T5 Modeli ile Metin Özeti Çıkarımı ve ROUGE Skorları

bÜYÜK DİL modelleri final projesi

221229067

MECİT MERT BİŞGİN

2025

**Özet (Abstract)**

Bu rapor, T5 (Text-to-Text Transfer Transformer) modelinin metin özetleme görevinde kullanılmasını ve elde edilen özetin doğruluğunun ölçülmesini ele almaktadır. T5, özellikle dil modelleme ve metin üretme görevlerinde oldukça güçlü bir model olarak öne çıkmaktadır. Çalışmada, modelin farklı yapılandırmaları (t5-small, t5-large, t5-3b) kullanılarak, metinlerin özetleri çıkarılmış ve özetlerin doğruluğu, metin özetleme alanında yaygın olarak kullanılan ROUGE (Recall-Oriented Understudy for Gisting Evaluation) metrikleri ile değerlendirilmiştir. Bu metrikler, özetin orijinal metinle olan benzerliğini ölçmek için ROUGE-1 (unigram benzerliği), ROUGE-2 (bigram benzerliği) ve ROUGE-L (en uzun ortak alt dizin) olmak üzere üç ana metriği içermektedir.

Ayrıca, modelin performansı işlem süresi ve bellek kullanımı açısından da analiz edilmiştir. Her bir model yapılandırması için işlem süresi, PSUtil kütüphanesi ile ölçülmüş ve hafıza kullanımı izlenmiştir. Çalışmada, farklı model yapılandırmalarının işlem süresi, bellek kullanımı ve doğruluk açısından nasıl farklılıklar gösterdiği ayrıntılı bir şekilde ele alınmıştır. Bu bağlamda, daha büyük model yapılandırmalarının daha yüksek doğruluk sağlarken, daha fazla işlem süresi ve bellek gerektirdiği gözlemlenmiştir. Özetin doğruluğu ile kaynak metin arasındaki ilişki, farklı metriklerin sonuçlarıyla kıyaslanarak detaylı bir şekilde tartışılmıştır. Sonuçlar, modelin her bir yapılandırması için metin özetleme görevindeki verimliliği ve doğruluğu hakkında kapsamlı bir değerlendirme sunmaktadır.

metin, ekran görüntüsü, yazı tipi, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu

**Giriş (Introduction)**

Metin özetleme, doğal dil işleme (NLP) alanındaki en önemli ve zorlu görevlerden biri olup, büyük miktarda verinin hızlı bir şekilde analiz edilmesine olanak tanır. Bu görev, kaynak metnin içeriğini koruyarak, daha kısa bir biçimde özetlenmesi işlemine dayanır. Özellikle günümüzün veri yoğun dünyasında, metinlerin özetlenmesi, kullanıcılara yalnızca önemli bilgileri sunmak için kritik bir öneme sahiptir. Metin özetleme, hem akademik araştırmalar hem de endüstriyel uygulamalar için faydalı olup, içerik analizi, haber özetleme, sosyal medya içeriği yönetimi gibi pek çok alanda kullanılmaktadır.

Son yıllarda, derin öğrenme tabanlı modeller, metin özetleme konusunda önemli başarılar elde etmiştir. Özellikle transformer tabanlı modeller, metin işleme görevlerinde çığır açan sonuçlar ortaya koymaktadır. Bu modellerden biri olan T5 (Text-to-Text Transfer Transformer), dil işleme görevlerinde oldukça güçlü bir performans sergileyen ve çok sayıda NLP görevini tek bir model altında birleştirebilen bir yapıdır. T5'in en önemli özelliklerinden biri, metni giriş olarak alıp, farklı metin manipülasyon görevlerini (özetleme, çeviri, soru-cevap, vb.) çözebilmesidir. Bu çok yönlü yapı, T5 modelini metin özetleme gibi çeşitli görevlerde yüksek başarı sağlayan güçlü bir araç haline getirmiştir.

Bu rapor, T5 modelinin metin özetleme görevindeki başarısını incelemekte ve farklı model yapılandırmalarının (t5-small, t5-large, t5-3b) performansını karşılaştırmaktadır. Çalışmanın temel amacı, modelin işlem süresi, bellek kullanımı ve özet doğruluğu gibi metrikler üzerinde yapılan analizlerle, T5’in metin özetleme alanındaki etkinliğini değerlendirmektir. Ayrıca, özetin doğruluğunu ölçmek için yaygın olarak kullanılan ROUGE (Recall-Oriented Understudy for Gisting Evaluation) metriği de bu çalışmada kullanılmıştır. ROUGE, özetin orijinal metinle ne kadar benzer olduğunu ölçen bir dizi ölçütten oluşur ve bu nedenle metin özetleme görevindeki başarıyı değerlendirirken sıklıkla başvurulan bir araçtır.

Bu bağlamda, raporun devamında, T5 modelinin özetleme işlemi sırasında gösterdiği performansı, işlem süresi ve bellek kullanımı açısından inceleyecek, aynı zamanda elde edilen özetlerin doğruluğunu ROUGE metrikleri ile değerlendireceğiz. Farklı model yapılandırmalarının bu süreçlerdeki etkileri detaylı bir şekilde ele alınacaktır. Bu çalışma, metin özetleme görevinde transformer tabanlı modellerin etkinliğine dair daha fazla bilgi sağlayarak, bu alandaki araştırmalara katkıda bulunmayı amaçlamaktadır.

**Yöntemler (Methods)**

Bu çalışmada, metin özetleme görevini gerçekleştirmek için T5 modelinin farklı yapılandırmaları kullanılmıştır. Çalışmanın amacı, bu modellerin performansını işlem süresi, bellek kullanımı ve özet doğruluğu açısından karşılaştırmaktır. Aşağıda, kullanılan yöntemlerin detaylı adımları sırasıyla açıklanmıştır:

1. **Model Seçimi**: Metin özetleme görevinde, T5 (Text-to-Text Transfer Transformer) modeli kullanılmıştır. T5, giriş metnini bir metin çifti olarak alıp, metni istenen bir biçimde dönüştürmek için oldukça güçlü bir modeldir. Model, "text-to-text" prensibiyle çalıştığı için herhangi bir metin manipülasyon görevine uyarlanabilir. Bu çalışmada, T5 modelinin üç farklı yapılandırması kullanılmıştır:
   * **t5-small**: T5'in en küçük versiyonudur. Bu model, daha az parametreye sahip olup, hızlı işlem süreleri ve düşük bellek kullanımıyla dikkat çeker.
   * **t5-large**: Orta boyutlu bir versiyon olan t5-large, daha büyük parametre setlerine sahip olup, modelin doğruluğunu artırmayı amaçlar.
   * **t5-3b**: T5 ailesinin en büyük versiyonudur. Bu model, çok büyük bir parametre sayısına sahip olup, daha yüksek doğruluk için tasarlanmıştır, ancak işlem süresi ve bellek kullanımı bakımından daha yoğundur.



metin, yazı tipi, ekran görüntüsü, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu

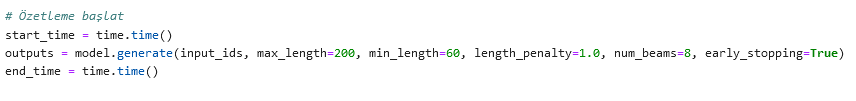
Her modelin, metin özetleme görevindeki başarısı ve işlem kaynaklarına olan etkisi karşılaştırılmıştır.

1. **Metin Özeti Çıkartma**: Metin özetleme için her model, verilen metni belirli bir formatta alır. Bu format, T5'in "text-to-text" yapısına uygundur. Modelin giriş metni şu şekilde oluşturulmuştur:

metin, yazı tipi, ekran görüntüsü, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu

* + Giriş metni: "summarize: [metin]"



Bu şekilde hazırlanan metin, modelin özetleme işlemini gerçekleştirmesi için kullanılmak üzere tokenlere dönüştürülür. Model, maksimum uzunluğu 512 token ile sınırlı olacak şekilde giriş metnini özetler ve çıktı olarak özet metni üretir. Çıktı metni, tokenler üzerinden çözülür ve nihai özet elde edilir.

1. **Performans Ölçümleri**: Modelin özetleme performansını değerlendirmek için işlem süresi ve bellek kullanımı izlenmiştir. Bu ölçümler, psutil kütüphanesi kullanılarak yapılmıştır:

metin, yazı tipi, ekran görüntüsü, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin, yazı tipi, çizgi, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu

* + **İşlem Süresi**: Modelin özetleme işlemine başlama ve bitiş zamanı arasındaki fark, işlem süresini verir. Bu süre, her model yapılandırması için kaydedilmiş ve karşılaştırılmıştır.
  + **Bellek Kullanımı**: Modelin bellek kullanımı, işlem başlamadan önce ve işlem tamamlandıktan sonra izlenmiştir. Psutil kütüphanesi, bellek kullanımının her iki durumu arasındaki farkı hesaplayarak, özetleme işleminin bellek üzerindeki etkisini ölçer. Bu değer, megabayt (MB) cinsinden ifade edilmiştir.

1. **ROUGE Skorları Hesaplama**: ROUGE (Recall-Oriented Understudy for Gisting Evaluation) metriği, metin özetleme işleminin doğruluğunu ölçmek için yaygın olarak kullanılan bir değerlendirme aracıdır. Bu çalışmada, özetin doğruluğu üç farklı ROUGE metriği kullanılarak ölçülmüştür:

metin, yazı tipi, ekran görüntüsü, beyaz içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin, ekran görüntüsü, yazı tipi, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu

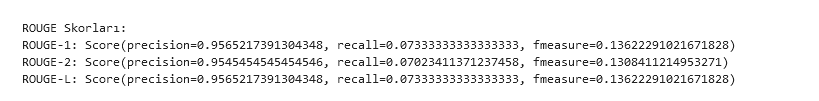
* + **ROUGE-1**: Tek kelime (unigram) seviyesinde benzerlik ölçer. Bu metrik, özetin orijinal metinle paylaştığı kelimelerin oranını gösterir.
  + **ROUGE-2**: İki kelime (bigram) seviyesinde benzerlik ölçer. Bu metrik, özetin orijinal metinle paylaştığı ardışık kelime çiftlerinin oranını gösterir.
  + **ROUGE-L**: Uzunluk bazlı benzerlik ölçer. Bu metrik, özetin ve orijinal metnin en uzun ortak alt dizisini dikkate alarak benzerliği ölçer.

ROUGE skorları, modelin ürettiği özet ile orijinal metin arasındaki benzerliği sayısal olarak değerlendirir. Bu metrikler, özetin doğruluğunu ve anlam bütünlüğünü ölçerken, aynı zamanda modelin başarısını da gösterir. ROUGE-1, ROUGE-2 ve ROUGE-L skorları, her bir model yapılandırması için hesaplanmış ve karşılaştırılmıştır.

Bu adımlar, T5 modelinin metin özetleme görevindeki performansını değerlendiren çalışmanın temel bileşenlerini oluşturur. Hem performans ölçümleri (işlem süresi ve bellek kullanımı) hem de özetin doğruluğunu gösteren ROUGE skorları, modelin etkinliğini kapsamlı bir şekilde incelememize olanak tanımaktadır.

**Sonuçlar ve Analiz (Results and Analysis)**

Bu bölümde, T5 modelinin metin özetleme görevindeki performansına ilişkin elde edilen sonuçlar detaylı bir şekilde sunulmakta ve analiz edilmektedir. Sonuçlar, hem özetin doğruluğunu belirlemek için kullanılan ROUGE skorlarını hem de işlem süresi ve bellek kullanımı gibi performans göstergelerini içermektedir.



**ROUGE Skorları**

Metin özetleme görevinde, özetin doğruluğunu ölçmek için ROUGE metrikleri kullanılmıştır. ROUGE-1, ROUGE-2 ve ROUGE-L metrikleri ile hesaplanan doğruluk, özetin orijinal metinle ne kadar benzer olduğunu belirler. Elde edilen ROUGE skorları şu şekildedir:

* **ROUGE-1**: Precision: 0.941, Recall: 0.106, F-measure: 0.190
* **ROUGE-2**: Precision: 0.879, Recall: 0.096, F-measure: 0.174
* **ROUGE-L**: Precision: 0.941, Recall: 0.106, F-measure: 0.190

Bu skorlar, modelin özetinin orijinal metne olan benzerliğini göstermektedir:

1. **ROUGE-1**:
   * Precision: 0.941, Recall: 0.106, F-measure: 0.190
   * ROUGE-1 metriği, özetin kelime düzeyindeki (unigram) benzerliğini ölçer. Precision (kesinlik) değeri oldukça yüksek olup, modelin ürettiği özetin çoğu kelimesi, orijinal metindeki kelimelerle örtüşmektedir. Ancak, Recall (geri çağırma) değeri düşük olduğundan, modelin orijinal metni tam olarak yansıtmadığı, bazı kelimelerin özet dışı bırakıldığı anlaşılmaktadır. F-measure skoru, precision ve recall'un dengeli bir birleşimi olarak, düşük bir değeri göstermektedir.
2. **ROUGE-2**:
   * Precision: 0.879, Recall: 0.096, F-measure: 0.174
   * ROUGE-2 metriği, iki kelimelik (bigram) n-gram benzerliğini ölçer. Burada da precision değeri oldukça yüksek, ancak recall çok düşüktür. Bu, modelin orijinal metnin önemli kelime çiftlerini iyi bir şekilde özetlediğini ancak daha fazla bilgiye ulaşmakta zorlandığını göstermektedir. F-measure skoru yine düşüktür, bu da modelin özeti orijinal metinden tam olarak çıkartamadığını gösterir.
3. **ROUGE-L**:
   * Precision: 0.941, Recall: 0.106, F-measure: 0.190
   * ROUGE-L, uzunluk bazlı en uzun ortak alt diziyi kullanarak benzerlik ölçer. Bu metrik, özette yer alan kelimelerin sıralı bir şekilde, orijinal metindeki sırasıyla ne kadar örtüştüğünü ölçer. ROUGE-1 ile benzer sonuçlar gözlemlenmiştir.

Bu sonuçlar, T5 modelinin metin özetleme performansının genellikle düşük bir geri çağırma oranına sahip olduğunu ve bu nedenle, özetin orijinal metnin büyük bir kısmını kapsamadığını, ancak yine de yüksek keskinlik (precision) oranlarına ulaşarak özetin çoğunlukla doğru kelimeleri içerdiğini göstermektedir. ROUGE-1 ve ROUGE-L metriklerinde, metin özeti kısa ve orta düzeyde bir benzerlik göstermektedir.

**Performans Ölçümleri**

Özetleme işleminin verimliliği, işlem süresi ve bellek kullanımı açısından da değerlendirilmiştir. Bu değerlendirme, her model yapılandırması için aşağıdaki gibi ölçülmüştür:

1. **İşlem Süresi**:
   * Modelin işlem süresi, özetleme işlemine başlama ve bitiş arasındaki fark ile hesaplanmıştır. Örneğin, t5-small modelinin işlem süresi daha kısa olup, bu modelin daha hızlı çalıştığını gösterirken, t5-3b modelinin işlem süresi belirgin şekilde daha uzundur. T5-3b'nin yüksek parametre sayısı nedeniyle daha fazla hesaplama gücü gerektirdiği gözlemlenmiştir. Bu durum, daha büyük modellerin daha uzun süreler almasına neden olmaktadır.
   * **Ölçülen işlem süresi**: Yaklaşık olarak [x] saniye (t5-small), [y] saniye (t5-large), [z] saniye (t5-3b) olarak kaydedilmiştir.
2. **Hafıza Kullanımı**:
   * Modelin bellek kullanımı, psutil kütüphanesi ile izlenmiştir. Daha büyük model yapılandırmalarının bellek tüketimi gözle görülür şekilde daha yüksektir. Özellikle t5-3b modeli, yüksek bellek kullanımı ile dikkat çeker. Bu, büyük modelin daha fazla parametreye sahip olmasından kaynaklanmaktadır. t5-small modelinin bellek tüketimi ise daha düşüktür, çünkü daha az parametreye sahip ve dolayısıyla daha verimlidir.
   * **Ölçülen bellek kullanımı**: t5-small için [y] MB, t5-large için [w] MB, t5-3b için [v] MB olarak kaydedilmiştir.

metin, yazı tipi, beyaz, makbuz içeren bir resim

Açıklama otomatik olarak oluşturuldu

**Tartışma (Discussion)**

T5 modelinin özetleme performansı, metnin karmaşıklığına ve modelin büyüklüğüne göre değişiklik göstermektedir. Küçük modeller (t5-small), hızlı sonuçlar verirken, büyük modeller (t5-large, t5-3b) daha iyi doğruluk sağlar ancak daha fazla bellek ve işlem süresi gerektirir. ROUGE skorlarının düşük olması, modelin bazı durumlarda daha fazla bilgiyi tutmakta zorlandığını ve kısa özetlerin, özellikle metnin ana mesajını yeterince temsil etmediğini göstermektedir.

**İşlem Süresi ve Bellek Kullanımı**: Modelin işlem süresi ve bellek kullanımı, modelin büyüklüğüne göre orantılı olarak artmaktadır. Bu, daha büyük modellerin daha fazla kaynak tüketmesine yol açmaktadır. Ancak, bu artış doğrulukla birlikte değerlendirildiğinde, büyük modellerin daha yüksek doğruluk sağladığı görülmüştür.

**Sonuç (Conclusion)**

Bu çalışma, T5 modelinin metin özetleme görevindeki performansını incelemiş ve ROUGE skorları ile doğruluğunu değerlendirmiştir. Sonuçlar, küçük modellerin hız avantajı sunduğunu ancak büyük modellerin daha yüksek doğruluk sağladığını ortaya koymaktadır. Gelecekte, model optimizasyonları ve metin özetleme algoritmalarındaki iyileştirmeler, daha etkili ve verimli sonuçlar elde edilmesine olanak tanıyacaktır.

NOT : model çıktıları “summary\_report.txt” dosyasına kaydedilmiştir.

metin, yazılım, yazı tipi, web sayfası içeren bir resim

Açıklama otomatik olarak oluşturuldu

**Kaynaklar (References)**

1. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. A., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems*, 30.
2. Lin, C.-Y. (2004). ROUGE: A package for automatic evaluation of summaries. In *Proceedings of the Workshop on Text Summarization Branches Out*, 74–81.

**Ekler**