

Doğal Dil Çıkarımı Modellerinde BERT Vektörlerinin Başarım Değerlendirmesi

Performance Evaluation of BERT Vectors on Natural Language Inference Models

İskender Ülgen OĞUL¹, Selma TEKİR¹

¹Bilgisayar Mühendisliği, İzmir Yüksek Teknoloji Enstitüsü, İzmir, Türkiye
iskenderogul@gmail.com, selmatekir@iyte.edu.tr

Özetçe —Doğal dil çıkarımı, düşünce ifade eden cümlelerin arasındaki ilişkiyi; karşıtlık, gerekseme veya tarafsızlık olarak sınıflandırmayı hedefler. Sınıflandırma görevini gerçekleştirmek için metinsel kaynaklar, vektör ya da gömme olarak adlandırılan matematiksel gösterimlere dönüştürülür. Bu çalışmada, hem statik (Glove, OntoNotes5) hem de bağlamsal (BERT) kelime gömme yöntemleri kullanılmıştır. Fikirsal cümleler arasındaki mantıksal ilişkilerin sınıflandırılması zordur zira cümleler karmaşık gramer yapılarına sahiptir ve cümlelerin işlenerek mantıksal gösterimlere dönüştürülmesi geleneksel doğal dil işleme çözümleri ile yetersiz kalmaktadır. Bu çalışma, sınıflandırma görevini gerçekleştirmek için ayrıştırılabilir ilgi ve doğal dil çıkarımı için gelişmiş LSTM (ESIM) derin öğrenme modellerini kullanmıştır. En iyi sonuç olan %88 doğruluk değeri SNLI veri kümesi üzerinde ESIM-BERT ile elde edilmiştir.

Anahtar Kelimeler—Doğal Dil Çıkarımı, BERT, Ayrıştırılabilir İlgi, ESIM, SNLI.

Abstract—Natural language inference aims to classify the binary relation between opinionated sentences as a contradiction, entailment, or neutral. To accomplish the task, classifiers transform textual data into numerical representations called vectors or embeddings. In this study, both static (Glove, OntoNotes5) and contextual (BERT) word embedding methods are used. Classifying the logical relationships between opinionated sentences is difficult. These sentences have complex grammatical structures to convert them into logical representations, and traditional natural language processing solutions are insufficient to meet the requirement. This study uses Decomposable Attention and Advanced LSTM for Natural Language Inference (ESIM) deep learning methods to perform this classification. The best accuracy score is achieved with 88% using ESIM - BERT on the SNLI corpus.

Keywords—Natural Language Inference, BERT, Decomposable Attention, ESIM, SNLI.

I. GİRİŞ

Teknolojide düşen maliyet ve yükselen hız ile çevrimiçi olan birey sayısının artmasıyla, üretilen veri miktarı katlanmaktadır. Bireylerin gizliliği gözetilerek toplanan ve dizinlenen bu veriler araştırma alanları, şirketler ve devletler için bilgi potansiyeline sahiptir. Bu veriler, tecrübe edilen bir olay, politik görüş veya günlük olgular çerçevesinde şekillenmektedir. Elde edilen verilerin, uygun yöntemler kullanılarak işlenmesi, problemlere çözüm sunabilir ve belirleyici sonuçlar ortaya

çıkabilir. Bu çalışma fikirsal cümleleri, doğal dil çıkarımı yöntemleri kullanarak sınıflandırmayı hedeflemektedir.

Doğal dil çıkarımı (*Natural Language Inference*) popüler, ve oldukça da karmaşık bir problemdir. Doğal dil çıkarımı, fikirsal cümleler arasındaki ilişkiyi, karşıtlık - çelişki (*contradiction*), örtüşme - gerekseme (*entailment*) ve tarafsızlık (*neutral*) olarak sınıflandırmayı hedefler. Çelişki ya da karşıtlık, iki fikirsal metnin, birbiri ile zıt görüşte olması veya semantik olarak birbirlerini desteklemediği bir durumdur. Örtüşme veya gerekseme, iki metnin birbirlerini destekleyen iddialara sahip olması, tarafsızlık ise metinlerin birbirinden bağımsız görüşler içermesi durumudur. Doğal dil çıkarımı, iki metin arasındaki semantik benzerlikten çok, birbirleri arasındaki semantik ilişkiyi ve alt bağlamı ortaya çıkarmaya odaklanmaktadır.

Öncül çalışmalarda, doğal dil çıkarım problemleri, sözcüksel ve yapısal bilgilere dayanılarak çözülmekteydi [1]. Yapısal bilgilerin çıkarımı WordNet, ardından SentiWordNet ile gerçekleştirilmiştir [2] [3]. Önceden eğitilmiş kelime vektörlerinin (*Glove*, *Word2Vec*, *Fasttext*) ortaya çıkması ile beraber, doğal dil işleme çalışmaları derin öğrenme yaklaşımlarına evrilmiştir [4] [5] [6]. Kelime vektörlerinin sabit değerlerle ifade edilmesi, sözcüklerin bağlamsal anlamını göz ardı etmektedir. Google AI ekibi ortaya çıkan bu eksikliğe dikkat çekerek, BERT (*Deep Bidirectional Transformers*) dil modelini geliştirmiştir [7]. BERT, kelimenin cümle içerisindeki bağlamını gözeterek vektör temsili üretmektedir.

Doğal dil çıkarımı araştırma alanı uzun bir süre RTE (*Recognizing Textual Entailment*) benzeri küçük boyutlu veri kümeleriyle kısıtlıydı [8]. Bu kısıt, 2015 yılında Stanford doğal dil işleme grubunun sunduğu, Stanford Natural Language Inference (SNLI) veri kümesi ile çözüme kavuşmuştur. Sunulmuş olan yeni veri kümesinin en önemli özelliği, tamamıyla alanında uzman kişiler tarafından manuel olarak etiketlenmiş olmasıdır [9].

Doğal dil çıkarımı çalışmaları, karmaşık gramer yapısından dolayı gelişmiş tasarıma sahip derin öğrenme mimarilerine ihtiyaç duymaktadır. Bu çalışmada, doğal dil çıkarımı için özel olarak hazırlanmış ve kendini kanıtlamış ayrıştırılabilir ilgi [10] ve ESIM [11] modelleri kullanılmıştır. Kullanılan modellerin yapısal mimarileri korunarak, yüksek doğruluklu kelime gömme yöntemleri olan, ön eğitilmiş Glove, OntoNotes5 ve bağlamsal kelime gömme yöntemi BERT uyarlanmıştır. Bu

çalışma, öncüllerinden farklı olarak ön eğitilmiş BERT ve bağlamsal BERT vektörlerinin, Ayrıştırılabilir İlgili ve ESIM mimarileri ile birlikte kullanıldığı özgün bir uyarlama çalışmasıdır. SNLI, güncel ve doğruluğu yüksek çalışmaları, karşılaştırmalı liderlik tablosu olarak sunmaktadır. Bu çalışmanın gerçekleştirildiği tarihte SNLI liderlik tablosunda raporlanan en yüksek doğruluk değeri olan %92.1 [12] referans noktası olarak alınmıştır. Çalışmamızda sunulan ESIM-BERT uyarlaması ile %88.0 doğruluk düzeyine, yönetilebilir ve karmaşıklığı az bir yaklaşım ile ulaşılmıştır.

II. İLGİLİ ÇALIŞMALAR

Harabagiu vd. [1], karşıtlık analizini gerçekleştirirken, makine öğrenmesi algoritmalarını, zıt anlamlılık, olumsuzluk ve pragmatik sözcüksel bilgilerini baz alarak kullanmıştır. Sundukları çalışma, zıt anlam ilişkisini incelemek için WordNet ve çözümün doğruluğunu sınamak için de 2006 RTE PASCAL veri kümesini kullanarak en yüksek %75.38 değerini elde etmiştir.

Marneffe vd. [13], metinler içindeki çelişkiyi bulmanın, örtüşmeyi bulmaktan daha zor olduğunu ifade etmiştir. Çelişkiyi; olumsuzluk, zıt anlamlılık, sayısal ve kelime bilgilerindeki tutarsızlık olarak tanımlamaktadır. Çalışmada, düğüm benzerliğini elde etmek için, metinler bağımlılık çizgelerine çevrilmiştir. Hizalanan çizgeler, lojistik regresyon modeli oluşturmak için kullanılmıştır.

Öncüllerinden farklı olarak Ritter vd. [14], karşıtlık analizinde, sözcüksel bilgi ve arka plan bilgisini beraber kullanmaktadır. Çalışmalarında, görünüşte karşıtlık koşulu içeren veri kümesi oluşturmuş ve gerçek karşıtlık koşulu içeren veri kümesi ile birleştirilmiştir. Sundukları çözüm, arka plan bilgisini kullanarak, görünüşte karşıtlık içeren cümleleri başarı ile ayırt etmiştir.

Ön eğitilmiş kelime vektörleri ile beraber, doğal dil çıkarımı problemlerinin çözümünde veri temelli yaklaşımlara yönelinmiştir. Parikh vd. [10], karmaşık dil çıkarımı problemini, ilgi (*attention*) kullanarak alt problemlere bölmüştür. Ardından, alt problemleri paralel şekilde çözerek, çok az parametre ile yüksek doğruluk değerine ulaşmıştır.

Kelime vektörleri, büyük bir veri kümesi içindeki kelimelerin, birbirleriyle olan ilişkilerinin çok boyutlu uzaya yoğun gösterimler olarak temsil edilmesidir. İlk olarak 2013 yılında Mikolov vd.'nin sunmuş olduğu Word2Vec yaklaşımı [4] ile doğal dil işleme alanına tanıtılmıştır. 2014 senesinde, Pennington vd. [5], rastgele ağırlık yaklaşımı yerine, kelimelerin beraber görülme oranlarının başlangıç noktası olması gerektiğini öne sürerek, Glove kelime gömmelerini oluşturmuştur. 2018 senesinde Mikolov vd., kelime vektörlerinin kelimeler içerisindeki karakter n-gram'larının toplamı olarak temsil edilebileceğini öne sürerek, Fasttext [6] ön eğitilmiş kelime vektörlerini geliştirmiştir.

2017 yılında Chen vd. [11] zincir LSTM mimarisini, çift yönlü olarak (*BiLSTM*) kullanmıştır. Oluşturdukları mimari, Glove kelime gömme yöntemini ilgi (*attention*) yapısıyla beraber kullanarak, doğal dil çıkarımı görevinde sunulduğu dönemdeki en iyi sonucu elde etmiştir.

En gelişmiş dil modeli olarak değerlendirilen BERT [7], maskelenmiş dil modeli kullanarak bilgiyi çift yönde işleyerek

dönüştürme ağlarına [15] uyarlamıştır. Çift yönlü yaklaşım, bilgiyi zaman içerisinde ileri ve geri yönde işleyerek, yüksek doğruluklu çözümler sunmaktadır. Sunulan çözüm sayesinde, bağlamsal kelime vektörleri elde etmenin yanında, bir çok doğal dil işleme probleminde de en iyi sonuçlara ulaşılmaktadır.

III. VERİ KÜMESİ VE ÖZNETELİK ÇIKARIMI

A. Veri Kümesi

Doğal dil çıkarımı görevinin makine öğrenmesi yöntemleriyle çözülmesinde, veri kalitesi ve miktarı büyük bir rol oynamaktadır. Bu konuya dikkat çeken Stanford doğal dil işleme ekibi, SNLI [9] veri kümesini sunmuştur. SNLI veri kümesi, Mechanical Turk altyapısı kullanılarak manuel etiketleme ile oluşturulmuştur. Alanında uzman katılımcılardan, her bir sınıf için üç örnek oluşturulması ve kendilerine ait olmayan beş örneği etiketlemesi istenmiştir. Örnekler için altın etiket ise beş etiketin uzlaşmasından elde edilmiştir.

B. Kelime Gömme Yöntemleri

Glove: Glove [5], Word2Vec'den farklı bir şekilde, amaç fonksiyonu olarak ağırlıklı en küçük kareler toplamını benimsemiştir. Bu yaklaşımda, kelime olasılıkları yerine, beraber görülme oranları başlangıç noktası olarak alınmıştır. Eğitilen modeller kelime benzerliği ve kelime analojisi ile test edilmiş ve öncüllerinden daha iyi başarımlar gösterdiği görülmüştür.

OntoNotes5 - Spacy: Kelime vektör temsilleri genel olarak, Wikipedia gibi bağlamı yapılandırılmış tekil kaynaklar ile oluşturulmaktadır. OntoNotes veri kümesi, birden fazla kaynaktan veri alımı gerçekleştirerek, çeşitliliği artırma yaklaşımını benimsemiştir. Bu kaynaklar; haber metni, canlı yayın konuşmaları, telefon konuşmaları ve internet dokümanlarıdır. Spacy, OntoNotes5 veri kümesi ile eğitilmiş vektör modelini açık kaynak olarak sunmaktadır [16] [17].

BERT: Ön eğitilmiş kelime gömme modelleri, vektör temsili gerçekleştirirken, kelimelerin cümle içindeki bağlamını koruyamamaktadır. BERT [7], öncüllerinden farklı olarak, maskeli dil modeli (MLM) yaklaşımıyla, tüm katmanlarda bilgiyi çift yönlü olarak işler ve temsil üretir. Bu sayede vektör temsili gerçekleştirirken, kelimenin cümle içindeki bağlamını koruyarak daha doğru temsiller üretmektedir.

IV. DERİN ÖĞRENME METODLARI

A. Ayrıştırılabilir İlgili

Parikh vd. [10] tarafından tasarlanmış olan ayrıştırılabilir ilgili (*decomposable attention*) modeli; problemi ilgili, karşılaştırma ve bütünleştirme alt problemlerine bölmektedir. Model, girdi cümlelerini, a dayanak ve b hipotez olmak üzere her a_i ve b_i 'nin kelime vektörünü birleştirerek $a = (a_1, a_2, a_3 \dots a_{l_a})$ ve $b = (b_1, b_2, b_3 \dots b_{l_b})$ üzerinden temsil etmektedir. Sonuç olarak, eğitim verisi $\{a^{(n)}, b^{(n)}, y^{(n)}\}_{n=1}^n$ oluşturulmakta ve etiketler $y^{(n)} = (y_1^{(n)}, y_2^{(n)}, \dots, y_c^{(n)})$ olarak ifade edilmektedir.

İlgili (Attention): Cümleler arasındaki ilgili, ReLU aktivasyonunu kullanan, ileri yönlü F fonksiyonundan dönen vektörlerin skaler çarpımından elde edilmektedir. Ardından, elde edilen ilgili vektörleri (e_{ij}), normalize edilerek hizalama işlemi gerçekleştirilmektedir.

$$e_{ij} = F'(\bar{a}_i, \bar{b}_j) := F(\bar{a}_i)^T F(\bar{b}_j) \quad (1)$$

$$\alpha_j := \sum_{i=1}^{l_a} \frac{\exp(e_{ij})}{\sum_{k=1}^{l_a} \exp(e_{kj})} \bar{a}_i \quad \beta_i := \sum_{j=1}^{l_b} \frac{\exp(e_{ij})}{\sum_{k=1}^{l_b} \exp(e_{ik})} \bar{b}_j \quad (2) \quad (3)$$

Karşılaştırma: Bu aşamada, hizalanmış olan ilgi vektörleri (α_j ve β_i) birleştirilerek, parametreleri F ile aynı olan bir diğer ileri yönlü ağ G kullanılarak işlenmektedir.

$$v_{1,i} := G([\bar{a}_i, \beta_i]) \quad \forall i \in [1, \dots, l_a] \quad (4)$$

$$v_{2,j} := G([\bar{b}_j, \alpha_j]) \quad \forall j \in [1, \dots, l_b] \quad (5)$$

Bütünleştirme: Karşılaştırma katmanından elde edilen, $\{v_{1,i}\}_{i=1}^{l_a}$ ve $\{v_{2,j}\}_{j=1}^{l_b}$ sonuçları, bütünleştirme katmanında, toplama işlemine tabi tutulur.

$$v_1 = \sum_{i=1}^{l_a} v_{1,i} \quad (6) \quad v_2 = \sum_{j=1}^{l_b} v_{2,j} \quad (7)$$

Elde edilen toplamalar, parametreleri G ile aynı olan H ileri yönlü ağ katmanında işleminden geçirilir ($sonuc = H([v_1, v_2])$). Çıktılar, *softmax* aktivasyonu kullanılan sınıflandırma katmanına gönderilerek sınıf tahmini elde edilir.

B. ESIM

ESIM modeli, BiLSTM mimarisinin çift yönlü bilgi işleme kapasitesinden yararlanmaktadır. ESIM, doğal dil çıkarımı problemini; girdi kodlama, yerel çıkarım modellemesi ve havuz katmanı olarak üç aşamada ele almaktadır.

Girdi Kodlama: ESIM, vektör temsiline çevrilmiş kelimeleri birleştirerek cümleleri $a = (a_1, a_2, a_3 \dots a_{l_a})$ ve $b = (b_1, b_2, b_3 \dots b_{l_b})$ şeklinde temsil etmektedir. Vektör temsilleri, BiLSTM katmanında işlenerek \bar{a}_i and \bar{b}_j çıktıları elde edilmektedir.

$$\bar{a}_i = BiLSTM(a, i), \quad \forall i \in [1, \dots, l_a] \quad (8)$$

$$\bar{b}_j = BiLSTM(b, j), \quad \forall j \in [1, \dots, l_b] \quad (9)$$

Yerel Çıkarım Modellemesi: Yerel çıkarım aşamasında, BiLSTM çıktılarının hizalanması (*ilgi*) gerçekleştirilir. Cümleler arasındaki ilgi, \bar{a}_i 'nin $\{\bar{b}_j\}_{j=1}^{l_b}$ değerlerinin ağırlıklı toplamı olarak ifade edilmektedir.

$$e_{ij} = \bar{a}_i^T \cdot \bar{b}_j \quad (10)$$

$$\tilde{a}_i = \sum_{j=1}^{l_b} \frac{\exp(e_{ij})}{\sum_{k=1}^{l_b} \exp(e_{ik})} \bar{b}_j, \quad \forall i \in [1, \dots, l_a] \quad (11)$$

$$\tilde{b}_j = \sum_{i=1}^{l_a} \frac{\exp(e_{ij})}{\sum_{k=1}^{l_a} \exp(e_{kj})} \bar{a}_i, \quad \forall j \in [1, \dots, l_b] \quad (12)$$

Elde edilen ilgi değerleri, BiLSTM çıktıları ile birleştirilmektedir.

$$m_a = [\tilde{a}; \bar{a}; \bar{a} - \tilde{a}; \bar{a} \odot \tilde{a}] \quad (13)$$

$$m_b = [\tilde{b}; \bar{b}; \bar{b} - \tilde{b}; \bar{b} \odot \tilde{b}] \quad (14)$$

Birleştirilmiş olan vektör temsilleri, zincir LSTM denilen ikinci BiLSTM katmanında işleme alınır.

$$v_{a,i} = BiLSTM(m_a, i), \quad \forall i \in [1, \dots, l_a] \quad (15)$$

$$v_{b,j} = BiLSTM(m_b, j), \quad \forall j \in [1, \dots, l_b] \quad (16)$$

Havuz Katmanı: İkinci BiLSTM katmanından elde edilen sonuçlar, parametre sayısı ve vektör boyutunu düşürmek için havuzlama işlemine alınır.

$$V_{a,ave} = \sum_{i=1}^{l_a} \frac{v_{a,i}}{l_a}, \quad V_{a,max} = \max_{i=1}^{l_a} v_{a,i} \quad (17)$$

$$V_{b,ave} = \sum_{j=1}^{l_b} \frac{v_{b,j}}{l_b}, \quad V_{b,max} = \max_{j=1}^{l_b} v_{b,j} \quad (18)$$

$$v = [v_{a,ave}; v_{a,max}; v_{b,ave}; v_{b,max}] \quad (19)$$

Maksimum ve ortalamaları alınan vektörler, birleştirilerek aktivasyonu *tanh* olan MLP katmanına, ardından *softmax* aktivasyonunu kullanan sınıflandırma katmanına gönderilerek tahmin sonuçları elde edilir.

V. YÖNTEM TASARIMI

Bu çalışma, doğal dil çıkarımı problemine çözüm sunmak için ön eğitilmiş ve bağlamsal kelime vektörlerini ayrıştırılabilir ilgi ve ESIM modelleri ile kullanmıştır. İlk olarak, BERT modelinin birincil katmanından kelime vektörleri çıkartılıp, sözlük ile eşleştirilerek, 1024 boyutunda ve 32 bin kelime içeren ön eğitilmiş vektör modeli elde edilmiştir. İlerleyen süreçte, bellek eniyilemesi sorunu çözülerek, BERT bağlamsal vektörleri araştırma ortamına uyarlanmıştır.

Kapalı kutu gibi davranan derin öğrenme modellerinin başarımının açıklanmasına yardımcı olması için bağlamsal BERT ilgi vektörleri görsel olarak sunulmuştur. Eğitilen modeller, SNLI verisine ek olarak MultiNLI [18] ve ANLI [19] veri kümeleri ile de sınanmıştır. Modellerin gerçek dünya verisi üzerindeki başarımının araştırılması için Habernal ve Gurevych [20] çalışmasında sunulan UKPConvArg2 derleminden seçilen iki alt başlık içerisindeki veriler mevcut probleme uyarlanmış ve mevcut yöntemlerin bir sınaması gerçekleştirilmiştir. Zira, UKPConvArg2 derleminde bir konunun (*TV - Kitap*), alt başlıkları ile ilgili olarak sunulan ikili argümanlardan daha ikna edici olan işaretlenmiştir.

VI. DENEYSEL SONUÇLAR

Tablo I sonuçlarında görüldüğü gibi en yüksek doğruluk oranı Glove vektörleri ile SNLI verisi üzerinde elde edilmiştir. Diğer sonuçlarda ise en yüksek değer BERT ön eğitilmiş vektörlerinden elde edilmiştir. BERT vektörlerinin başarımı, yüksek boyutlu vektör temsili ve tam ayrıştırma (*Full tokenizer*) yaklaşımıyla ilişkilendirilmektedir. Tam ayrıştırma, kelimeleri kök ve eklerine detaylı bir şekilde parçalamaktadır. Deneysel değerlendirmeler, yüksek boyutlu (1024 *boyut*) BERT vektörlerinin kelimeleri kapsamlı bir şekilde temsil ettiğini göstermektedir. Ayrıştırılabilir ilgi modeli 1024 yığın (*batch*) değeri ile en iyi başarımı vermektedir, ancak BERT'in karmaşık tasarımı gereği bu denli yüksek yığın değerlerinde bellek hatası oluşturmaktadır. Düşük yığın değerlerinde ise model, bağlamsal vektörleri yeterince genelleştiremeyip, içerdiği bilgiyi öğrenememektedir.

MNLI ve ANLI veri kümelerinde doğruluk sonuçları düşme eğilimindedir. SNLI yapılandırılmış veri kümesidir, MNLI ve ANLI ise farklı konular ve bağlamlardan derlenmiş

TABLE I: AYRIŞTIRILABİLİR İLGi SONUÇLARI

	Train	Dev	SNLI Test	MNLI Matched	MNLI Mismatched	ANLI Test
OntoNotes5	84.05	85.43	85.39	56.95	58.93	28.59
Glove	84.23	85.48	85.40	57.35	59.60	30.21
BERT Ön Eğitim	88.45	86.15	85.23	61.30	61.16	30.87
BERT Bağlamsal	—	—	—	—	—	—

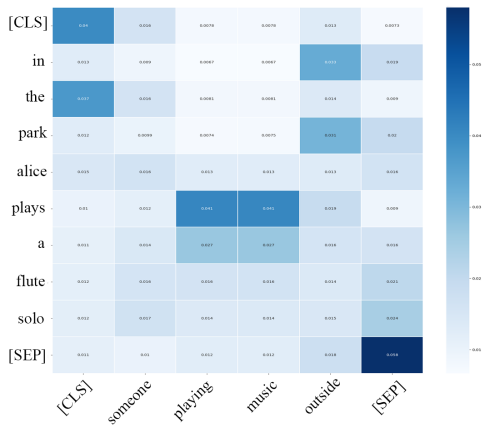
örnekler içerir. MNLI matched ve mismatched alt veri kümeleri on farklı konudan derlenmiş ve eşit şekilde paylaşılmış örnekler barındırır. ANLI günümüzün en zor veri kümesi olarak değerlendirilmektedir.

Tablo II sonuçlarına bakıldığında, ESIM, BiLSTM yapısını çok iyi şekilde kullanarak, karmaşık bağlamsal vektör temsillerini en iyi şekilde öğrenmektedir. Model üzerindeki en yüksek doğruluk oranları BERT bağlamsal vektörleri ile elde edilmiştir.

TABLE II: ESIM SONUÇLARI

	Train	Dev	SNLI Test	MNLI Matched	MNLI Mismatched	ANLI Test
OntoNotes5	87.36	86.22	86.14	57.58	59.22	30.28
Glove	89.34	85.79	86.02	55.73	57.21	30.00
BERT Ön Eğitim	86.93	87.49	86.76	61.00	62.23	30.46
BERT Bağlamsal	89.45	88.44	88.00	66.90	67.75	31.59

Şekil 1’de, ESIM-baglamsal BERT modelinin doğal dil çıkarımı gerçekleştirirken, hangi kelimelere daha çok önem verdiği, sıcaklık haritası ile gösterilmiştir. Şekilden de görülebildiği gibi "plays - playing music" ve "in, park - outside" kelimelerine daha fazla önem gösterilmektedir. Verilen örnek için, BERT bağlamsal vektörleri ile eğitilen ESIM modelinden karşıtlık: 0.009, gerekseme 0.651, tarafsızlık 0.34 sonuçları elde edilmiş ve gerekseme tahminlemesi yapılmıştır.



Şekil 1: Bağlamsal BERT ilgi görselleştirme örneği

TABLE III: UKP VERİSİ TEST SONUÇLARI

Konu	Karşıtlık	Gerekseme	Tarafsızlık
Tv - Kitap Karşıtlık	%48.3	%2.80	%48.9
Tv - Kitap Gerekseme	%46,9	%3.30	%49.8
Evrım - Yaradılış Karşıtlık	%53.1	%7.00	%39.9
Evrım - Yaradılış Gerekseme	%55.1	%4.5	%40.4

Tablo III, en yüksek doğruluk sonucu veren ESIM - BERT doğal dil çıkarımı modelinin gerçek dünya verisi üzerindeki davranışını göstermektedir. Bunun için iki farklı konuya ait

karşıtlık ve gerekseme verileri oluşturulmuştur. Bu konular televizyon - kitap ve evrim - yaradılış tartışmalarıdır. Sonuçlara bakıldığında model karşıtlık çıkarımında daha iyi başarımlar gösterirken, gerekseme ve tarafsızlık arasındaki ayrımı yeterince iyi yapamamaktadır.

UKP verisi incelendiğinde konuşma diline yakın bir dil kullanılmakta olduğu ve görüşlerin uzun cümleler ile ortaya konulduğu görülmektedir. Bu durum SNLI gibi yapılandırılmış, dayanak ve hipotezin kısa ve düzgün cümlelerden kurulduğu veri kümelerinde eğitilen modellerin başarımları olumsuz etkilemektedir. Dolayısıyla doğal dil çıkarımı alanında, sosyal medya verisi gibi gerçek veri üzerinde genelleştirme başarımları yüksek çözümlere ihtiyaç vardır.

KAYNAKLAR

- [1] S. M. Harabagiu, A. Hickl, and V. Lacatusu, "Negation, contrast and contradiction in text processing," in *AAAI*, 2006.
- [2] G. Miller, "Wordnet: a lexical database for english," *Commun. ACM*, vol. 38, pp. 39–41, 1995.
- [3] A. Esuli and F. Sebastiani, "Sentiwordnet: A publicly available lexical resource for opinion mining," in *LREC*, 2006.
- [4] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," *ArXiv*, vol. abs/1310.4546, 2013.
- [5] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *EMNLP*, 2014.
- [6] T. Mikolov, E. Grave, P. Bojanowski, C. Puhresch, and A. Joulin, "Advances in pre-training distributed word representations," *ArXiv*, vol. abs/1712.09405, 2018.
- [7] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *NAACL-HLT*, 2019.
- [8] E. Voorhees, "Contradictions and justifications: Extensions to the textual entailment task," in *ACL*, 2008.
- [9] S. R. Bowman, G. Angeli, C. Potts, and C. D. Manning, "A large annotated corpus for learning natural language inference," in *EMNLP*, 2015.
- [10] A. P. Parikh, O. Täckström, D. Das, and J. Uszkoreit, "A decomposable attention model for natural language inference," in *EMNLP*, 2016.
- [11] Q. Chen, X.-D. Zhu, Z. Ling, S. Wei, H. Jiang, and D. Inkpen, "Enhanced lstm for natural language inference," in *ACL*, 2017.
- [12] J. Pilault, A. Elhattami, and C. Pal, "Conditionally adaptive multi-task learning: Improving transfer learning in nlp using fewer parameters & less data," *ArXiv*, vol. abs/2009.09139, 2020.
- [13] M.-C. de Marneffe, A. N. Rafferty, and C. D. Manning, "Finding contradictions in text," in *ACL*, 2008.
- [14] A. Ritter, S. Soderland, D. Downey, and O. Etzioni, "It's a contradiction - no, it's not: A case study using functional relations," in *EMNLP*, 2008.
- [15] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *NIPS*, 2017.
- [16] R. Weischedel, E. Hovy, M. Marcus, and M. Palmer, "Ontonotes: A large training corpus for enhanced processing," 2017.
- [17] M. Honnibal, I. Montani, S. Van Landeghem, and A. Boyd, "spacy: Industrial-strength natural language processing in python," 2020.
- [18] A. Williams, N. Nangia, and S. R. Bowman, "A broad-coverage challenge corpus for sentence understanding through inference," in *NAACL-HLT*, 2018.
- [19] Y. Nie, A. Williams, E. Dinan, M. Bansal, J. Weston, and D. Kiela, "Adversarial nli: A new benchmark for natural language understanding," *ArXiv*, vol. abs/1910.14599, 2020.
- [20] I. Habernal and I. Gurevych, "What makes a convincing argument? empirical analysis and detecting attributes of convincingness in web argumentation," in *EMNLP*, 2016.