

**ÜSKÜDAR ÜNİVERSİTESİ**

Fen Bilimleri Enstitisü

**2023-2024 Bahar Dönemi**

Doğal Dil İşleme

(YZM509/1)

Vize Araştırma Ödevi

Konu : Makine Çevirisi ve Görüntü Altyazılama

**Danışman Öğretim Elemanı**

 Dr. Öğr. Üyesi FATİH ŞAHİN

**Öğrenci Adı-Soyadı:** Emre AKBULUT

**Öğrenci Numarası :** 234329051

**Öğrenci Bölüm Adı:** Yapay Zeka Mühendisliği(Yüksek Lisans)

**MODERN YAKLAŞIMLARLA MAKİNE ÇEVİRİSİ**

**Öz**

Makine çevirisi, farklı diller arasında anlamlı metin çevirisi yapabilen bilgisayar destekli teknolojilerin genel adıdır. Son yıllarda, özellikle yapay zeka ve derin öğrenme tekniklerinin ilerlemesi ile bu alanda önemli gelişmeler kaydedilmiştir. Modern makine çevirisi yaklaşımları, dil modellerini ve çeviri kalitesini büyük ölçüde iyileştirmiştir. Bu araştırma, Transformer gibi nöral ağ tabanlı modellerin, dil işleme ve makine çevirisi alanındaki en son gelişmeleri ve bu modellerin çeviri sürecinde nasıl kullanıldığını ele almaktadır. Yazının ana odak noktası, çeviri hızı ve doğruluğunu artırma potansiyeline sahip yeni nesil teknolojilerdir. Özellikle Transformer modeli, paralel veri işleme kabiliyeti ve daha efektif öğrenme dinamikleri ile dikkat çeker. Bu modelin makine çevirisinde nasıl bir devrim yarattığı, çeviri süreçlerindeki verimlilik artışı ve dil engellerini nasıl daha etkin aştığı üzerinde urulmaktadır.

Ayrıca makine çevirisi yapabilien bir uygulamanın, moden mimarilerle nasıl hayata geçirileceği, kapsamlı bir şekilde ele alınmıştır. Bununla beraber uygulamada çeviriye gelecek verinin de İmage Captioning kullanılarak başka bir uygulamadan gelmesi de işleri biraz daha karmaşık hale getirmiştir. Gerekli olan ana yol haritasının ve bilinmesi gereken teknolojilerin temel manada açıklandığı br ararştırma yazısı oluşturulmaya çalışılmıştır.

**Giriş**

Günümüzde, küresel bir köy haline gelen dünyamızda iletişim, sınırları aşan bir önem kazanmıştır. Farklı dillerde konuşan insanlar arasındaki iletişim engelini aşmanın anahtarı ise makine çevirisidir. Teknolojinin hızla ilerlemesiyle birlikte, makine çevirisi, dil bariyerlerini aşmak ve kültürlerarası anlayışı teşvik etmek için vazgeçilmez bir araç haline gelmiştir. Bu süreçte, yapay zeka ve öğrenme algoritmalarının etkisiyle, makine çevirisinin doğruluğu ve etkinliği önemli ölçüde artmıştır. Ancak, her ne kadar büyük gelişmeler kaydedilse de, makine çevirisi hâlâ çeşitli zorluklarla karşı karşıyadır. Bu makalede, makine çevirisinin kısa bir tarihçesi, teknolojik gelişmeleri, özellikle mevcut durumu ve karşılaşılan zorluklar ele alınacak, Böylece, dil engellerinin üstesinden gelinerek daha anlaşılır ve erişilebilir bir dünya hedefine nasıl yaklaşıldığına dair bir çerçeve sunulmuş olacak. Ayrıca ileri aşamalarda geliştirecek olduğumuz bir uygulamanın ana hatları çizilerek neler yapmayı planladığımızdan bahsedilecektir. Bu bağlamda projemizde direk bir makine çevirisi yapılması yerine “Görüntü Altyazılama” ile alakalı da bir geliştirme yaparak, buradan alacağımız çıktıyı oluşturduğumuz makine çevirisi modeline göndereceğiz; Bu şekilde yine **“Doğal Dil İşlemenin(NLP / DDİ)”** bir alt kolu olan **“Info Extraction(Bilgi Çıkarımı)”** konusuna da kısaca değinerek gelişimizmize katkı sağlamış olmayı beklemekteyiz. Fakat, yazımızın ana konusunu teşkil eden konu NLP ve **“Makine Çevirisi”** olduğu için, **“Görüntü İşleme”** konusunun detaylarına bu yazımızda girilmeyecektir.

**LİTERATÜR TARAMASI**

**Neural Machine Translation: A Review of Methods, Resources, and Tools** - Bu makale, NMT'nin çerçevelerini, zorluklarını ve bu teknolojinin gelecekteki yönlerini tartışan kapsamlı bir genel bakış sunmaktadır.​ ([ar5iv](https://ar5iv.org/pdf/2012.15515.pdf))​.

**Transforming Machine Translation: A Deep Learning System Reaches News Translation Quality Comparable to Human Professionals** - Nature dergisinde yayınlanan bu makale, makine çevirisinde önemli bir kilometre taşı olan, insan benzeri çeviri kalitesine ulaşan bir sistemi tartışıyor([Nature](https://www.nature.com/articles/s41467-020-18073-9.pdf#:~:text=URL%3A%20https%3A%2F%2Fwww.nature.com%2Farticles%2Fs41467))​.

**Overcoming Language Barriers in Academia: Machine Translation Tools and Their Impact** - Akademik bir dergide yer alan bu makale, makine çevirisinin diller arası akademik iletişimi kolaylaştırmadaki rolünü araştırıyor ​ ([Oxford Academic](https://academic.oup.com/bioscience/article/72/10/988/6653151#:~:text=URL%3A%20https%3A%2F%2Facademic.oup.com%2Fbioscience%2Farticle%2F72%2F10%2F988%2F6653151%0ALoading...%0AVisible%3A%200%25%20))​.

**Fully Quantized Transformer for Machine Translation** - ​ EMNLP 2020'de sunulan bu çalışma, kaliteden ödün vermeden makine çevirisi verimliliğini artırmak için transformatörlerin nicelenmesini araştırıyor([ACL Anthology](https://aclanthology.org/2020.findings-emnlp.1/))​.

**Is Machine Translation a Dim Technology for Its Users? An Eye Tracking Study** - Bu çalışma, göz takibi yoluyla makine çevirisinin kullanıcı etkileşimini araştırıyor ve bunun etkinliği ve geliştirilebilecek alanları hakkında fikir veriyor([Frontiers](https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2023.1076379/full))​.

**In Neural Machine Translation, What Does Transfer Learning Transfer?** - Bu makale, bilginin diller arasında nasıl aktarıldığının daha iyi anlaşılmasına katkıda bulunarak, NMT'de transfer öğreniminin bileşenlerini ve etkinliğini araştırmaktadır([ACL Anthology](https://aclanthology.org/2020.acl-main.688/))​.

**Participatory Research for Low-resourced Machine Translation: A Case Study in African Languages** - ​ Düşük kaynaklı dillere odaklanan bu çalışma, makine çevirisi çözümlerinin geliştirilmesinde topluluk katılımının önemini vurguluyor ([ACL Anthology](https://aclanthology.org/2020.findings-emnlp.195/))​.

**Neural Machine Translation for Low Resource Languages** - ​ Bu araştırma, sınırlı veri kaynaklarına sahip dillerde NMT performansını iyileştirmeye yönelik stratejileri derinlemesine inceleyerek odak kaybı ve geri çeviri potansiyelini ortaya koyuyor([ar5iv](https://ar5iv.org/abs/2304.07869))​.

**How Good Are GPT Models at Machine Translation? A Comprehensive Evaluation** - ​Bu makale, makine çevirisi görevlerini gerçekleştirmede Üretken Önceden Eğitimli Transformatör(GPT) modellerinin yeteneklerini eleştirel bir şekilde değerlendiriyor([ar5iv](https://ar5iv.org/abs/2302.09210)).

**Deep Learning Approaches on Image Captioning: A Review** - ​ Bu inceleme, görüntü altyazısı oluşturmadaki derin öğrenme devrimini ele alıyor ve yöntem karmaşıklığını ve performansını önemli ölçüde geliştiren görme dili ön eğitim tekniklerini vurguluyor ([ar5iv](https://ar5iv.org/abs/2201.12944)).

**A Comprehensive Survey on Image Captioning** - Klasik makine öğrenimi temelli yaklaşımlardan çağdaş derin öğrenme yöntemlerine geçişe odaklanan bu anket, robotikten görsel yardıma kadar çeşitli uygulamaları kapsamaktadır([SpringerOpen Journal of Big Data](https://journalofbigdata.springeropen.com/articles/10.1186/s40537-022-00571-w)).

**Image Caption Generation Using Visual Attention Prediction and Contextual Spatial Relation Extraction** - ​ Dalgacık dönüşümü tabanlı CNN modellerini ve görsel dikkat tahmincilerini kullanan bu çalışma, daha doğru altyazılar için görüntülerden anlamsal özelliklerin çıkarılmasını geliştirir([SpringerOpen Journal of Big Data](https://journalofbigdata.springeropen.com/articles/10.1186/s40537-023-00693-9))​.

**A Systematic Literature Review on Image Captioning** - ​ Bu literatür taraması, yeni bulguları şu ana kadar elde edilen en iyi sonuçlarla karşılaştırmayı amaçlayan resim yazısı oluşturmadaki teknikleri, zorlukları ve son araştırmaları ele almaktadır([MDPI](https://www.mdpi.com/2076-3417/9/10/2024))​.

**Image Captioning Based on Deep Learning Methods: A Survey** – ​ Yazı görüntü açıklamaları oluşturmada nesne algılama ve diziden diziye çerçevelerin rolünü vurgulayarak görüntü altyazısı sorununu bir makine çevirisi merceği aracılığıyla yeniden ele alıyor ([ar5iv](https://ar5iv.org/abs/1905.08110))​.

**Makine Çevirisinin Kökeni Nereden Geliyor?**

İnsanların genel inanışı makine çevirisinin oldukça modern bir buluş olduğuna yönünde olmasına karşın, makine çevirisi teknolojisi on yıllardır kullanılmakta. Makine çevirisi kavramı, bilim adamlarının bilimsel belgelerin ve askeri iletişimlerin çevirisi için güvenilir bir yöntem bulma göreviyle 1940'larda ortaya çıktı. 1954'te Georgetown-IBM Deneyi, düzinelerce cümleyi Rusça'dan İngilizceye çeviren basit, kurala dayalı bir sistemle makine çevirisinin potansiyelini ortaya çıkardı

Kural tabanlı makine çevirisi son birkaç on yılda gelişti. Bununla birlikte, 1980'lere kadar, en gelişmiş sistemler bile dil kurallarının karmaşıklığıyla başa çıkmakta zorlanırken, insan dilbilimcilerden gelen kapsamlı manuel girdi talebi, çeviri motorlarının çoğunu kullanışsız hale getirdi. 1990'lara gelindiğinde kurala dayalı makine çevirisinin yerini istatistiksel makine çevirisi aldı. 2010'larda ise, nöral makine çevirisinin önelenemez yükselişini görmekteyiz

Makine çevirisinde herkese uygun tek bir yaklaşım yoktur. Bugün, içeriği bir dilden diğerine yeniden dönüştürmek istiyorsanız kullanabileceğiniz birkaç seçenek bulunmakta. Hangi rotayı seçeceğiniz gereksinimlerinize bağlıdır. Aşağıda en yaygın bazı çözüm çözümler izah edilmiştir.

**1-Kural Tabanlı Makine Çevirisi (RBMT)**

Sonuç elde etmek için katı kurallara ve programlanmış sözlüklere dayanır. Bu sistem önceden belirlenmiş bilgi, hedef dildeki en iyi kelime veya kelime öbeğine karar vermek için orijinal metni dikkate alır. Bir RBMT uygulaması bu kararı vermek için cümle yapısını, dil kurallarını ve anlamı dikkate alır. Kural tabanlı makine çevirisi motorları etkili olabilse de bazı sınırlamalar bulunmaktadır. Birincisi, eksiksiz bir sözlüğe ihtiyaç duyulurken aynı zamanda kapsamlı dil kurallarının da mevcut olması gerekir. Bu kuralların hem kaynak hem de hedef dildeki cümle yapısına ilişkin ayrıntılı bilgileri ortaya koyması gerekmektedir.

Bu tür çeviri araçlarının bir diğer büyük dezavantajı, dil değişikliklerini hesaba katmak için sürekli manuel giriş gerektirmesidir. Kullandığımız dil zaman içinde gelişir, dolayısıyla eski kurallar çevirilerin güvenilirliğini ve alaka düzeyini etkileyebilir.

**2-İstatistiksel Makine Çevirisi (SMT)**

İstatistiksel makine çevirisi, önceden çevrilmiş metni referans noktası olarak kullanan RBMT sistemlerine daha gelişmiş bir alternatiftir. SMT sistemleri, tek tek kelimeler yerine tüm cümleleri değerlendirmek için insan tarafından oluşturulan bu çevirileri kullanır. Sonuç olarak daha doğal ve doğru çeviriler elde edilir.‍ Ancak istatistiksel makine çevirisinde bundan daha fazlası vardır. SMT sistemleri, insanlar tarafından çevrilen cümleleri analiz etmenin yanı sıra, nispeten güvenilir sonuçlar modellemek ve üretmek için hesaplamalı dil bilimini kullanarak kaynak ve hedef dilleri kapsamlı bir şekilde analiz eder. Google Translate gibi en eski ve yaygın olarak kullanılan makine çevirisi motorlarından bazılarına öncülük eden SMT dir.‍

Bununla birlikte, bu tür makine çevirisi motorunun da sınırlamaları bulunmaktadır. SMT sistemlerinin doğru sonuçlar verebilmesi için eğitilmesi gerekir ve yalnızca bir referans noktası varsa çeviri sağlayabilmektedir.

**3-Sinirsel Makine Çevirileri (NMT)**

Nöral makine çevirisi (NMT), çeviri yapmak için yapay zekadan yararlanan bir yaklaşımdır. Makine öğrenimi, çevirilerini zaman içinde öğrenme ve iyileştirme yeteneği sayesinde NMT'ye benzersiz bir esneklik ve uyarlanabilirlik kazandırır.

Bu nedenle, NMT şu anda makine çevirisi, NLP ve diğer metin teknolojilerinde lider teknolojidir. Karmaşık cümleler ve ifadelerde bile RBMT ve SMT'den son derece daha doğru sonuçlar verir. Hepsinden iyisi, nüansları yakalayan çıktılar üretebilir ve insan gibi görünmesini sağlar. Ancak bazı dezavantajlar devam etmektedir. Bir nöral makine çevirisi platformunun başarısı, eğitim verilerine bağlıdır. Eğer veri kötüyse, çeviriler de o kadar doğru olmayacaktır. Ayrıca, NMT hesaplama açısından yoğundur ve bakımı pahalı olabilen kapsamlı donanım ve kaynaklar gerektirir.

Sınırlamalarına rağmen, NMT en iyi makine çevirmeni olmaya devam etmektedir ve araştırmacılar zorluklarını aktif olarak ele aldıkça daha fazla gelişme için umut vaat etmektedir.

**4-Hibrit Makine Çevirileri (HMT)**

Hibrit makine çevirisi (HMT) yazılımı, yukarıda açıklanan diğer çeviri metodolojilerinden iki veya daha fazlasını birleştirir. Amaç, zayıflıklarının üstesinden gelirken her birinin güçlü yönlerinden yararlanmaktır. Örneğin, hibrit bir sistem birincil çevirisi için bir sinir ağı kullanabilir ve yalnızca belirli alanlar veya dil nüansları için kural tabanlı bir yaklaşım kullanabilir. Alternatif olarak, öncelikli olarak kural tabanlı bir sistem, bu kuralları otomatik olarak uyarlamak ve değiştirmek için yapay zekayı kullanabilir.

HMT birden fazla yaklaşım kullandığından, doğruluk büyük ölçüde geliştirilebilir. Ayrıca, sınırlamalar ve özel durumlar kolayca ele alınabilir. Asıl sorun, hibrit makine çevirmenlerinin geliştirilmesi ve bakımının daha karmaşık olabilmesidir. Birden fazla sistemi entegre etmek, sürdürmek ve ince ayar yapmak bir programlama kabusu olabilir.

Ayrıca, HMT sistemi yapay zeka kullanıyorsa, NMT yaklaşımından daha da yoğun kaynak gerektirecektir.

**Makine Çevirisi Konusunda Modern Yaklaşımlar ve Bağlı Teknolojiler**

Yukarıda bahsettiğimiz makine çevirisi yöntemlerinden sadece “Nöral Makine Çevirisi (NMT)” bu araştırmada detaylı bir şekilde incelenecektir. Araştrma bölümü tamamlandıktan sonra projenin uygulama aşamasına geçildiğinde bahsi geçen bu teknolojilerin bazılarından da istifade edilecektir. Bu noktada nöral makine çevirisi teknolojilerin detaylarına inmeden önce **“Yapay Zeka(Artificial Intelligence)”**  onun alt dallarından biri olan **“Doğal Dil İşleme(Natural Language Processing)”** ayrıca yapay zeka teknolojilerinde kullandığımız **“Yapay Sinir Ağları(Artificial Neural Network)”** ve **“Derin Öğrenme(Deep Learning)”** gibi kavramlardan bahsetmemiz araştırdığımız konunun bağlamını doğru zemine yerleştirmek ve konu bütünlüğünü korumak açısından fayda sağlayacaktır.

Konuyu belirlerken öğrenme dürtüsünü tetiklemesi etmesi açısından(Chalange) araştırma alanına eklenen “**Görüntü Altyazılama”** teknolojisine ise ilgili bölüm geldiği zaman gerekli bilgiler verilip, minik bir uygulama demosu yapılarak, konu açıklanacaktır ve ilerleyen bölümlerde tekrardan değinilmeyecektir.

**Yapay Zeka Kavramı**

Yapay zeka (AI), insan benzeri zeka gösteren makineleri, özellikle bilgisayar sistemlerini tasarlama ve programlama bilimidir. Bu alandaki sistemler, öğrenme, karar verme, problem çözme ve algılama gibi insan zekâsıyla özdeşleştirilen işlevleri yerine getirebilirler. Yapay zeka, birçok alt dala ayrılır ve bu dallar geniş bir yelpazede teknolojik uygulama ve teorik araştırma fırsatları sunar

1. **Makine Öğrenmesi (Machine Learning)**: Veri analizi yoluyla öğrenme yeteneğini geliştiren algoritmaların geliştirilmesini içerir. Bu dal, özellikle büyük veri setlerinden öğrenim yaparak tahminlerde bulunma yeteneğine sahiptir.
2. **Derin Öğrenme (Deep Learning)**: Makine öğrenmesinin bir alt dalı olan ve çok katmanlı yapay sinir ağlarından faydalanan bir yöntemdir. Görüntü ve ses işleme gibi karmaşık görevlerde özellikle etkilidir.
3. **Doğal Dil İşleme (Natural Language Processing - NLP)**: İnsan dilinin anlaşılması ve üretilmesi üzerine çalışır. Bu dal, çeviri, duygu analizi ve metin özetleme gibi uygulamalara olanak tanır.
4. **Robotik**: Fiziksel olarak hareket edebilen ve çeşitli görevleri otomatik olarak yerine getirebilen makinelerin ve yazılımların geliştirilmesidir.
5. **Uzman Sistemler (Expert Systems)**: Belirli bir alanda uzmanlık düzeyinde karar verme yeteneği sunan bilgisayar sistemleridir. Tıp, finans ve hukuk gibi alanlarda destek sağlamak için kullanılır.
6. **Bilişsel Bilgisayar Bilimi (Cognitive Computing)**: İnsan düşünce süreçlerini taklit etmeye çalışan sistemler geliştirmek için bilişsel bilim prensiplerinin uygulanmasıdır.

Yapay zeka, teknoloji ve bilim dünyasında devrim yaratma potansiyeline sahip olup, sürekli olarak gelişmekte olan bir alandır ve bu alt dallar, onun çeşitli uygulama alanlarını ve potansiyelini yansıtmaktadır.

**Yapay Sinir Ağları(Artiffical Neural Networks)**

Yapay zeka, makine öğrenmesi özellikle de derin öğrenmeyi açıkladıktan sonra, geldiğimiz geldiğimiz noktada “Yapay Sinir Ağları” larından söz etme zamanı gelmiş demektir.

Yapay sinir ağları (YSA), insan beyninin bilgi işleme şeklini taklit etmeye çalışan matematiksel modeller ve algoritmalar bütünüdür. İnsan beynindeki nöronların işlevselliğinden esinlenerek geliştirilen bu sistemler, genellikle öğrenme ve hafıza işlevlerini yerine getirebilirler.

**Yapay Sinir Ağlarının Temel Yapısı**

Yapay sinir ağları, genellikle üç ana katmandan oluşur:

1. **Giriş Katmanı (Input Layer)**: Dış dünyadan alınan ham verilerin ağa girip ilk işlendiği katmandır.
2. **Gizli Katmanlar (Hidden Layers)**: Giriş katmanı tarafından alınan verileri işleyen ve çıktı katmanına bilgi aktaran bir veya birden fazla katmandan oluşur. Bu katmanlar, ağın öğrenme ve karmaşık ilişkileri modelleme yeteneğini sağlar.
3. **Çıkış Katmanı (Output Layer)**: Sonuçların elde edildiği, işlenmiş verilerin son kullanıcıya sunulduğu katmandır.

A diagram of a network

Description automatically generated

**Çalışma Prensibi**

Yapay sinir ağları, veriler arasındaki karmaşık ilişkileri ve desenleri tanımak için "öğrenme" sürecinden geçer. Bu süreçte genellikle büyük miktarda veri kullanılır ve ağ, verilen girdilerle beklenen çıktılar arasındaki ilişkiyi kurmaya çalışır. Öğrenme süreci sırasında, ağın her bir nöronundaki ağırlıklar (weights), en iyi sonucu vermek üzere ayarlanır. Bu ayarlama işlemi genellikle geri yayılım (backpropagation) algoritması kullanılarak yapılır.

**Bazı Uygulama Alanları**

Yapay sinir ağları pek çok farklı alanda kullanılmaktadır:

* **Görüntü İşleme**: Otomatik plaka tanıma sistemleri veya yüz tanıma sistemleri gibi.
* **Ses Tanıma**: Akıllı asistanlar veya otomatik çeviri sistemlerinde kullanılır.
* **Tahmin Sistemleri**: Hava durumu tahmini, hisse senedi fiyatları gibi zaman serilerinin tahmin edilmesinde kullanılır.
* **Otomasyon ve Robotik**: Sensör verilerini işleyerek robot hareketlerini kontrol etme.

Yapay sinir ağları, öğrenme yetenekleri sayesinde karmaşık problemlere çözüm üretebilir ve sürekli olarak geliştirilmekte olan bir teknolojidir. Bu alandaki ilerlemeler, hem teorik bilgilerin artması hem de pratik uygulamaların genişlemesi anlamına gelmektedir. Bu teknoloji, bilgisayar bilimi, mühendislik, sağlık bilimleri gibi birçok farklı disiplinde devrim yaratma potansiyeline sahiptir.

**Doğal Dil İşleme (Natural Language Processing - NLP)**

Yukarıda tarifini verdiğimiz bu kavrama derinlemesine bir giriş yapıyoruz.

Doğal Dil İşleme (NLP), yapay zeka içinde dinamik ve geniş bir alan olup, bilgisayarlar ile insanlar arasında doğal dil yoluyla etkileşime odaklanır. NLP'nin amacı, bilgisayarların insan dilini anlamasını, yorumlamasını ve hem anlamlı hem de kullanışlı bir şekilde üretebilmesini sağlamaktır.

**NLP'nin Bazı Temel Alanları:**

1. **Veri Ön İşleme:** Metin verilerini temizlemeyi ve hazırlamayı içerir. Yaygın görevler arasında tokenizasyon (metni kelimelere veya cümlelere ayırma), gereksiz kelimeleri çıkarma (örneğin "ve", "bir" gibi bilgi vermeyen yaygın kelimeler) ve metni normalize etme (örneğin küçük harfe çevirme veya lemmatization yoluyla kelimeleri kök veya sözlük formuna indirgeme) bulunur.
2. **Özellik Çıkarma:** Bag-of-Words, TF-IDF (Terim Frekansı-Ters Belge Frekansı) ve Word2Vec gibi teknikler, metni makine öğrenimi algoritmalarının çalışabileceği bir formata dönüştürmek için kullanılır. Bu modeller, veri seti içindeki kelimelerin önemini veya sıklığını belgeler arasında veya belgeler içindeki görünümlerine göre anlamada yardımcı olur.
3. **Modelleme ve Algoritmalar:** İstatistiksel yöntemler, makine öğrenimi ve giderek daha fazla derin öğrenme kullanılarak, NLP görevleri duygu analizi, makine çevirisi ve adlandırılmış varlık tanıma gibi karmaşık işlemleri gerçekleştirecek şekilde modellenir. Örneğin, büyük metin veya konuşma verilerinden bağlamı anlamak gereken görevlerde kullanılmak üzere RNN'ler (Tekrarlayan Sinir Ağları) ve CNN'ler (Evrimsel Sinir Ağları) gibi derin öğrenme modelleri kullanılır.
4. **Uygulamalar:** NLP, çeşitli gerçek dünya uygulamalarında kullanılmaktadır. Yaygın kullanımlar arasında spam tespiti (NLP, istenmeyen e-postalardaki tipik dil kalıplarını tanır), sohbet robotları ve sanal asistanlar (NLP'yi kullanarak insan sorgularını anlamak ve yanıtlamak) ve Google Translate gibi makine çeviri araçları (bir dildeki metni başka bir dile çevirirken orijinal anlamı, tonu ve niyeti korumaya çalışır) bulunmaktadır.

NLP, on yıllar boyunca basit kural tabanlı algoritmalarından karmaşık sinir ağlarına kadar büyüyerek insan benzeri metinleri anlayabilen ve üretebilen gelişmiş algoritmalar için yol açmıştır. NLP'nin evrimi, sürekli olarak gelişmekte olan hesaplama gücü, büyük veri miktarları ve yapay zeka algoritmalarındaki ilerlemeler tarafından yönlendirilmektedir. Bu sürekli gelişme, insan-makine etkileşimlerini daha da ileriye taşıma ve NLP teknolojilerinin çeşitli sektörlerdeki uygulanabilirliğini genişletme vaadinde bulunmaktadır.

**Image Captioning**

**(Görüntü Altyazılama/Görüntü Betimlemesi/Görüntüden Bilgi Çıkarımı)**

Bu noktada NLP’ nin bir başka alt alanı olan, görüntü altyazılama ya da görüntü betimleme olarak da adlandırabileceğimiz kavramın detaylarına inmemiz gerekiyor.

Görüntü betimlemesi, bilgisayar görüşü ve doğal dil işleme (NLP) tekniklerini birleştirerek görüntülerden betimleyici metin üretmeyi amaçlayan yapay zeka alanında heyecan verici bir sahadır. Bu teknoloji, görüntülerin içeriğini derinlemesine anlamak ve bunları insan dilinde açık bir şekilde ifade etmek için derin öğrenme modellerini kullanır.

**Görüntü Betimleme Unsurları:**

1. **Encoder-Decoder Framework:** Bu yaygın kullanılan çerçeve, genellikle bir Evrişimli Sinir Ağı (CNN) olan bir kodlayıcı ve karşılık gelen metni üreten genellikle bir Tekrarlayan Sinir Ağı (RNN) olan bir çözücü içerir. Kodlayıcı, görüntüyü bir dizi özelliğe dönüştürürken, çözücü bu özellikleri genellikle doğruluğu ve ilgiliği artırmak için Uzun Kısa Süreli Bellek (LSTM) ağları ve dikkat modelleri kullanarak tutarlı cümlelere dönüştürür.
2. **Attention Mechanisms:** Modelin her bir kelimeyi oluştururken görüntünün belirli bölümlerine odaklanmasını sağlayarak önemli bir rol oynar. Bu, insan görüşünün bir sahneyi anlamak için farklı nesnelere nasıl odaklandığını taklit eder.
3. **Training Strategies:** Modeller genellikle görsel öğeleri metin açıklamalarıyla etkili bir şekilde ilişkilendirmeyi öğrenmelerine yardımcı olmak için tanımlayıcı başlıklarla eşleştirilmiş büyük görüntü veri kümelerinde eğitilir. Eğitimde kullanılan popüler veri kümeleri arasında Flickr8k, MS COCO ve daha yakın zamanlarda Yelp veri kümesi gibi gerçek dünya görüntülerini ve kullanıcı tarafından oluşturulan başlıkları içeren çeşitlilik ve zorluklar sunan veri kümeleri bulunmaktadır.
4. **Advanced Models:** Temel LSTM ve CNN modellerinin ötesinde, özellikle dikkat mekanizmalarını görüntü özellikleri ve metin üretimi ile daha ince bir şekilde hizalayan Transformer modelleri gibi daha yeni mimariler, görüntü betimleme için araştırılmıştır. Bu, üretilen başlıkların doğallığını ve ilgililiğini artırır.

Görüntü betimlemenin uygulamaları geniş ve etkileyicidir; görme engelli bireylerin çevrelerini anlamalarına yardımcı olmaktan sosyal medya platformlarında kullanıcı etkileşimlerini otomatik olarak üretilen tanımlayıcı başlıklarla geliştirmeye kadar uzanır. Ayrıca, içerik tabanlı görüntü alma sistemleri gibi daha işlevsel görevlerde de rol oynar.

Genel olarak, görüntü betimleme, görsel veri yorumlaması ve dil üretimi arasında önemli bir kesişim oluşturarak teknolojiler geliştikçe heyecan verici gelişmeler vaat eder.

**Mini Demo**

Burada Python'u ve görüntü altyazıları için transformatör mimarisini kullanan basit bir mini demo örneği oluşturacağız. Resimler için altyazı oluşturabilen bir model oluşturmak için popüler Hugging Face transformatör kütüphanesini torchvision ile birlikte kullanacağız.

from transformers import VisionEncoderDecoderModel, ViTImageProcessor, AutoTokenizer

import torch

from PIL import Image

model = VisionEncoderDecoderModel.from\_pretrained("nlpconnect/vit-gpt2-image-captioning")

feature\_extractor = ViTImageProcessor.from\_pretrained("nlpconnect/vit-gpt2-image-captioning")

tokenizer = AutoTokenizer.from\_pretrained("nlpconnect/vit-gpt2-image-captioning")

device = torch.device("cuda" if torch.cuda.is\_available() else "cpu")

model.to(device)

max\_length = 16

num\_beams = 4

gen\_kwargs = {"max\_length": max\_length, "num\_beams": num\_beams}

def predict\_step(image\_paths):

  images = []

  for image\_path in image\_paths:

    i\_image = Image.open(image\_path)

    if i\_image.mode != "RGB":

      i\_image = i\_image.convert(mode="RGB")

    images.append(i\_image)

  pixel\_values = feature\_extractor(images=images, return\_tensors="pt").pixel\_values

  pixel\_values = pixel\_values.to(device)

  output\_ids = model.generate(pixel\_values, \*\*gen\_kwargs)

  preds = tokenizer.batch\_decode(output\_ids, skip\_special\_tokens=True)

  preds = [pred.strip() for pred in preds]

  return preds

predict\_step(['futbolcu.png']) # ['a man kicking a soccer ball on a field']

Bu script, bir transformatör modeliyle görüntü altyazısının temel bir gösterimini sağlar. Farklı veya daha karmaşık senaryolara göre geliştirmek ve uyarlamak için modeli ayarlamayı, belirli verilerle ince ayar yapmayı veya daha büyük bir uygulama çerçevesine entegre etmeyi hayal edebiliriz. Ayrıca burada önceden eğitilmiş veri setlerini kullanmış olduk. Kendimiz de veri setlerini eğitebilirdik. Burada örnek olarak kullandığım resmi de paylaşıyorum, bu sayede kodun çalışabirliği test edilebilsin.

A group of men playing football

Description automatically generated

“a man kicking a soccer ball on a field”

**Konu ile Alakalı Bilinmesi Gereken DiğerKavramlar ve Teknolojiler**

1. **CNN (Evrimsel Sinir Ağı)**: CNN'ler, görüntü işleme ve bilgisayar görüşünde başta olmak üzere kullanılan derin sinir ağları türüdür. Mekansal verilerdeki desenleri algılama konusunda özellikle iyidirler. Bu ağlardaki evrimsel katmanlar, girişe evrişim işlemi uygular ve sonucu sonraki katmana aktarır. Bu özellikleri, görüntülerdeki yüzleri veya nesneleri tanıma gibi görevlerde yüksek verimlilik sağlar.
2. **RNN (Tekrarlayan Sinir Ağı)**: Tekrarlayan Sinir Ağları, sıralı tahmin problemleriyle çalışmak üzere tasarlanmıştır. RNN'ler, dizinin elemanları üzerinden yinelemeli olarak işlem yapar ve şimdiye kadar gördükleriyle ilgili bilgi içeren bir durum tutar. Bu, dil modelleme ve metin üretimi gibi, dizinin daha önceki kısımlarından gelen bağlamın daha sonraki elemanlarını anlamak veya tahmin etmek için gereken görevlerde kullanışlıdır.
3. **LSTM (Uzun Kısa Süreli Bellek)**: LSTM'ler, uzun vadeli bağımlılıkları öğrenme kapasitesine sahip özel bir RNN türüdür. Standart RNN'lerde ortaya çıkabilecek kaybolan gradyan problemi ile başa çıkmak için geliştirilmiştir. LSTM'ler, bir hücre durumu sürdürür ve hücreye bilgi akışını düzenlemek için kapılar kullanır, bu da uzun girdi dizilerinden bağımlılıkları daha etkili bir şekilde yakalamalarına yardımcı olur.
4. **GRU (Kapılı Tekrarlayan Birim)**: GRU'lar, LSTM'lerle benzer olan fakat daha basit yapıya sahip bir RNN çeşididir. Unutma ve giriş kapılarını tek bir "güncelleme kapısı"nda birleştirir ve hücre durumu ile gizli durumu birleştirir, bu da daha az hesaplama yükü ile çoğu zaman LSTM kadar iyi performans sergileyen bir model sunar.
5. **Seq2Seq (Diziden Diziye)**: Seq2Seq modelleri, bir dildeki cümleleri başka bir dile çevirmek gibi, bir alanın dizilerini başka bir alana dönüştürmek için NLP'de kullanılır. Bu modeller genellikle iki parçadan oluşur: girdi dizisini işleyen ve bilgiyi bir bağlam vektörüne sıkıştıran bir kodlayıcı ve bu vektörü kullanarak çıktı dizisini üreten bir çözücü.
6. **GloVe (Global Vektörler ile Kelime Temsili)**: GloVe, kelime temsillerini elde etmek için gözetimsiz bir öğrenme algoritmasıdır. Eğitim, bir korpus içinde küresel kelime-kelime birliktelik istatistikleri üzerinde gerçekleştirilir ve elde edilen temsiller, kelime vektör uzayının ilginç lineer alt yapılarını gösterir.
7. **Hugging Face:** Başlangıçta 2016 yılında Julien Chaumond, Clément Delangue ve Thomas Wolf tarafından kurulan, yapay zeka topluluğu ve makine öğrenimi platformudur. Hugging Face, özellikle doğal dil işleme (NLP) alanında devrim yaratmayı amaçlayan, Transformers kütüphanesi ile tanınır. Bu platform, veri bilimcilerine, AI uygulayıcılarına ve mühendislere, 20.000'den fazla önceden eğitilmiş model sunarak NLP'yi demokratikleştirmeyi hedefler. Hugging Face'in sunduğu modeller arasında, metin, ses, görüntü ve tablo verileri için çeşitli görevlerde kullanılabilen geniş bir yelpazede yer almaktadır. Bu modeller, metin sınıflandırma, bilgi çıkarımı, soru cevaplama, metin üretimi ve çeviri gibi görevlerde kullanılır​. Platform, kullanıcıların modellerini ve veri kümelerini paylaşmalarına olanak tanıyan bir "Hub" barındırır. Ayrıca, kullanıcıların makine öğrenimi modellerinin demo versiyonlarını oluşturup paylaşmaları için "Spaces" adı verilen web uygulamalarını içerir. Hugging Face, bu altyapıyı sağlayarak, AI topluluğunun işbirliği yapmasını ve yenilikçi projeler geliştirmesini teşvik eder​.Eğer AI ve makine öğrenimi konularında çalışıyorsanız veya bu alanlarda yeniyseniz, Hugging Face platformu, kaynaklar ve araçlar açısından değerli bir başlangıç noktası olabilir.

**Transformers Mimarisi Kullanılarak Makine Çevirisi**

Transformer mimarisi, yapay zeka alanında özellikle makine çevirisi için geliştirilmiş bir derin öğrenme modelidir. 2017 yılında "Attention is All You Need" başlıklı çalışmayla tanıtılan bu mimari, dikkat mekanizmalarını temel alır ve RNN (Tekrarlayan Sinir Ağları) ile CNN'lerin (Evrimsel Sinir Ağları) yerine geçmek için tasarlanmıştır. Transformer'ın temel avantajlarından biri, girdi ve çıktı arasında karmaşık bağımlılıkları modelleyebilmesi ve işlemlerin paralel olarak yapılabilmesi sayesinde eğitim süreçlerinin hızlandırılmasıdır.

**Transformer Mimarisi Nasıl Çalışır?**

Transformer, esas olarak bir kodlayıcı (encoder) ve bir çözücü (decoder) olmak üzere iki ana bileşenden oluşur:

* **Encoder**: Girdi dizisini alır ve bu diziyi sürekli bir temsile dönüştürür. Her bir encoder katmanı, kendisine özgü dikkat mekanizmaları ve tam bağlantılı ağlar içerir. Encoder, girdi dizisindeki her elemanın bağlamını diğer elemanlarla ilişkilendirerek anlamlandırır.
* **Decoder**: Encoder'dan alınan temsilleri kullanarak çıktı dizisini adım adım oluşturur. Her decoder katmanı, encoderdan gelen bilgileri kullanarak çıktı dizisinin sonraki elemanını tahmin etmeye çalışır.

Her iki bileşen de çoklu dikkat başlıkları içerir, bu başlıklar modelin farklı bilgi parçalarına odaklanmasına olanak tanır. Ayrıca, pozisyonel kodlamalar sayesinde model, dizilerdeki sıra bilgisini koruyabilmektedir.

Transformer mimarisi, makine çevirisi başta olmak üzere doğal dil işleme, bilgisayarla görü işleme ve hatta ses işleme gibi birçok farklı alanda kullanılmaktadır. Eğitim sürecini hızlandırması ve uzun mesafeli bağımlılıkları etkili bir şekilde öğrenebilmesi, Transformer'ı son yıllarda popüler hale getirmiştir.

Bu mimari, GPT ve BERT(Gemini) gibi önceden eğitilmiş modellerin geliştirilmesinde de kullanılmaktadır, bu da onu çeşitli dil işleme görevlerinde devrim yaratmasını sağlamıştır​.

A diagram of a process

Description automatically generated

**Bu mimari kullanılarak nasıl makine çevirisi modeli oluşturabiliriz?**

Makine çevirisi için bir Transformer modeli oluşturmak, son yıllarda yapay zeka ve dil işleme alanlarında popüler bir yaklaşım haline gelmiştir. Transformer mimarisi, özellikle dil modellerinde etkileyici sonuçlar elde etmiştir ve bu, onu makine çevirisi gibi karmaşık NLP görevleri için ideal kılar. İşte adım adım bir Transformer tabanlı makine çevirisi modeli oluşturma süreci

**1. Veri Hazırlığı**

Makine çevirisi modeli eğitmek için büyük bir paralel dil veri setine ihtiyacınız olacak. Bu, genellikle kaynak dilde cümleler ve hedef dildeki karşılıkları içeren çiftlerden oluşur. Örneğin, İngilizce'den Türkçe’ye çeviri yapmak istiyorsanız, her iki dilde de benzer cümleleri içeren bir veri seti gereklidir.

**2. Tokenizasyon**

Verileri modeliniz tarafından işlenebilir hale getirmek için tokenizasyon işlemi yapılmalıdır. Tokenizasyon, cümleleri daha küçük parçalar olan token'lara böler. Bu token'lar kelime, kelime kökü veya harfler olabilir. Çoğu durumda, subword tokenizasyonu tercih edilir çünkü bu yöntem bilinmeyen kelimelerle başa çıkma konusunda daha etkilidir.

**3. Model Mimarisi**

Transformer modeli, temelde bir encoder ve bir decoder içerir:

* **Encoder**, girdi cümlesini bir dizi vektöre dönüştürür. Her bir token, model tarafından daha sonra işlenebilmek üzere bir temsile kodlanır.
* **Decoder**, encoder'dan alınan bu temsilleri kullanarak hedef dilde bir çıktı üretir. Her adımda, decoder bir sonraki token'ı tahmin etmek için önceki token'ları ve encoder çıktılarını kullanır.

**4. Attention Mekanizması**

Transformer modellerinin temel özelliklerinden biri, attention mekanizmasıdır. Bu mekanizma, modelin kaynak cümledeki önemli bilgilere "dikkat etmesini" sağlar ve bu bilgileri çeviri sırasında kullanır. Multi-head attention, modelin farklı bilgi türlerine aynı anda dikkat etmesine olanak tanır.

**5. Eğitim**

Modeli eğitmek için genellikle cross-entropy loss fonksiyonu kullanılır. Bu aşamada, gerçek çeviri ile modelin tahminleri arasındaki fark minimize edilmeye çalışılır. Eğitim süreci, genellikle GPU veya TPU gibi yüksek performanslı hesaplama birimlerinde yapılır çünkü büyük veri setleri ve kompleks model yapıları önemli miktarda hesaplama gücü gerektirir.

**6. Değerlendirme ve İyileştirme**

Modelin performansı, genellikle BLEU skoru gibi metriklerle değerlendirilir. Eğitimden sonra modelin daha da iyileştirilmesi için hiperparametre ayarlamaları, daha fazla eğitim verisi eklenmesi veya model mimarisinde değişiklikler yapılabilir.

**Örnek Kod ve Kaynaklar**

Python'da Transformer modeli geliştirmek için TensorFlow veya PyTorch gibi kütüphaneler kullanılabilir. Örneğin, Hugging Face'in “**Transformers”** kütüphanesi, çeşitli ön eğitimli Transformer modelleri sağlar ve bu modeller makine çevirisi gibi görevlerde kolayca kullanılabilir.

Makine çevirisi modeli oluşturmak için ayrıntılı bilgi ve kod örnekleri için Hugging Face dökümantasyonu incelemek ileri aşamada gerçekleştireceğimiz projede kolaylıklar sağlayacaktır.

**Sonuç**

Gerek araştırılacak konuların çokluğu, zorluğu, birbiriyle olan bağlantılarını anlayabilme, gerekse bilgilerin sürekli güncellenip yeni teknolojilere adapte olabilme zorluğu NLP alanında da kendisini göstermektedir. Bütün bu olumsuzluklara karşı daha fazla araştırma yapmak, azimle, yılmadan konuları analamaya çalışarak uygulamalarlar geliştirmek ve bilgilerimizi pekiştirmemiz gerekmektedir.

Teşekkürler Saygılarımla

Emre AKBULUT 12.04.2024

Üsküdar/İstanbul

**Kaynakça**

1. Vaswani, A. et al. (2017). Attention Is All You Need. *NeurIPS*.
2. Brown, T. et al. (2020). Language Models are Few-Shot Learners. *NeurIPS*.
3. Raffel, C. et al. (2020). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *JMLR*.
4. Lewis, M. et al. (2020). BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. *ACL*.
5. Radford, A. et al. (2021). Learning Transferable Visual Models From Natural Language Supervision. *ICML*.
6. Lample, G., & Conneau, A. (2019). Cross-lingual Language Model Pretraining. *NeurIPS*.
7. Xue, L. et al. (2021). MT5: A Massively Multilingual Pre-trained Text-to-Text Transformer. *ACL*.
8. Tang, Y. et al. (2020). Multilingual Translation with Extensible Multilingual Pretraining and Finetuning. *EMNLP*.
9. Fan, A. et al. (2020). Beyond English-Centric Multilingual Machine Translation. *JMLR*.
10. Zhang, J. et al. (2020). Bridging the Gap between Training and Inference for Neural Machine Translation. *WMT*.
11. Kasai, J. et al. (2021). Deep Encoder, Shallow Decoder: Reevaluating Non-autoregressive Machine Translation. *ICLR*.
12. Ott, M. et al. (2020). FairSeq: A Fast, Extensible Toolkit for Sequence Modeling. *NAACL*.
13. Wu, S. et al. (2020). Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. *arXiv*.
14. Kudugunta, S. et al. (2021). Investigating Multilingual NMT Representations at Scale. *EMNLP*.
15. Qian, Y. et al. (2021). Exploring the Limits of Domain Adaptation for Neural Machine Translation. *NAACL*.