

## TRANSORMATÖR TABANLI MODELLER

Yoğun hesaplamalar içeren, yüksek boyutları ve zaman öncelikli işlerde transformatör tabanlı modeller, temel olarak kendini yineleyen sinir ağlarına göre en gelişmiş performansı araştırmacılara sunuyor.

Başlangıçta dil görevleri için tasarlanan Transformatör mimarisi; görüntü oluşturma, ses, müzik uygulamalarıyla Üretken Yapay Zekâ alanında önemli bir bileşen haline geldi.

Derin sinir ağı mimarileri arasında tekrarlayan sinir ağları (RNN'ler, Elman,1990) dil modelleme uygulamalarında özellikle kullanılmaktadır. Kendini yineleyen sinir ağları yapısı gereği sıralı veriler üzerinde daha iyi performans vermektedir. Fakat daha uzun girdi dizilerinde geri besleme sürecinde önceki kelimeleri unutma eğilimi söz konusudur. Yaygın olarak kaybolan gradyan problemi (Hochreiter, 1991) olarak adlandırılan bu olgu uzun girdi dizilerinde performansı düşürüyor.

Uzun kısa süreli bellek modelleri (LSTM; Hochreiter & Schmidhuber, 1997; Jozefowicz ve diğerleri, 2015 ) tekrarlayan sinir ağı mimarisini genişleterek uzun mesafeli bağımlılık sorununu , yani kelimeler olmasa bile kelimeler arasındaki ilişkileri öğrenmeyi çözer. LSTM'ler, modelin hangi bilgilerin hatırlanması veya unutulması gerektiğini öğrenmesine yardımcı olan durum vektörlerini (ağın önceki adımdaki çıktısı) özel bir yapıdan geçirerek çalışır.

Her ne kadar LSTM'ler, uzun mesafeli bağımlılıklarda geleneksel yinelenen sinir ağlarından daha iyi performans gösterse de, özellikle uzun diziler verildiğinde kaybolan gradyanlardan muzdariptirler ve büyük miktarda donanım kaynağı gerektirme eğilimi gösterirler, bu da çoğu araştırmacının daha büyük modelleri eğitme olanağını engeller.

Transformatör model mimarisi, doğal dil gibi sıralı verilere uygulandığında çok sayıda avantaja sahiptir. İlk olarak, sıralı veriler transformatör modelleri tarafından paralel olarak işlenebilir ve bu şekilde model eğitmek için gerekli kaynaklar azaltılarak avantaj sağlanır. Sıralı bilgi (kelimelerin sırası), konumsal kodlama adı verilen bir işlemle korunur. Her kelimeyi amaçlanan sıra konumuna göre cümleye taşır. Ve böylece daha büyük ve daha yetkin dil modelleri eğitilebilmektedir. İkincisi ve tasarım açısından merkezi öneme sahip olan transformatör modelleri, öz-dikkat olarak adlandırılan bir mekanizma aracılığıyla öğrenir.

Transformatör mimarisi 'Kodlayıcı' ve 'Kod Çözücü' olmak üzere ikiye ayrılabilir. Transformatör mimarisi, özel ihtiyaçlara dayalı olarak farklı varyasyonları içerecek şekilde gelişmiştir.

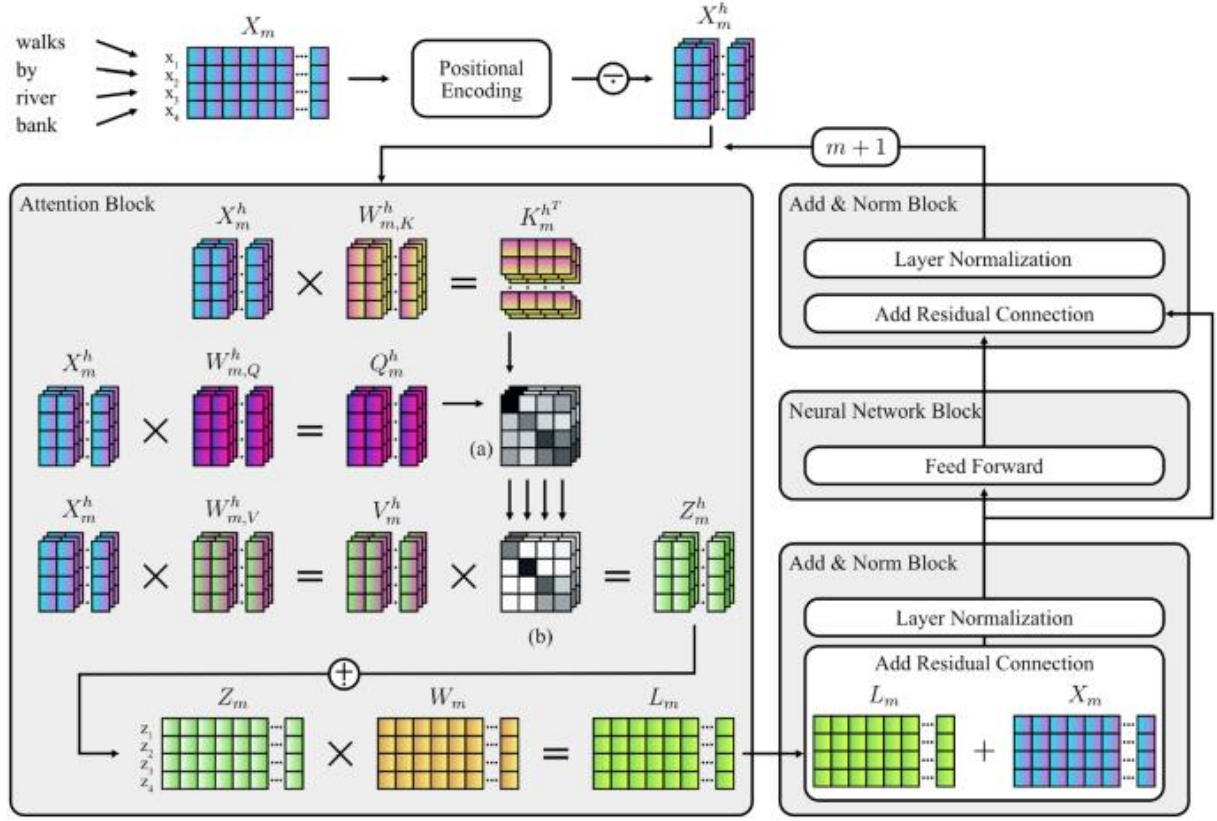
1. Kodlayıcı Ön Eğitimi: Bu modeller tam cümleleri veya pasajları anlamaya odaklanır. Ön eğitim sırasında kodlayıcı, giriş cümlesindeki maskelenmiş belirteçleri yeniden oluşturmak için kullanılır. Bu, modelin genel bağlamı anlamayı öğrenmesine yardımcı olur. Bu tür modeller, metin sınıflandırma, gereklilik ve çıkarımsal soru yanıtlama gibi görevler için kullanışlıdır.

2. Kod Çözücü Ön Eğitimi: Kod çözücü modelleri, önceki belirteç dizisine dayalı olarak bir sonraki belirteci oluşturmak üzere eğitilir. Otomatik gerileyen dil modelleri olarak bilinirler. Kod çözücüdeki öz-dikkat katmanları, yalnızca cümlede belirli bir belirteçten önceki belirteçlere erişebilir. Bu modeller, metin oluşturmayı içeren görevler için idealdir.

3. Transformatör (Kodlayıcı-Kod Çözücü) Ön Eğitimi: Bu varyasyon hem kodlayıcı hem de kod çözücü bileşenlerini birleştirir. Kodlayıcının öz-dikkat katmanları tüm girdi belirteçlerine erişebilirken, kod çözücünün öz-dikkat katmanları yalnızca belirli bir belirteçten önceki belirteçlere erişebilir. Bu mimari, kod çözücünün kodlayıcı tarafından öğrenilen temsilleri kullanmasını sağlar. Kodlayıcı-kod çözücü modelleri özetleme, çeviri veya üretken soru yanıtlama gibi görevler için çok uygundur.

Her ne kadar önceden eğitilmiş dönüştürücü modelleri oldukça tutarlı metin gövdeleri oluşturabilse de çoğu zaman dilsel yeteneklerinin belirli uygulama alanları için özelleştirilmesi arzu edilir. İlgili bir görev ailesini çözmek için önceden elde edilen bilgiyi uygulama süreci, transfer öğrenme olarak adlandırılır ve özellikle kıt eğitim verilerine sahip uygulamalar için güçlüdür (Zhuang ve diğerleri, 2020). Temel varsayım, sinir ağlarının, ilgili görevlerin geniş bir ailesi için iyi düşük seviyeli özellikler olan, erken katmanlardaki nispeten evrensel temsilleri öğrendiğidir. Bu düşük seviyeli özelliklerin genel doğası, genel bir modelden özel modeller üretmek için gereken eğitim süresini veya veri miktarını azaltarak bunları ilgili görevler için yeniden kullanmanın mümkün olması gerektiğini göstermektedir.

Şekil 1, transformatör mimarisinin merkezi yönlerinin şematik bir tasvirini göstermektedir.



Tipik olarak çift yönlü ve tek yönlü transformatör modelleri arasında ayrım yapılır. Çift yönlü modeller, mevcut hedeften hem önce gelen hem de başarılı olan jetonları kullanarak her bir jetonu sırayla tahmin etmeye çalışır.

Belirteçler, belirli bir sözcük dağarcığında yararlı bir anlamsal birim olarak gruplandırılmış karakter dizileridir (örneğin sözcükler, heceler, önekler, noktalama işaretleri, vb.; Manning ve diğerleri, 2008). Bu, bu tür modelleri ikili metin sınıflandırması veya makine çevirisi gibi görevlere uygun hale getirir (Camacho-Collados ve Pilehvar, 2018); González-Carvajal ve Garrido-Merchán, 2021). Bununla birlikte, tek yönlü modeller, belirteç tahminlerini bir sırayla yalnızca önceki sözcükler kümesine dayandırıyor ve bu da onları otoregresif hale getiriyor. Bu nedenle bazen nedensel dönüştürücü modeller olarak anılırlar ve metin oluşturma alanındaki çeşitli uygulamalarda son derece yararlı olduklarını kanıtlamışlardır.

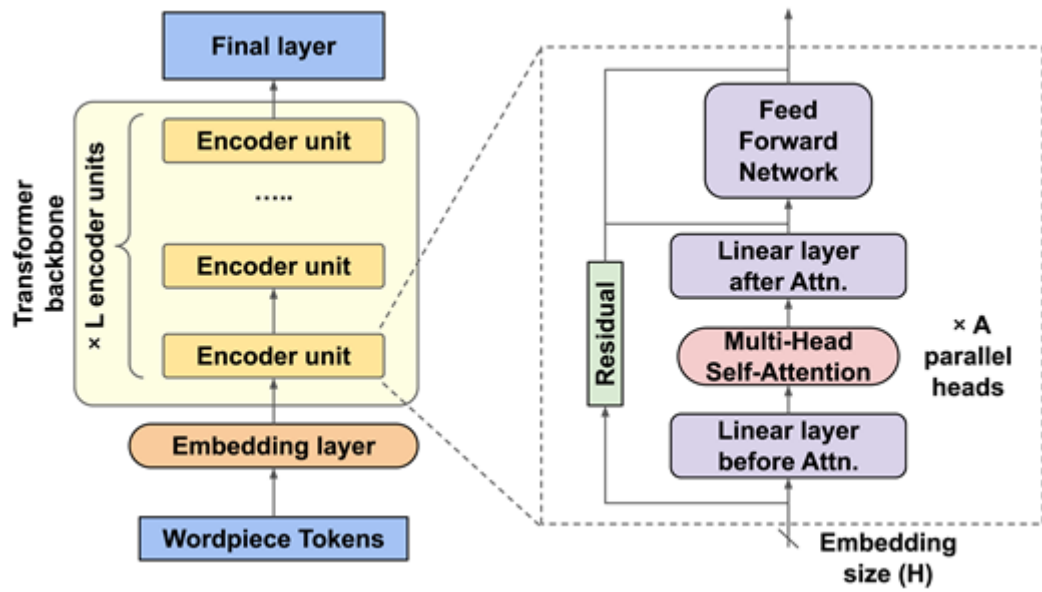
İki yönlü ve tek yönlü transformatör modellerinin dikkate değer sürümleri arasında Transformatörlerden Çift Yönlü Kodlayıcı Temsilleri (BERT; Devlin ve diğerleri, 2018) ve Üretken Önceden Eğitimli Transformatör (GPT; Radford ve diğerleri, 2018 ) yer alır.

BERT mimarisi, aşağıdaki üç hiper parametre kullanılarak belirlenebilir: kodlayıcı birimlerinin sayısı (L ), yerleştirme vektörünün boyutu ( H ) ve her bir

kişisel dikkat katmanındaki dikkat kafalarının sayısı (  $A$  ).  $L$  ve  $H$  modelin derinliğini ve genişliğini belirlerken, her kodlayıcının odaklanabileceği bağlamsal ilişkilerin sayısını etkileyen dahili bir hiper parametredir. BERT'in yazarları önceden eğitilmiş iki model sağladı:

- BERT BAZI (  $L = 12$ ;  $H = 768$ ;  $A = 12$ );
- BERT BÜYÜK (  $L = 24$ ;  $H = 1024$ ;  $A = 16$ ).

Şekil 2' de BERT modelinin akış şeması gösterilmektedir.



Şekil 2

## SONUÇ

Transformatör makalesinin yayınlanmasından bu yana, popüler modeller Bert ve GPT, kodlayıcı veya kod çözücü bileşenlerini kullanarak orijinal mimarının özelliklerini benimsemiştir. Bu modeller arasındaki temel benzerlik, kendi kendine dikkat mekanizmalarını ve ileri beslemeli katmanları içeren katman mimarisinde yatmaktadır. Transformatör' de, her girdi belirteci, girdi dizisindeki diğer tüm belirteçlerle doğrudan bağımlılıklarını korurken, katmanlar arasında kendi yolundan geçer. Bu benzersiz özellik, bağlamsal belirteç temsillerinin paralel ve verimli bir şekilde hesaplanmasına izin verir; bu, RNN'ler gibi sıralı modellerle mümkün olmayan bir yetenektir.

Yapay zekadaki en ünlü trafo modelleri arasında BERT (Transformatör' den Çift Yönlü Kodlayıcı Temsilleri), GPT (Üretken Önceden Eğitilmiş Transformatör) ve T5 (Metinden Metne Aktarım Transformatörü) bulunur. Bu modeller, çeşitli doğal

dil işleme görevlerinde dikkate değer sonuçlar elde etti ve yapay zekâ araştırma topluluğunda önemli bir popülerlik kazandı.

## KAYNAKÇA

[https://direct.mit.edu/tac1/article/doi/10.1162/tac1\\_a\\_00413/107387/Compressing-Large-Scale-Transformer-Based-Models-A](https://direct.mit.edu/tac1/article/doi/10.1162/tac1_a_00413/107387/Compressing-Large-Scale-Transformer-Based-Models-A)

<https://mpost.io/tr/top-30-transformer-models-in-ai-what-they-are-and-how-they-work/>

[https://link.springer.com/article/10.1007/s11336-021-09823-9\]](https://link.springer.com/article/10.1007/s11336-021-09823-9)