**BERT Modelinin Stack Overflow Soru Sınıflandırılmasındaki Performans Değerlendirmesi**

**Özet**

Bu çalışma, doğal dil işleme (NLP) alanında son yıllarda büyük ilgi gören BERT (Bidirectional Encoder Representations from Transformers) modelinin performansını incelemektedir. BERT, iki yönlü transformer mimarisini kullanarak metinlerdeki bağlam bilgisini etkili bir şekilde yakalamaktadır. Bu çalışmanın amacı, BERT modelinin metin sınıflandırma ve anlamsal benzerlik gibi çeşitli metin işleme görevlerindeki doğruluğunu ve etkinliğini değerlendirmektir. Çalışma kapsamında, BERT modeli Stack Overflow veri seti üzerinde eğitilmiş ve test edilmiştir. Modelin performansını değerlendirmek için doğruluk, F1 skoru, geri çağırma ve kesinlik gibi ölçütler kullanılmıştır. Çalışma sonucunda, BERT modelinin metin sınıflandırma görevlerinde yüksek doğruluk oranlarına ulaştığı ve başarılı sonuçlar verdiği görülmüştür. Bu çalışma, BERT modelinin doğal dil işleme görevlerindeki etkinliğini vurgulamakta ve gelecekteki çalışmalar için önemli bir referans noktası oluşturmaktadır. Elde edilen sonuçlar, BERT modelinin metin analizi ve işleme alanında güçlü bir araç olduğunu göstermektedir.

**Giriş**

NLP, bilgisayarların insan dilini anlaması, yorumlaması ve oluşturması için kullanılan bir teknoloji alanıdır[1]. Son yıllarda, NLP alanındaki gelişmeler, büyük ölçekli dil modellerinin kullanımını yaygınlaştırmıştır[2]. Bu modeller, metinlerin daha iyi anlaşılmasını ve işlenmesini sağlamakta, dil tabanlı görevlerde önemli başarılar elde etmektedir. BERT, NLP alanında devrim niteliğinde bir modeldir. Google tarafından geliştirilen BERT, iki yönlü transformer mimarisi kullanarak metinlerin bağlam bilgisini yakalamaktadır[3]. Bu yaklaşım, modelin her kelimeyi hem sol hem de sağ bağlamıyla birlikte anlamasına olanak tanır. Bu özellik, BERT'i özellikle metin sınıflandırma, anlamsal benzerlik ve dil modelleme gibi görevlerde etkili kılar[4].

BERT'in temel özelliklerinden biri, önceden eğitimli bir model olmasıdır[5]. Bu, modelin geniş bir metin korpusu üzerinde eğitildiği ve daha sonra çok özel görevler için ince ayar yapılabildiği anlamına gelir[6]. Bu yaklaşım, modelin çeşitli NLP görevlerinde yüksek performans göstermesini sağlar. BERT, sadece dil modeli olarak değil, aynı zamanda ince ayar yapılabilir bir model olarak da kullanılır ve bu, modelin esnekliğini ve uygulama alanını genişletir[7]. BERT modeli, çift yönlü dil temsillerini kullanarak metinlerin anlamını daha iyi kavrayabilen ve bu sayede metin sınıflandırma performansını artıran bir modeldir[8].

Bu çalışmada, BERT modelinin güçlü temsillerini elde etmek için Hugging Face Transformers kütüphanesi kullanılmış ve bu temsiller Logistic Regression, Random Forest, Support Vector Machine (SVM) makine öğrenmesi algoritmaları ile birleştirilmiştir. Çalışmanın ilk adımlarında, gerekli Python kütüphaneleri yüklenmiş ve metin ön işleme süreci tamamlanmıştır. Bu süreçte, metinler temizlenmiş, etiketler sayısallaştırılmış ve BERT modeli ile metin gömülmeleri elde edilmiştir. Ardından, elde edilen BERT gömülmeleri kullanılarak makine öğrenmesi modelleri eğitilmiş ve test veri seti üzerinde değerlendirilmiştir.

**İlgili Çalışmalar**

NLP alanında, transformer tabanlı modeller son yıllarda büyük ilgi görmüştür. Transformer mimarisi, dikkat mekanizması kullanarak uzun bağımlılıkları modellemek için geliştirilmiş ve bu sayede dil modellerinin performansında önemli iyileşmeler sağlamıştır[9]. Bu bölümde, BERT modeli ve benzer modeller hakkında literatürde yer alan çalışmalardan bazıları incelenmiştir.

**Transformer Modelleri ve BERT**

Vaswani ve arkadaşları 2017 yılında yaptıkları çalışmalarında Transformer mimarisi, sıralı verilerin işlenmesinde devrim yaratmıştır. Transformer, tamamen dikkat mekanizmalarına dayanmakta olup, dil modellerinin paralel olarak eğitilmesine olanak tanımaktadır[10]. Bu, modellerin daha hızlı ve verimli bir şekilde eğitilmesini sağlamıştır. BERT, Devlin ve arkadaşları tarafından tanıtılmıştır[11]. BERT, çift yönlü (bidirectional) bir transformer mimarisi kullanarak bağlam bilgisini her iki yönden de yakalamaktadır. Bu özellik, BERT'in metin anlama görevlerinde üstün performans göstermesine olanak tanır. BERT'in önceden eğitimli bir model olması, farklı NLP görevleri için ince ayar yapılabilmesini mümkün kılmaktadır[12].

**BERT'in Performans Değerlendirmeleri**

Sun ve arkadaşları 2019 yılında yaptıkları çalışmalarında, BERT modelini çeşitli NLP görevlerinde incelemiş ve modelin metin sınıflandırma, adlandırılmış varlık tanıma (NER), ve soru-cevap gibi görevlerde üstün performans sergilediğini göstermiştir[13]. Özellikle metin sınıflandırma görevlerinde, BERT'in doğruluk ve F1 skoru gibi ölçütlerde geleneksel yöntemlere kıyasla daha yüksek performans sağladığı görülmüştür.

**Benzer Modeller ve Karşılaştırmalar**

BERT'in başarısının ardından, çeşitli varyantlar ve alternatif modeller geliştirilmiştir. Liu ve arkadaşları, 2019 yılında yaptıkları RoBERTa isimli çalışmalarında BERT modelinin eğitim sürecini iyileştirerek daha yüksek performans elde etmiştir[14]. Yang ve arkadaşları, 2019 yılında yaptıkları çalışmalarında, BERT'in eksikliklerini gidermeyi amaçlayan XLNet isimli başka modeli geliştirmişler ve permütasyon bazlı dil modelleme yaklaşımı kullanarak bağlam bilgisini daha etkin bir şekilde yakalamışlardır[15].

**Stack Overflow Soru Sınıflandırılması**

Stack Overflow üzerindeki soruların kalitesini değerlendirmek ve sınıflandırmak, topluluk yönetimi ve bilgi paylaşımı açısından önemlidir[16]. Faseeh ve arkadaşları çalışmalarında, NLP'de yinelenen soruları tespit etme konusunda literatüre önemli katkılar sağlamışlardır. Derin öğrenme tabanlı modellerin başarılı bir şekilde kullanılması ve metin verilerindeki bağımlılıkları yakalamadaki yetkinlikleri, araştırmanın öne çıkan unsurları arasındadır. Ayrıca, farklı kelime yerleştirme tekniklerinin uygulanması, metodolojinin esnekliğini ve uyarlanabilirliğini artırmaktadır[17].

**Yöntem ve Veri Seti**

**Veri Seti**

Bu çalışmada kullanılan veri seti, 2016-2020 yılları arasında Stack Overflow'da sorulmuş 60,000 sorudan oluşmaktadır. Sorular, yüksek kaliteli (HQ), düşük kaliteli ve düzenlenmiş (LQ\_EDIT), ve düşük kaliteli ve kapatılmış (LQ\_CLOSE) olmak üzere üç kategoriye ayrılmıştır. Bu kategorilere dayalı olarak, BERT modelinin performansı değerlendirilmiş ve modelin soru sınıflandırma görevindeki etkinliği analiz edilmiştir. Bu veri seti, BERT modelinin eğitim ve test süreçlerinde kullanılmak üzere hazırlanmıştır.

**Yöntem**

BERT modelinin metin sınıflandırma görevindeki performansını değerlendirmek için aşağıdaki adımlar izlenmiştir:

**Veri Ön İşleme:**

1. Metinler küçük harflere dönüştürüldü.
2. Noktalama işaretleri ve özel karakterler temizlendi.
3. Sayılar kaldırıldı.
4. Veri setindeki Y sütunu sayısallaştırıldı. Sınıflar 0,1,2 yapıldı.

**BERT Modelinin Kullanımı:**

1. BERT modelinin bert-base-uncased versiyonu kullanıldı.
2. Tokenizer, metinleri BERT modeline uygun hale getirmek için kullanıldı.
3. Metinler tokenize edilerek BERT modeline girdi olarak verildi.

**Model Eğitimi ve Değerlendirme:**

1. Model, eğitim veri seti üzerinde eğitildi ve doğrulama veri seti üzerinde değerlendirildi.
2. Performans ölçütleri olarak doğruluk, geri çağırma, F1 skoru ve kesinlik kullanıldı.

**Sonuçlar**

Yapılan deneyler sonucunda, BERT modelinin metin sınıflandırma ve anlamsal benzerlik görevlerinde yüksek doğruluk ve etkinlik sağladığı görülmüştür.

Tablo 1. Model karşılaştırma.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Doğruluk** | **F1-Skoru (0)** | **F1-Skoru (1)** | **F1-Skoru (2)** |
| Logistic Regression | 0.847 | 0.81 | 0.79 | 0.95 |
| Random Forest | 0.738 | 0.72 | 0.66 | 0.83 |
| SVM | 0.847 | 0.8 | 0.79 | 0.95 |

Tablo 2. Geri çağırma ve kesinlik değerleri.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Kesinlik** | | | **Geri Çağırma** | | |
| **Sınıf (0)** | **Sınıf (1)** | **Sınıf (2)** | **Sınıf (0)** | **Sınıf (1)** | **Sınıf (2)** |
| Logistic Regression | 0.8 | 0.8 | 0.94 | 0.81 | 0.77 | 0.96 |
| Random Forest | 0.71 | 0.68 | 0.81 | 0.73 | 0.64 | 0.85 |
| SVM | 0.79 | 0.8 | 0.95 | 0.81 | 0.77 | 0.96 |

Logistic Regression ve SVM modelleri, Random Forest modeline göre daha yüksek doğruluk ve f1-score değerleri elde etmiştir. Bu sonuçlar, BERT modelinin metin sınıflandırma görevlerinde güçlü bir temsil yeteneğine sahip olduğunu ve Logistic Regression ile SVM modellerinin bu temsilleri başarılı bir şekilde kullandığını göstermektedir.

**Kaynaklar**

[1] D. Jurafsky and J. H. Martin, “Speech and language processing: An introduction to speech recognition,” *Comput. Linguist. Nat. Lang. Process. Edn., Prentice Hall, ISBN*, vol. 10, no. 0131873210, pp. 794–800, 2008.

[2] E. Sadikoğlu *et al.*, “The Evolution and Impact of Large Language Model Chatbots in Social Media: A Comprehensive Review of Past, Present, and Future Applications,” vol. 6, no. 2, pp. 67–76, 2023, Accessed: Jun. 06, 2024. [Online]. Available: https://www.researchgate.net/publication/376713755

[3] M. ÖZKAN and G. KAR, “TÜRKÇE DİLİNDE YAZILAN BİLİMSEL METİNLERİN DERİN ÖĞRENME TEKNİĞİ UYGULANARAK ÇOKLU SINIFLANDIRILMASI,” *Mühendislik Bilim. ve Tasarım Derg.*, vol. 10, no. 2, pp. 504–519, Jun. 2022, doi: 10.21923/jesd.973181.

[4] Ö. Özyurt, S. Barış, S. Rzayev, and H. Özyurt, “Makale Bir İnsan Bilgisayar Etkileşimi Örneği: Sesli Komutlar İle Veri Tabanı Sorgulama Uygulaması,” *Karadeniz Fen Bilim. Derg.*, vol. 14, no. 1, pp. 211–223, 2024, doi: 10.31466/kfbd.1384401.

[5] Ş. Bayrak, A. Karaca, F. Toson, A. Kocabey, and F. B. Arslanoǧlu, “Detection of Hate Speech in Turkish Social Media Posts with BERT-Base Model,” *31st IEEE Conf. Signal Process. Commun. Appl. SIU 2023*, no. 2022, pp. 2022–2025, 2023, doi: 10.1109/SIU59756.2023.10224040.

[6] S. Yücesoy Kahraman, A. Durmuşoğlu, and T. Dereli, “Ön eğitimli Bert modeli ile patent sınıflandırılması,” *Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Derg.*, vol. 4, pp. 2485–2496, 2023, doi: 10.17341/gazimmfd.1292543.

[7] Ö. Yiğit and Ş. Demİr, “Ön Eğitimli Dil Modelleriyle Duygu Analizi Sentiment Analysis with Pre-Trained Language Models,” 2023.

[8] S. S. Helli, S. Tanberk, and S. N. Cavsak, “Resume Information Extraction via Post-OCR Text Processing,” *2023 Innov. Intell. Syst. Appl. Conf. ASYU 2023*, pp. 1–5, 2023, doi: 10.1109/ASYU58738.2023.10296715.

[9] T. B. Brown *et al.*, “Language Models are Few-Shot Learners”.

[10] A. Vaswani *et al.*, “Attention is all you need,” *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Nips, pp. 5999–6009, 2017.

[11] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 1, no. Mlm, pp. 4171–4186, 2019.

[12] Ş. Ozan, U. Özdil, D. Taşar, B. Arslan, and G. Polat, “BERT Modeli ’ nin Sınıflandırma Doğruluğunu n Sıfır - Atış Öğrenmesi ile Artırılması Increasing the Classification Accuracy of BERT Model with Zero-Shot Learning,” pp. 99–108, 2021.

[13] X. Chen, M. X. Sun, D. Wu, and X. Y. Song, “Information-Sharing Behavior on WeChat Moments: The Role of Anonymity, Familiarity, and Intrinsic Motivation,” *Front. Psychol.*, vol. 10, no. November, pp. 1–15, 2019, doi: 10.3389/fpsyg.2019.02540.

[14] Y. Liu *et al.*, “RoBERTa: A Robustly Optimized BERT Pretraining Approach,” no. 1, 2019.

[15] Z. Yang, Z. Dai, Y. Yang, and J. Carbonell, “XLNet : Generalized Autoregressive Pretraining for Language Understanding,” no. NeurIPS, pp. 1–18, 2019.

[16] I. Chehreh, “Advanced Automated Tagging for Stack Overflow : A Multi-Stage Approach Using Deep Learning and NLP Techniques,” *2024 20th CSI Int. Symp. Artif. Intell. Signal Process.*, pp. 1–6, 2024, doi: 10.1109/AISP61396.2024.10475258.

[17] M. Faseeh, M. A. Khan, N. Iqbal, F. Qayyum, A. Mehmood, and J. Kim, “Enhancing User Experience on Q&A Platforms: Measuring Text Similarity Based on Hybrid CNN-LSTM Model for Efficient Duplicate Question Detection,” *IEEE Access*, vol. 12, no. March, pp. 34512–34526, 2024, doi: 10.1109/ACCESS.2024.3358422.