosyal medya siteleri ve forumlar öne çıkmaktadır. Forumlarda sosyal medyadan farklı olarak, her başlık altında sadece o başlık ile ilgili konuşmalar yapılması beklenmektedir. Konu kısıtlılığı olan ve sözlü iletişimin son yıllarda en hızlı geliştiği alanlardan biri de çağrı merkezleridir. Belirli konuların dışına çıkılması ya da ana konunun değiştirilmesinin otomatik tespiti özellikle çağrı merkezleri ve teknik forumlar gibi mecraların iletişim performansının değerlendirilmesi ve otomatik olarak yönetilebilmesi açısından önemlidir. Bu çalışma ile diyalog tabanlı Türkçe metinler içerisinde konu değişimini otomatik olarak algılayabilen sınıflandırıcılar geliştirilmiştir. Bu sınıflandırıcıların geliştirilebilmesi için öncelikle Türkçe forumlardan konu tabanlı karşılıklı konuşma verileri tasnif edilerek ham bir veri kümesi elde edilmiştir. Oluşturulan veri kümesi üzerinde klasik bir yöntem (TF-IDF) ile bir derin öğrenme modeli (LSTM) otomatik konu değişimi tespiti problemi için karşılaştırılmıştır. Klasik yöntem ile test kümesinde %80'lere varan başarı elde edilirken, derin öğrenme yönteminin performansının %76 seviyesinde kaldığı gözlenmiştir.

Anahtar Kelimeler—Konu Değişimi, Konu Takibi, Diyalog Veri Kümesi, Doğal Dil İşleme

Abstract—It has become easier to reach and share data thanks to the communication methods (internet, social media, smartphone, etc.) which have been advancing in the last years. Especially in recent years, oral and written sharing channels have developed rapidly. Social media sites and forums are among the areas where written sharing is the fastest. Unlike social media, in forums it is expected that people will discuss specific issues under corresponding topics. Call centers are also among rapidly developing oral communication channels in the recent years where the scope of the conversation is restricted. The automatic determination of digressing from certain subjects or changing the main subject is especially important for the evaluation of the communication performance and automatic management of the media such as call centers and technical forums. With this study, classifiers that can automatically detect the subject change within the dialogue-based Turkish texts have been developed. In order to develop these classifiers, first of all, the subject-based conversation data from Turkish forums were compiled and a raw data set was obtained. In this study, a classical method (TF-IDF) and a deep learning model (LSTMs) have been applied to the generated dataset for topic change detection task. Results show that 80% accuracy can be achieved by the classical method while deep learning model achieves 76%.

Keywords—Topic Change, Topic Tracking, Dialog Dataset, Natural Language Processing

I. GİRİŞ

Sosyal medya platformları insanların serbest formda paylaşım yaptıkları ve metin tabanlı haberleştikleri ortamlardır. Son dönemde hızla gelişen internet erişimi ve akıllı cihaz teknolojileri ile bu tip platformlara karşı olan ilgi de hızla artmıştır. Sosyal medya platformlarından hem işleyiş hem de tarihsel olarak farklı bir gelişim gösteren bir diğer platform da forumlardır. Konuya özel farklı sayfalara erişim imkanı veren forum yapıları karşılıklı konuşma benzeri bir yapıda iletişim imkanı sunmaktadır. Belirli bir konuda sorulan bir soru ya da yapılan bir yorum ile başlayan iletişim, her kullanıcının önceki girilerle ilgili yaptığı yorumlar ile devam eder. Forumlarda en çok kaçınılmaya çalışılan olay sayfa konusunun değişimidir çünkü sayfa içerisinde değişen konu hem forumun kullanılabilirliğini azaltacak hem de kullanıcıların aradığını bulmasını zorlaştıracaktır. Forum sayfalarındaki konu dışı girdilerin tespiti ve sayfadan kaldırılması güncel olarak kullanıcı şikayeti, yönetici tespiti ve eylemi ile gerçekleştirilmektedir. Özellikle yazılım dünyasının vazgeçilmezlerinden olan teknik forumlar kullanıcı deneyimini iyileştirebilmek için bu konuda en katı kuralları uygulayan forumlardır.

Yazılı iletişim dışında performans endeksine önem veren sözlü uygulamaların başında çağrı merkezleri uygulamaları gelmektedir. Çağrı merkezi aramalarında hem kullanıcı deneyimini iyileştirmek hem de çağrının hedefinden sapmasını önlemek amaçları ile son dönemde konu bazlı aramalar yapılmaktadır. Sadece belirli bir bilgilendirme ya da işlem için aranan kullanıcı istese dahi farklı bir işleme geçmesi engellenmekte, bunun için yeni bir çağrı oluşturmaya yönlendirilmektedir. Çağrı merkezi sistemlerinin performans analizlerinde bu kriterlere ne derece uyulabildiğinin tespiti forumlara benzer şekilde ancak şikayet ya da amir tespiti ile gerçekleştirilmektedir. Çağrı içerisinde konu değişiminin tespiti ile saptanabilecek bir diğer unsur da kötü niyetli kullanıcıların çağrı merkezini boş yere meşgul etmesidir.

Doğal dil işleme literatüründe diyalog tabanlı metinler üzerine yapılmış çok çeşitli çalışmalar mevcuttur. Diyalog durumu takibi [1], [2], diyalog üretimi [3] veya sohbet robotu geliştirilmesi çalışmaları [4], [5] bunlara örnek olarak verilebilir. Literatürde metinlerde konu değişimi tespitine yönelik olarak

TABLO I: Veri kümesi içeriği

Konu	Başlık Sayısı	Girdi Sayısı
Basketbol	159	52490
Belgesel-TV	122	129173
Bilim-Teknoloji	251	21229
Evciil Hayvanlar	351	9840
Formula 1	280	21935
Futbol	2786	724770
Kitap	333	38187
Müzik	262	81111
Sinema	837	121156
Yemek-Mutfak	281	12607
Toplam	5662	1212498
Android Oyun ve Uygulama	283	28743
Araba Markaları	486	177802
Beyaz Eşya	425	78774
Laptop	142	5592
Toplam	1336	290911

yapılan çalışmalar genellikle konu tespiti ve takibi (TDT) çalışmaları olarak bilinir [6]. Ancak bu çalışmalar çoğunlukla İngilizce için yapılmış olup, haberler yayınlarındaki hikayeler [7] veya dökümanlar [8] gibi nispeten uzun ve karşılıklı diyalog içermeyen metinlere odaklanmıştır. Literatürdeki konu tespiti ve takibi çalışmalarında, kelime frekanslarına dayalı TF-IDF [9] benzeri yöntemler ile metinleri içerdikleri konuların olasılıksal karışımları şeklinde modelleyen Latent Dirichlet Allocation (LDA) [10] tabanlı yöntemler sıklıkla kullanılmaktadır. Bildiğimiz kadarıyla Türkçe dili için diyalog tabanlı metinlerde konu değişimi tespitine yönelik yapılmış bir çalışma mevcut değildir.

Çalışmanın geri kalanı özetle şu şekildedir: İkinci kısımda otomatik konu değişimi tespiti için öğrenme tabanlı geliştirme yapabilmek adına hazırlanmış olan veri kümesi hakkında detaylara yer verilmiştir. Yine aynı kısımda, konu değişimini tespit edebilmek amacıyla kullanılan yöntemler anlatılmıştır. Üçüncü kısımda ise önerilen yöntemler kullanılarak elde edilen sonuçlar sunulmuş ve dördüncü kısımda bu sonuçlar üzerine bir tartışma yapılarak çalışma sonlandırılmıştır.

II. YÖNTEMLER

A. Veri Kümesi

Bildiğimiz kadarıyla metine dökülmüş ve etiketlenmiş karşılıklı Türkçe diyalog veri kümesi bulunmamaktadır. Derin öğrenme algoritmaları için yeterli büyüklükte ve konularına göre etiketli gerçek karşılıklı konuşma verisi oluşturmak önemli miktarda insan eforu ve zaman gerektirmektedir. Bu nedenle bu çalışmada gerçek karşılıklı konuşma verisine benzetmek amacıyla internet üzerindeki 'forum' tipindeki sitelerin kullanılmasına karar verilmiştir. Genel olarak bu sitelerde kullanıcılar farklı konuların (bazen alt konuların) altında başlıklar açarak, başlığa veya başlığın altındaki bir başka girdiye cevap verecek şekilde girdiler girerler. Bir başka deyişle, bu sitelerdeki her bir başlık, belirli bir konuda çok katılımcılı bir diyaloga karşılık gelmektedir. Bu nedenle bu tip internet sitelerindeki içeriğin gerçek diyalog verisini simüle etmede kullanılabileceği öngörülmüştür.

İnternette çok sayıda Türkçe forum sitesi mevcuttur. Bu sitelerde en çok kullanılan *Donanım Haber*¹ forum sitesi bu çalışmada diyalog veri kümesini oluşturmak için kaynak olarak kullanılmıştır. Donanım Haber sitesindeki başlıkları ve içerindeki girdileri indirmek amacıyla Python dilinde bir program yazılmıştır. Yazılan program ile seçilen 14 farklı ana konu içerisindeki en az 10 girdiye sahip başlıkların içerisindeki 2000 karakterden uzun olmayan girdiler indirilmiştir. Ana konular arasında içerik miktarı olarak çok fazla dengesizlik oluşturmak adına *Araba Markaları* gibi çok fazla girdiye sahip konulardaki içeriğin tamamı yerine bir kısmı indirilmiştir.

Ön işleme adımı olarak indirilen girdilerdeki büyük harfler küçük harfe çevrilip URL'ler, alfanümerik olmayan bütün karakterler, gereksiz kelimeler (stop words), tek başına bulunan harfler ve ekler kaldırılmıştır. Bu ön işleme adımlarından sonra uzunluğu 5 kelimeden altında kalan girdiler veri kümesinden kaldırılmıştır. İndirilen girdilerden bir örneğin ön işlemeden önceki ve sonraki halleri aşağıda verilmiştir:

Ö: *bilmeyenler için <http://www.last.fm/> bence dünyanın en iyi radyosu hertürlü tarz grup mevcut bence üye olun türk grupları bile var pentagram felan çalıyo*

S: *bilmeyenler bence dünyanın radyosu hertürlü tarz grup mevcut bence üye olun türk grupları pentagram felan çalıyo*

İndirilen 14 ana konudan 4 tanesi, modelin daha önce görmediği yeni konulardaki performansının değerlendirilmesi için veri kümesinden ayrılmıştır. Ön işlemeden sonra elde edilen veri kümesinin detayları Tablo I'de verilmiştir. Tablo I'nin üst kısmında verilen 10 konudan her birinin içerisindeki başlıkların %60'ı eğitim, %20'si geliştirme ve %20'si test veri kümelerini oluşturmak üzere ayrılmıştır.

B. Konu Değişimi Tespiti

Konu değişimi tespiti problemi, bir diyalog içerisindeki karşılıklı n adet konuşmaya bakarak son konuşmanın konuyu değiştirip değiştirmediğini anlamaya dayalı bir ikili sınıflandırma problemi olarak formalize edilebilir. Bu prensip kullanılarak geliştirilen bir sınıflandırıcı, herhangi bir uzunluktaki gerçek diyalog verisi üzerinde kaydırılarak konunun değiştiği yerleri tespit etmek için kullanılabilecektir.

Bu çalışma prensibine sahip bir modelin öğrenme teknikleri ile eğitilebilmesi için, öncelikle veri kümesinin uygun bir formata getirilmesi gerekmektedir. Eğitim ve değerlendirme için uygun bir veri kümesi oluşturmak amacıyla, önce 10 ana konudan biri, seçilme ihtimali konunun veri kümesi içerisindeki boyutunun (byte cinsinden) logaritması ile orantılı olacak şekilde, rastgele seçilmiştir. Sonra seçilen konunun içerisinde, boyutu ile orantılı şekilde rastgele bir başlık seçilmiş ve seçilen bir başlıktan ardışık n ($n = 3, 5$) tane girdi seçilip arka arkaya eklenerek (her bir girdinin sonuna "<EOS>" ifadesi eklenmiştir) bir örnek oluşturulmuştur. Dengeli bir veri kümesi oluşturmak için, oluşturulan örnekteki son girdi %50 ihtimalle başka bir konudan rastgele seçilen bir girdi ile değiştirilmiştir. Bu işlem eğitim kümesi için 100.000, geliştirme ve test kümeleri için 10.000 kere tekrarlanmıştır. Aynı işlemler ayrılan 4 ana konu için diğer 10 ana konudan bağımsız bir şekilde tekrarlanmıştır.

¹<https://forum.donanimhaber.com>

C. TF-IDF

Terim frekansı - ters metin frekansı (term frequency - inverse document frequency, TF-IDF), bir kelimenin döküman kümesi içerisindeki bir döküman için ne kadar önemli olduğunu ifade eden sayısal bir istatistiktir. Bir döküman içerisinde seçilen bir kelime için TF-IDF değeri, kelimenin döküman içerisindeki frekansı (tf) ile kelimenin ters metin frekansının (idf) çarpılması ile elde edilir. Bu değerlerin hesaplanması için literatürde farklı ağırlıklandırma yöntemleri mevcuttur. Bu çalışmada kullanılan ağırlıklandırma yöntemi Denklem (1)'de verilmiştir.

$$\begin{aligned} \text{TF-IDF}(t, d, D) &= \text{TF}(t, d) \cdot \text{IDF}(t, D) \\ &= \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}} \cdot \log\left(\frac{N}{1 + n_t}\right) \end{aligned} \quad (1)$$

Bu denklemde t odak kelimeyi, d odak dökümanı, D ise bütün dökümanların oluşturduğu kümeyi ifade etmektedir. $f_{t,d}$ d dökümanında t kelimesinin geçme sayısına, N toplam döküman sayısına, n_t ise t kelimesini içeren döküman sayısına karşılık gelmektedir.

TF-IDF bir kelimenin, bulunduğu dökümana ne derece özgü olduğunu hesaplamaktadır. Ancak bu çalışmada, bu yöntem farklı dökümanlar yerine farklı konulardaki diyaloglardan oluşan bir veri kümesinde, diyalog içerisindeki bir kelimenin, konular için ne derece önemli olduğunu hesaplamakta kullanılmıştır. Bir başka ifadeyle, Denklem (1)'deki d gerçek bir döküman yerine bir konu içerisindeki bütün girdilerin birleşimini, D ise bütün konuların oluşturduğu kümeyi ifade etmektedir. Eğitim kümesindeki her bir kelime için hesaplanan TF-IDF değerleri test kümesi kullanılarak oluşturulan 10.000 örnekten her birindeki ilk $n-1$ girdi ile son girdinin aynı konu ile alakalı olup olmadığını tahmin etmekte kullanılmıştır. Test kümesindeki bir örneği oluşturan n adet girdi içerisinde, ilk $n-1$ girdinin ve son girdinin konuları Denklem (2) kullanılarak tahmin edilmiştir.

$$\begin{aligned} \hat{d}_1 &= \arg \max_{d \in D} \sum_{t \in S_1} \text{TF-IDF}(t, d, D) \\ \hat{d}_2 &= \arg \max_{d \in D} \sum_{t \in S_2} \text{TF-IDF}(t, d, D) \end{aligned} \quad (2)$$

Burada S_1 test kümesindeki bir örnekteki ilk $n-1$ girdinin tamamına, S_2 ise örnekteki son girdiye karşılık gelmektedir. Eğer $\hat{d}_1$ ile $\hat{d}_2$ farklı ise konunun değiştiği sonucuna varılmıştır.

D. LSTM

Diyalog metinlerinde konu değişimini tespit edebilmek amacıyla Facebook tarafından geliştirilen ve desteklenen Pytorch derin öğrenme kütüphanesi incelenmiş ve bu kütüphane kullanılarak uzun-kısa süreli bellek (long-short term memory, LSTM) temelli bir derin öğrenme modeli geliştirilmiştir. Geliştirilen model, her kelime için 300 boyutlu kelime vektörü öğrenen bir temsil katmanından, 5 adet 300 gizli birim içeren LSTM katmanından ve son LSTM katmanının 300 boyutlu çıktısını bir tane sayıya çeviren tam bağlı (fully connected, FC) katmandan oluşmaktadır. Tam bağlı katmandan çıkan değer sigmoid fonksiyonu ile olasılığa çevrilmiş ve bu olasılık konu değişimi olup olmadığına karar vermek için kullanılmıştır. Eğitim sırasında eğitim kümesindeki örnekler modele 100'lü gruplar halinde verilmiştir. Modelin genelleme performansını

Şekil 1: Eğitim için kullanılan LSTM tabanlı model

arttırmak için her bir grubun eğitimi sırasında model parametrelerinin yarısı rastgele olarak atılmıştır (0.5 olasılık dropout). Kullanılan model Şekil 1'de gösterilmiştir.

Modelin ilk katmanı olan kelime temsil katmanını rastgele ağırlıklarla başlatmak yerine, önceden büyük ve etkisiz bir derlemden gözetimsiz olarak öğrenilmiş olan kelime temsillerini [11]–[13] kullanmanın bir çok doğal dil işleme uygulamasında modelin öğrenmesini kolaylaştırdığı ve performansını arttırdığı gösterilmiştir. Burada, kelime temsil katmanını başlatmada kullanılan bu vektörler eğitim sırasında model tarafından güncellenebileceği gibi eğitim süresince değişmeden sabit olarak da kullanılabilir. Bu çalışmada önceden eğitilmiş kelime vektörlerinin modelin performansını nasıl etkilediğini incelemek için rastgele başlatma, Türkçe için Wikipedia ve Common Crawl derlemlerinden öğrenilmiş olan fastText kelime temsilleri [14] ile başlatıp eğitime, ve eğitim boyunca bu vektörleri sabit tutma yöntemleri denenmiştir. Modellerin eğitimi sırasında eğitim örnekleri üzerinden 30 iterasyon yapılmış, devamında geliştirme kümesinin maliyet fonksiyonunun değeri azalmayı bırakana kadar eğitime devam edilmiştir.

III. SONUÇLAR

TF-IDF ve LSTM yöntemleri konu değişimi tespiti için $n = 3$ ve $n = 5$ seçilerek denenmiştir. Tablo II incelenen yöntemlerin 10 konudan oluşan veri kümesinden oluşturulan test kümesindeki konu değişimi tespiti için doğruluk, kesinlik ve geri çağırma performanslarını göstermektedir. Bu tabloda LSTM FT fastText kelime vektörleri ile başlatılıp eğitim sırasında vektörlerin güncellendiği modeli, LSTM FTS ise fastText ile başlatılıp vektörlerin sabit tutulduğu modeli ifade etmektedir.

Tablo II'teki sonuçlara göre TF-IDF konu değişimini tespit etmekte LSTM tabanlı derin öğrenme yöntemlerinden daha iyi performans göstermektedir. TF-IDF yöntemi geçmişe dönük daha az sayıda girdi ($n = 3$) kullandığında daha iyi performans gösterirken LSTM daha fazla sayıda girdiyi ($n = 5$) kullandığında konu değişimini daha başarılı tespit etmektedir. LSTM tabanlı modelde kelime temsili katmanını önceden eğitilmiş vektörler ile başlatıp eğitim sırasında bu vektörleri güncelleme rastgele başlatmaya göre bir miktar iyileştirmeye sağladığı görülmektedir. Tablo II'teki dikkat çekici

TABLO II: İncelenen modellerin 10 konuluk veri kümesindeki test performansları

		Doğruluk (%)	Kesinlik (%)	Geri Çağırma (%)
n = 3	TF-IDF	80.2	73.2	94.0
	LSTM	69.6	68.5	70.6
	LSTM FT	72.9	72.3	72.8
	LSTM FTS	70.2	69.3	70.6
n = 5	TF-IDF	77.4	70.4	94.4
	LSTM	76.1	78.2	71.3
	LSTM FT	76.7	78.8	72.2
	LSTM FTS	76.3	78.8	71.2

bir diğer nokta ise TF-IDF yönteminin %94'e ulaşan geri çağırma performansıdır. Bu sonuç, yöntemin konu değişimini kaçırmamanın hatalı tespit yapmaktan daha önemli olduğu uygulamalarda kullanıma potansiyelinin olduğunu göstermektedir. Hatalı tespit yapmamanın kritik olduğu uygulamalarda ise kesinlik değerleri önem kazanmaktadır. Sonuçlar %79'a yaklaşan bir performansla LSTM modellerinin TF-IDF'ten daha yüksek kesinlik sağladığını göstermektedir. Modelin konu değişimi tespiti için kullandığı karar eşiği (0.5) kesinlik ile geri çağırma arasında bir ödünleşim sağlamaktadır, bu eşik yükseltilecek daha yüksek kesinlik performansı, geri çağırma performansından ödün verilerek, elde edilebilir.

İncelenen yöntemlerin daha önceden görmedikleri yeni konuların değişimlerini tespit etmede nasıl bir başarı göstereceklerini test etmek amacıyla, 4 farklı konu içeren ikinci veri kümesi kullanılarak oluşturulan test örneklerindeki performansları değerlendirilmiştir ve sonuçlar Tablo III'te gösterilmektedir. Tablo III'teki sonuçlar değişimi tespit edilecek olan konuların yöntemler tarafından daha önce görülmemiş olmasının tespit performansını önemli derecede düşürdüğünü göstermiştir. TF-IDF yönteminin tespit doğruluğunun %80'den %60'a düşmüş olmasına karşın, bu yöntem konu değişimlerinin %87'ye yakınına tespit etmeyi başararak etkileyici bir geri çağırma performansı göstermiştir. Bu tablodan yapılabilecek bir diğer gözlem ise LSTM modellerinin eğitim veri kümesinde görmediği konulara genellemediği ve neredeyse rastgele seviyesinde bir doğruluk gösterdiği.

IV. TARTIŞMA

Bu çalışmada Türkçe diyalog metinlerinde konuşulan konunun değiştiği yeri otomatik olarak tespit etmek amacıyla iki farklı yöntem test edilmiştir. Öncelikle, etiketli bir Türkçe diyalog veri kümesinin mevcut olmaması sebebiyle Donanım Haber forum sitesinden belirli konularda yazılmış başlıkların altındaki girdiler indirilip ön işlemeden geçirilerek bir veri kümesi oluşturulmuştur. Bu veri kümesi kullanılarak konu değişimini simüle eden örnekler oluşturulmuş ve konu değişimi problemi ikili sınıflandırma problemine dönüştürülmüştür. TF-IDF ve LSTM tabanlı iki farklı yöntem ile örneklerde konu değişimi olup olmadığı tespit edilmeye çalışılmıştır. TF-IDF yöntemi ile %80'in üzerinde bir doğruluk ve %94'ün üzerinde geri çağırma performansı elde edilmiştir. LSTM modeli ile ise %76'nın üzerinde doğruluk ve %79'a yakın kesinlik elde edilmiştir. LSTM modellerindeki kelime temsil katmanını rastgele yerine önceden eğitilmiş vektörler ile başlatmanın performansı bir miktar artırdığı gözlemlenmiştir.

Önerilen yöntemlerin yeni konular için genelleme performansları test edilmiştir. TF-IDF yönteminin doğruluk performansı %60'lara düşerken LSTM modellerinin ise neredeyse

TABLO III: İncelenen modellerin ayrılan 4 konuluk veri kümesindeki test performansları

		Doğruluk (%)	Kesinlik (%)	Geri Çağırma (%)
n = 3	TF-IDF	59.2	55.4	86.9
	LSTM	50.1	48.8	33.8
	LSTM FT	51.4	50.7	38.3
	LSTM FTS	51.5	51.0	33.2
n = 5	TF-IDF	60.6	57.2	86.5
	LSTM	49.7	50.1	28.1
	LSTM FT	49.9	50.6	24.3
	LSTM FTS	50.6	51.9	24.9

hiç genelleme göstermediği ve rastgeleye yakın bir performans gösterdiği görülmüştür. LSTM modellerinin görece düşük performans göstermelerinin sebeplerinden biri oluşturulan veri kümelerinin çok sayıda yazım hatası barındırıyor olması olabilir. Yazım hataları sebebiyle veri kümesindeki eşsiz kelime sayısı artmakta ve derin öğrenme modellerinin eğitilmesi zorlaşmakta ve genelleme performansları düşmektedir. Bu nedenden ötürü, ilerdeki çalışmalarda veri kümesinin yazım hatalarından temizlenmesinin veya yazım hatası daha az olan yeni bir veri kümesi oluşturulmasının, konu değişimi tespiti problemi açısından faydalı olacağı öngörülmektedir.

KAYNAKLAR

- [1] J. Williams, A. Raux, D. Ramachandran, and A. Black, "The dialog state tracking challenge," in *Proc. of the SIGDIAL 2013 Conference*, 2013, pp. 404–413.
- [2] M. Henderson, B. Thomson, and S. Young, "Deep neural network approach for the dialog state tracking challenge," in *Proc. of the SIGDIAL 2013 Conference*, 2013, pp. 467–471.
- [3] J. Li, W. Monroe, A. Ritter, M. Galley, J. Gao, and D. Jurafsky, "Deep reinforcement learning for dialogue generation," *arXiv preprint arXiv:1606.01541*, 2016.
- [4] J. Huang, M. Zhou, and D. Yang, "Extracting chatbot knowledge from online discussion forums," in *IJCAI*, vol. 7, 2007, pp. 423–428.
- [5] B. A. Shawar and E. Atwell, "Using dialogue corpora to train a chatbot," in *Proc. of the Corpus Linguistics 2003 conf.*, 2003, pp. 681–690.
- [6] J. Allan, J. G. Carbonell, G. Doddington, J. Yamron, and Y. Yang, "Topic detection and tracking pilot study final report," 1998.
- [7] J. M. Schultz and M. Liberman, "Topic detection and tracking using idf-weighted cosine coefficient," in *Proc. of the DARPA broadcast news workshop*. San Francisco: Morgan Kaufmann, 1999, pp. 189–192.
- [8] L. AlSumait, D. Barbará, and C. Domeniconi, "On-line lda: Adaptive topic models for mining text streams with applications to topic detection and tracking," in *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*. IEEE, 2008, pp. 3–12.
- [9] H. P. Luhn, "A statistical approach to mechanized encoding and searching of literary information," *IBM Journal of research and development*, vol. 1, no. 4, pp. 309–317, 1957.
- [10] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.
- [11] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135–146, 2017.
- [12] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in neural information processing systems*, 2013, pp. 3111–3119.
- [13] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proc. of the 2014 conf. on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.
- [14] E. Grave, P. Bojanowski, P. Gupta, A. Joulin, and T. Mikolov, "Learning word vectors for 157 languages," in *Proc. of the International Conference on Language Resources and Evaluation (LREC 2018)*, 2018.