



Derin Öğrenme Teknikleri Kullanılarak Türkçe-İngilizce Nöral Makine Çevirisi

Yazılım Mühendisliği Ana Bilim Dalı

Dönem Projesi

Altuğhan Özengi

ORCID 0009-0004-3770-2118

Proje Danışmanı: Prof. Dr. Ayşegül Alaybeyoğlu Soy

Kasım 2023

Derin Öğrenme Teknikleri Kullanılarak Türkçe-İngilizce Nöral Makine Çevirisi

ÖZ

Nöral makine çevirisi son yıllarda yapay zekâ ve doğal dil işleme alanında büyük bir ilerleme kaydeden, bir dilden başka bir dile yüksek kaliteli çeviriler sunma potansiyeline sahip yapay sinir ağları tabanlı bir teknolojidir. Derin öğrenme modellerini kullanan bu sistem, doğru bir şekilde tasarlandığında geleneksel istatistiksel ve kural tabanlı çeviri sistemlerine oranla metinlerin daha doğru ve doğal bir şekilde çevrilmesini sağlamaktadır.

Bu proje çalışmasında, Türkçeden İngilizceye makine çevirisi için derin öğrenme tabanlı bir yöntem olan ve diziden diziye dönüşüm yapmaya olanak sağlayan kodlayıcı-kod çözücü tabanlı mimari benimsenmiştir. Bu mimari içerisinde RNN, LSTM ve GRU olmak üzere üç farklı tekrarlayan yapay sinir ağı modeli uygulanarak her bir modelin makine çevirisindeki etkinliği farklı performans metrikleri kullanılarak değerlendirilerek güçlü ve zayıf yönleri tartışılmıştır. Çalışmanın yapay sinir ağları ile makine çevirisi alanında yürütülen araştırmalara katkısının olacağı düşünülmektedir.

Anahtar Sözcükler: Nöral makine çevirisi, derin öğrenme, yapay sinir ağları, tekrarlayan sinir ağları, RNN, LSTM, GRU, kodlayıcı-kod çözücü mimarisi

Turkish-English Neural Machine Translation Using Deep Learning Techniques

Abstract

Neural machine translation is a neural network-based technology that has made great progress in artificial intelligence and natural language processing in recent years offering high-quality translations from one language to another. When properly designed using deep learning models, this system provides more accurate and natural translations compared to traditional statistical and rule-based translation systems.

In this project, the encoder-decoder architecture, which is a deep learning-based method, was adopted for machine translation from Turkish to English. In this architecture, three different recurrent neural network models, namely RNN, LSTM and GRU, were implemented to evaluate the effectiveness of each model in machine translation using different performance metrics, based on which their strengths and weaknesses were discussed. The study is expected to contribute to the research conducted with artificial neural networks in the field of machine translation.

Keywords: Neural machine translation, deep learning, artificial neural network, recurrent neural network, RNN, LSTM, GRU, encoder-decoder architecture

Teşekkür

Bu çalışmanın gerçekleştirilmesinde emeđi geen saygıdeđer danıřman hocam Prof. Dr. Ayřegöl ALAYBEYOĐLU SOY'a yardımları iin itenlikle teřekkür ederim.

İçindekiler

Öz	i
Abstract	ii
Teşekkür	iii
Şekiller Listesi.....	vi
Tablolar Listesi.....	vii
Kısaltmalar Listesi	viii
1 Giriş	1
1.1 İlgili Çalışmalar	2
1.2 Projenin Amacı ve Kapsamı	3
2 Yapay Sinir Ağları.....	4
2.1 Yapay Sinir Ağlarının Tanımı	4
2.2 Yapay Sinir Ağlarının Tarihçesi	4
2.3 Biyolojik Sinir Ağlarının Yapısı.....	5
2.4 Yapay Sinir Ağlarının Yapısı ve Bileşenleri	7
2.4.1 Girişler	7
2.4.2 Ağırlıklar	7
2.4.3 Toplama Fonksiyonu	8
2.4.4 Aktivasyon Fonksiyonu	8
2.4.5 Çıkış İşlevi.....	8
2.5 Yapay Sinir Ağları ile Nöral Makine Çevirisi	8
2.6 Yapay Sinir Ağı Türleri ve Kullanım Alanları	9
2.6.1 Tekrarlayan Sinir Ağları.....	10

2.6.1.1	Tekrarlayan Sinir Ağları Çalışma Prensibi	10
2.6.1.2	Tekrarlayan Sinir Ağı Mimarileri.....	10
2.6.2	Uzun Kısa Süreli Bellek Ağları (LSTM).....	11
2.6.3	Geçitli Tekrarlayan Birim Ağları (GRU)	12
3	Materyal ve Yöntem	13
3.1	Veri Okuma ve Ön İşleme	14
3.1.1	Veri Seti.....	14
3.1.2	Veri Okuma	14
3.1.3	Veri Ön İşleme.....	14
3.1.3.1	Veri Temizleme	14
3.1.3.2	Tokenizasyon.....	16
3.2	Eğitim ve Test Verisinin Hazırlanması	16
3.3	Modellerin Oluşturulması	16
3.3.1	RNN ile Kodlayıcı-Kod Çözücü Tabanlı Model Oluşturma	17
3.3.2	LSTM ile Kodlayıcı-Kod Çözücü Tabanlı Model Oluşturma.....	18
3.3.3	GRU ile Kodlayıcı-Kod Çözücü Tabanlı Model Oluşturma	19
3.4	Model Performans ve Hatalarının Karşılaştırılması	20
3.5	Model Çeviri Performansı ve BLEU Puanı	21
4	Bulgular ve Tartışma	24
5	Sonuçlar.....	26
	Kaynaklar	27

Şekiller Listesi

Şekil 2.1	Biyolojik bir beyin sinir hücresi.....	6
Şekil 2.2	Yapay sinir hücresi.....	7
Şekil 2.3	Çince'den İngilizce'ye makine çevirisi için RNN tabanlı bir sinir ağı modelinin şematik gösterimi	9
Şekil 2.4	Tekrarlayan sinir ağının yapısı.....	10
Şekil 3.1	Araştırmanın deneysel aşamaları	13
Şekil 3.2	Veri setindeki cümle uzunluklarının dağılımı.....	15
Şekil 3.3	RNN tabanlı kodlayıcı-kod çözücü modeli.....	17
Şekil 3.4	LSTM tabanlı kodlayıcı-kod çözücü modeli.....	18
Şekil 3.5	GRU tabanlı kodlayıcı-kod çözücü modeli.....	19
Şekil 3.6	RNN, LSTM ve GRU tabanlı kodlayıcı-kod çözücü tabanlı modellerin doğruluk grafiği.....	20
Şekil 3.7	RNN, LSTM ve GRU tabanlı kodlayıcı-kod çözücü tabanlı modellerin hata grafiği.....	21

Tablolar Listesi

Tablo 2.1	Ađlar ve uygulama alanları	9
Tablo 3.1	İşlenmemiş veri setinden örnek cümleler	14
Tablo 3.2	İşlenmiş veri setinden örnek cümleler	15
Tablo 3.3	Modeller kullanılarak yapılan çeviri örnekleri	22
Tablo 3.4	Her bir model için n-gram BLEU puanı	24

Kısaltmalar Listesi

ADALINE	Adaptive Linear Elements (Uyarlanabilir Doğrusal Elemanlar)
CGM	Convolutional N-gram Model (Evrışimli N-gram Modeli)
CNN	Convolutional Neural Network (Evrışimli Sinir Ağı)
GRU	Gated Recurrent Unit (Geçitli Tekrarlayan Birim)
LSTM	Long Short-Term Memory (Uzun Kısa Süreli Bellek)
MADALINE	Multiple Adaptive Linear Elements (Çoklu Uyarlanabilir Doğrusal Elemanlar)
NLP	Natural Language Processing (Doğal Dil İşleme)
RNN	Recurrent Neural Network (Tekrarlayan Sinir Ağı)
YSA	Yapay Sinir Ağları

Bölüm 1

Giriş

Globalleşen dünyada kültürler arası iletişimin ön plana çıkması, farklı diller arasındaki bilgi aktarımının daha hızlı ve yüksek kaliteli sistemlere ihtiyaç duymasına yol açmıştır. Bu ihtiyaca cevap vermek amacıyla ortaya çıkan ve doğal dil işlemenin bir alt alanı olan makine çevirisi, başlangıçta kural tabanlı ve istatistiksel yöntemlerle gerçekleştirilirken son yıllarda derin öğrenme yöntemleri ve yapay sinir ağlarının ön plana çıkmasıyla beraber büyük bir dönüşüm geçirmiştir. Yapay sinir ağları kullanılarak gerçekleştirilen ve nöral makine çevirisi olarak bilinen bu teknoloji, farklı dillerdeki yazılı metinlerin daha doğru, hızlı ve doğal bir şekilde çevrilmesini sağlama potansiyeline sahiptir.

Yapay sinir ağları ile gerçekleştirilen doğal dil işleme uygulamalarında ve özellikle makine çevirisinde, cümlede yer alan bir kelimedenden sonra hangi kelimenin gelmesi gerektiği tahmin edildiğinden dizilim biçiminde ele alınan cümlelerdeki kelime bağımlılıklarının korunarak doğru ve tutarlı çeviriler yapılması büyük bir öneme sahiptir [1]. Bu görev için yaygın olarak kullanılan sinir ağlarından biri de RNN, LSTM ve GRU olarak bilinen tekrarlayan yapay sinir ağlarıdır. Tekrarlayan yapay sinir ağları kelimeler arasındaki ilişkileri bir bağlam içinde ele aldıklarından çevirinin tutarlılığı üzerine de doğrudan etki etmektedirler.

Bu proje çalışmasında, bahsi geçen tekrarlayan yapay sinir ağlarından her birinin yapısını incelenerek Türkçeden İngilizceye makine çevirisi üzerindeki etkileri, performansları, sundukları avantaj ve dezavantajlarının kapsamlı bir şekilde analiz edilmesi hedeflemiştir. Elde edilen modellerin etkinliği farklı performans metrikleri kullanılarak değerlendirilerek güçlü ve zayıf yönleri tartışılmış ve olası iyileştirmeler

üzerinde durulmuştur. Bu sayede yapay sinir ağları ile makine çevirisi alanında yürütülen diğer araştırmalara katkısının olması amaçlanmıştır.

1.1 İlgili Çalışmalar

Geçtiğimiz on yılda literatürde tekrarlayan yapay sinir ağları ile nöral makine çevirisi üzerine birçok araştırma ve makale yayınlanmıştır. Bu alandaki öncü çalışmalardan ilki Kalchbrenner & Blunsom [2] tarafından makine çevirisi için sunulan bir encoder-decoder (kodlayıcı-kod çözücü) mimarisi olmuştur. Bu mimarinin kodlayıcı kısmında kaynak dildeki cümleler Evrişimli N-gram Modeli (CGM) kullanılarak vektörlere çevrilmiştir. Kod çözücü kısmında ise ters CGM ve Tekrarlayan Sinir Ağı (RNN) kombinasyonu uygulanarak hedef dile dönüşüm sağlanmıştır. Bu yaklaşım nöral makine çevirisinin başlangıcı olarak kabul görmüştür.

Cho ve diğ. [3] çalışmalarında, iki RNN'den oluşan kodlayıcı-kod çözücü sinir ağı mimarisi kullanılmıştır. Bu mimari kullanılarak yapılan çevirilerde hem anlamsal hem de söz dizimsel yapının iyi bir şekilde korunduğu sonucuna varılmıştır. Buna rağmen Cho ve diğ. [4] tarafından yapılan bir diğer çalışmada özellikle uzun cümle çevirilerinde RNN'in zayıf performans gösterdiğini belirlenmiştir.

Sutskever ve diğ. [5] yaptıkları çalışmada İngilizceden Fransızcaya makine çevirisi için benzer bir mimari üzerinde giriş dizisini sabit boyutlu bir vektöre dönüştürmek için çok katmanlı bir Uzun Kısa Süreli Bellek (LSTM) ve tekrar hedef diziyi vektörden çözmek için başka bir LSTM katmanı kullanmışlardır.

Chung ve diğ. [6] ise LSTM ve Geçitli Tekrarlayan Birim (GRU) gibi farklı tekrarlayan yapay sinir ağlarının kullanımına odaklanmışlardır. Çok sesli müzik ve konuşma sinyali veri setleri üzerinde yaptıkları bir dizi modelleme deneyleri sonucu LSTM ve GRU gibi gelişmiş tekrarlayan yapay sinir ağlarının sıradan ağlara göre daha üstün sonuçlar verdiğini ortaya koymuşlardır.

Bahdanau ve diğ. [7] kodlayıcı-kod çözücü yapıları için dikkat mekanizmasını tanıtmışlardır. Bu mekanizma makine çevirisi için daha doğru çeviriler üretmek adına modelin kaynak cümlesinin belirli kısımlarına "dikkat etmesini" sağlayarak uzun cümlelerde bağlamı daha iyi yakalamasına olanak sağlamıştır.

Vaswani ve diğ. [8] özgün dikkat mekanizmalarını kullanarak sıralı bilgileri modellemede önemli bir başarı sağlayan Transformer adında yeni bir mimari sunarak kodlayıcı-kod çözücü yapısına yeni bir perspektif getirmiştir. Bu mimari özellikle dizi elemanları arasındaki uzun mesafeli bağımlılıkları daha etkili bir şekilde ele alabilme ve öğrenebilme kapasitesine sahip olmasıyla dikkat çekmiştir. Bu sayede modelin anlama ve dil işleme kabiliyetlerini arttırdığı sonucuna varılmıştır.

1.2 Projenin Amacı ve Kapsamı

Literatürde bu alanda yapılan çalışmalarda elde edilen ve bulgu ve sonuçlar doğal dil işleme ve nöral makine çevirisi gibi konularda doğru kabul edilen tek bir doğru yöntemin olmadığına işaret etmektedir. Dikkat mekanizması ve Transformer gibi yeni nesil yapılar her ne kadar bu alandaki gelişmeleri hızlandırıp dikkatleri bu yöne çekse de klasik kodlayıcı-kod çözücü tabanlı sinir ağı mimarisinin önemi azalmamış, aksine birçok uygulama ve araştırmada yol gösterici olmaya devam etmiştir.

Bu proje beş ana bölümden oluşmaktadır. Giriş bölümünde araştırma konusu olan nöral makine çevirisine giriş yapılarak bu alanla ilgili literatürde yapılmış önde gelen çalışmalar hakkında bilgi verilmiştir. Bölüm 2’de yapay sinir ağları ile tekrarlayan yapay sinir ağları ve yapılarına dair teorik bilgiler ele alınarak çalışma mekanizmaları hakkında özet bilgi verilmiştir. Bölüm 3’te projenin pratik yönüne odaklanılarak kullanılan veri seti, veri ön işleme adımları ve model mimarilerinin tasarım süreci detaylı bir şekilde açıklanmıştır. Bu kapsamda oluşturulan üç farklı kodlayıcı-kod çözücü tabanlı modelin nasıl oluşturulduğu ve eğitim süreçlerinin nasıl yürütüldüğü açıklanmıştır. Bölüm 4’te oluşturulan modellerin performanslarının karşılaştırmalı olarak değerlendirilmesi için kullanılan metrikler ve analiz yöntemleri ile elde edilen sonuçlar sunulmuştur. Son bölümde model performansları üzerine genel bulgular tartışılarak araştırmanın kısıtlılıkları incelenmiş ve gelecek çalışmalar için önerilere yer verilmiştir.

Bölüm 2

Yapay Sinir Ağları

2.1 Yapay Sinir Ağlarının Tanımı

Yapay sinir ağı (YSA) insan beynindeki biyolojik sinir ağlarının işleyişini taklit eden bilgisayar algoritmalarıdır. Makine öğreniminin bir alt kümesi olan ve derin öğrenme olarak da adlandırılan bu yaklaşım adını insan beynindeki biyolojik nöronların birbirleri arasında sinyalleri iletme şekline, çevreden gelen uyarıları (verileri) öğrenme ve işleme sürecinden ilham almıştır [9].

YSA'nın temel bileşeni olan nöronlar bilgiyi işleyen ve depolayan birimlerdir. Sayısal bilgisayarlarda bilgi ikili kod (binary code) şeklinde 1 ve 0 serileri ile temsil edilirken, YSA'larda matematiksel işlev ve bağlantı ağırlıklarıyla temsil edilir. Bu bağlantı ağırlıkları ağı boyunca bilgiyi dağıtmak için kullanılırlar. Bu yönüyle bilgiyi belleğin belirli bir kısmında saklayan klasik bilgisayarlardan ayrılırlar [10].

2.2 Yapay Sinir Ağlarının Tarihçesi

YSA üzerine ilk çalışmalar 1943 yılında Warren McCulloch ve Walter Pitts tarafından elektrik devreleriyle oluşturulan basit bir sinir ağı modeli ile başlamıştır [11]. Bu çalışmayı takiben 1948 yılında Wiener "Cybernetics" adlı eserinde sinirlerin işleyişine değinmiştir [12]. 1949 yılında ise Hebb, "Organization of Behavior" kitabında öğrenme teorisini inceleyerek adaptif sinir ağları için temel teşkil eden Hebb Kuralı'nı tanıtmıştır [13]. Bu çalışmalar biyolojik sinir sisteminin bileşenlerinin yapay ve eşdeğer bileşenlerini yaratma girişimlerine büyük katkıda bulunmuştur.

1957 yılına gelindiğinde Frank Rosenblatt tarafından Algılayıcı (Perceptron) adı verilen ve birçok değeri girdi olarak alıp yalnızca bir çıktı üretmeye yarayan ilk tek katmanlı yapay sinir ağı algoritması geliştirilmiştir [14].

Bundan iki yıl sonra Stanford Üniversitesi'nden Bernard Widrow ve Marcian Hoff ADALINE (Adaptive Linear Elements - Uyarlanabilir Doğrusal Elemanlar) ve MADALINE (Multiple Adaptive Linear Elements - Çoklu Uyarlanabilir Doğrusal Elemanlar) şeklinde adlandırdıkları iki ağ modeli geliştirmişlerdir. Bunlardan ilki yapısal olarak algılayıcıya benzese de algılayıcı aktivasyon fonksiyonu olarak eşik fonksiyonu kullanırken ADALINE doğrusal fonksiyon kullanır [15]. Onun çoklu hali olan MADALINE ise telefon hatlarının yankılarını yok etmede kullanılarak gerçek dünya sorunlarına uygulanmış ilk sinir ağıdır [10].

1969 yılında ise Minsky ve Papert standart bir tek katmanlı algılayıcının XOR problemi gibi doğrusal olarak sınıflandırılmayan mantık problemini çözmede yetersiz kaldığını ispatlayarak bu sınırlamaları aşmak için iki katmanlı ileri beslemeli ağların kullanılabileceğini ileri sürmüşlerdir [16].

Belirli bir süre araştırmaların durulması sonrasında 1980'lerde yapay sinir ağları araştırmaları tekrar artmış, Anderson, Wilshaw, Kanerva ve Kohonen bilim insanları bu alanda katkıda bulunmuşlardır.

1982'de Hopfield optimizasyon gibi teknik problemleri çözmek için doğrusal olmayan Hopfield ağını, 1984'te ise Kohonen gözetimsiz öğrenen Kendini Düzenleyen Haritalar (Self Organizing Maps) ağını geliştirmişlerdir [17].

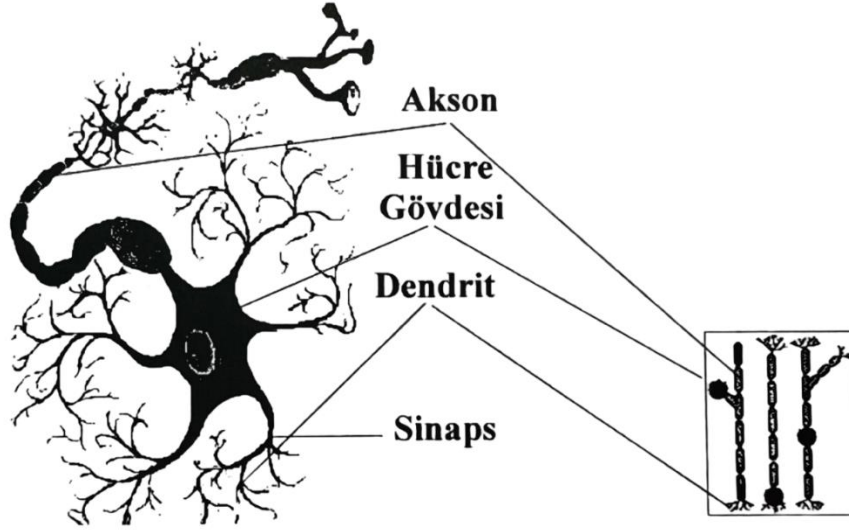
1986 yılında Rumelhart ve McClelland karmaşık ve çok katmanlı ağlar için geriye yayılmalı öğrenme algoritmasını ortaya koyarak ilerideki birçok araştırma ve çalışmaya ışık tutmuşlardır [18].

2.3 Biyolojik Sinir Ağlarının Yapısı

İnsan beyni mantıksal düşünme, ezberleme ve hesaplama yeteneğine sahip son derece gelişmiş ve karmaşık bir yapıya sahiptir. Yapay sinir ağları üzerine yapılan araştırma ve çalışmalar insan beyninin bu gelişmiş mimarisine dayalı yüksek seviyede veri işleme kapasitesine sahip bir algoritma oluşturmaya odaklanmıştır.

Yapay sinir ağlarının çalışma mekanizmasının daha iyi anlaşılabilmesi için öncelikle biyolojik sinir ağlarının yapısının ve işleyişinin bilinmesi gerekmektedir.

Nöron olarak bilinen sinir hücresi beynin en temel birimidir. Bir insan beyni yaklaşık 100 milyar sinir hücresinin birleşiminden oluşmaktadır. Şekil 2.1’de biyolojik bir beyin sinir hücresinin yapısı gösterilmiştir.



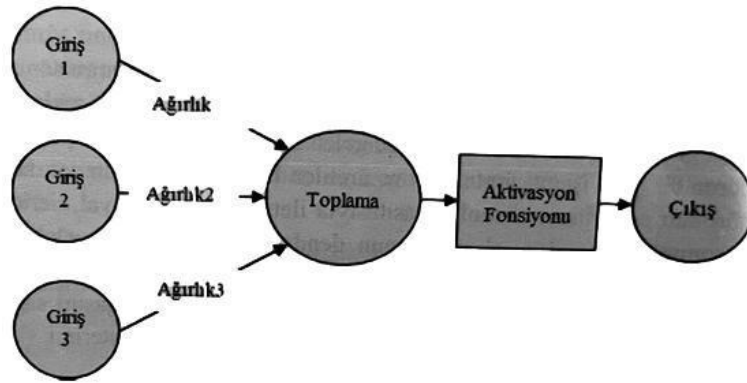
Şekil 2.1: Biyolojik bir beyin sinir hücresi [10]

Bir sinir hücresinin çalışması komşu sinir hücrelerinden alınan uyarılara dayanır. Dendritler aracılığıyla sinir hücresine giriş yapan bu uyarılar soma adı verilen gövde kısmına iletilerek burada doğrusal olmayan bir şekilde işlenirler. İşlenen bu sinyaller akson adı verilen kanallar aracılığıyla sinaps adı verilen uç kısma iletilirler. Sinapsta nörotransmittere dönüştürülerek diğer sinir hücrelerine gönderilirler.

Yapay sinir ağları beynin yapısı örnek alınarak geliştirilmelerine karşın bazı özellikleri ile beynin yapısı ve işleyişinden ayrılırlar. Bu farklar takip eden alt başlıkta ayrıntılı bir şekilde ele alınmıştır.

2.4 Yapay Sinir Ağlarının Yapısı ve Bileşenleri

Yapay sinir ağları biyolojik sinir ağlarına benzer şekilde yapay nöronlardan oluşmaktadır. Yapay bir nöronda giriş alındıktan sonra bu girişlerin her biri ağırlıklar ile çarpılarak toplanmakta ve çıkış için sınırlamaların belirlenmesi amacıyla aktivasyon fonksiyonuna gönderilmektedir. Şekil 2.2’de yapay sinir hücresinin yapısı şematik olarak gösterilmiştir.



Şekil 2.2: Yapay sinir hücresi [19]

Temel olarak yapay sinir ağları girişler, ağırlıklar, toplama fonksiyonu, aktivasyon fonksiyonu ve çıkış işlevi olmak üzere beş temel yapıdan oluşmaktadır.

2.4.1 Girişler

Girişler, yapay sinir hücresinin aldığı veriyi ifade etmekle beraber matematiksel olarak $x_1, x_2, x_3, \dots, x_n$ şeklinde ifade edilmektedir.

2.4.2 Ağırlıklar

Ağırlıklar, girişler aracılığıyla gelen verinin sinir üzerindeki etkisini belirlemeye yarayan katsayılardır. Matematiksel olarak $w_1, w_2, w_3, \dots, w_n$ şeklinde ifade edilmektedir.

2.4.3 Toplama Fonksiyonu

Toplama fonksiyonu, her bir ağırlığın ait olduğu girdi değeri ile çarpımının toplamlarını eşik değerini de dahil ederek aktivasyon fonksiyonuna göndermeye yarar. Matematiksel olarak aşağıdaki şekilde ifade edilir [20].

$$z_i = \sum_{i=0}^n (w_i x_i) + b_j \quad (2.1)$$

2.4.4 Aktivasyon Fonksiyonu

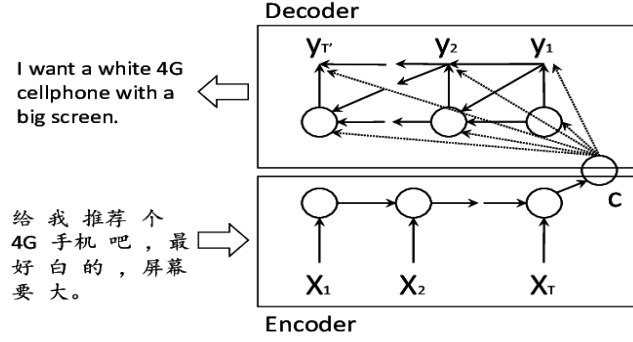
Aktivasyon fonksiyonu belirleyici bir fonksiyon olarak nöronun çıkışının nasıl şekilleneceğinin belirlendiği kısımdır. Yapay sinir ağlarında yaygın olarak kullanılan aktivasyon fonksiyonları lineer, eşik değer, hiperbolik tanjant, sigmoid ve softmax olarak sıralanabilir. Çıkışı doğrudan etkileyen bir faktör olduğundan dikkatle seçilmesi önemlidir. Örneğin pozitif bir çıkış isteniyorsa sigmoid, hem pozitif hem de negatif bir çıkış isteniyorsa hiperbolik tanjant aktivasyon fonksiyonu tercih edilmelidir [19].

2.4.5 Çıkış İşlevi

Çıkış işlevi, aktivasyon fonksiyonu sonucunun gönderildiği kısımdır. Her yapay sinirin yalnızca bir çıkışı bulunmaktadır. Bu çıkış aynı zamanda bağlı bulunan komşu sinirlerin giriş kısmıdır.

2.5 Yapay Sinir Ağları ile Nöral Makine Çevirisi

Nöral makine çevirisi, diller arası metin çevirilerini insan müdahalesi olmadan otomatikleştirmek için yapay zekâ ve derin öğrenme tekniklerini kullanan bir disiplindir. Bu alanda kullanılan yapay sinir ağları kaynak dildeki metni hedef dile doğru bir şekilde çevirmek için eğitilir.



Şekil 2.3: Çince'den İngilizce'ye makine çevirisi için RNN tabanlı bir sinir ağı modelinin şematik gösterimi [21]

Nöral makine çevirisi globalleşen dünyada dil engellerini aşmada kritik bir rol oynamaktadır. Hızlı ve etkili çeviriler sunma kapasitesi bu teknolojinin uluslararası iş birliği, eğitim ve kültür alışverişi gibi birçok alanda popüler olmasına neden olmuştur. Geleneksel çeviri yaklaşımlarından farklı olarak, sinir ağları temelli çeviri, metindeki bağlamları ve nüansları daha iyi yakalamak için öğrenme mekanizmalarını kullanır.

Günümüzde yapay sinir ağları kullanılarak gerçekleştirilen bir diğer adıyla nöral makine çevirisi, uluslararası haber ajanslarından e-ticaret sitelerine, akademik araştırmalardan sosyal medya platformlarına kadar geniş bir yelpazede aktif olarak kullanılmaktadır.

2.6 Yapay Sinir Ağı Türleri ve Kullanma Alanları

Yapay sinir ağları çözülmek istenen problemin niteliğine ve istenen çözüme bağlı olarak göre çeşitlere ayrılır. Bunlara örnek olarak Evrişimli Sinir Ağları, Tekrarlayan Sinir Ağları ve Derin İnanç Ağları verilebilir.

Yapay Sinir Ağları ve uygulanma alanları Tablo 2.1'de görüldüğü gibidir.

Tablo 2.1: Ağlar ve uygulama alanları [10]

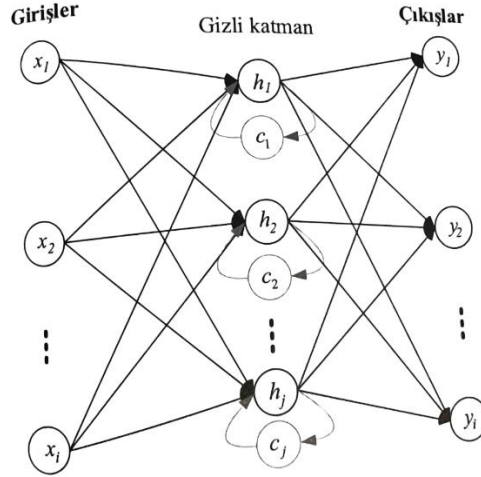
Ağ Türü	Uygulama Alanı
Evrişimli Sinir Ağları	Görüntü ve video işleme, örüntü tanıma
Tekrarlayan Sinir Ağları	Doğal dil işleme
Derin İnanç Ağları	Örüntü tanıma, doğal dil işleme

Doğal dil işleminin uygulama alanlarından olan makine çevirisi problemlerinin çözümü için sıralı dizileri analiz ederek sınıflandırabilecek bir sinir ağı yapısına ihtiyaç vardır. Bu problemlerin çözümünde Tekrarlayan Sinir Ağları en sık kullanılan ağ mimarisi olarak ön plana çıkmaktadır. Aynı zamanda bu çalışmanın ana metodolojisini oluşturan Tekrarlayan Sinir Ağlarının yapısı aşağıdaki başlıklarda ayrıntılı bir şekilde ele alınmıştır.

2.6.1 Tekrarlayan Sinir Ağları

Tekrarlayan sinir ağları (Recurrent Neural Network - RNN), sıralı veriyi işleyerek önceki adımlardan gelen bilgiyi hafızasında tutabilen ve bu bilgiyi sonraki adımlar için kullanabilen derin öğrenme algoritmasıdır. Genel olarak farklı uzunluktaki giriş dizilerini yine farklı uzunluktaki çıkış dizilerine dönüştürürken kullanıldıklarından metin işleme gibi doğal dil işleme uygulamalarında sıklıkla tercih edilirler.

Şekil 2.4'te tekrarlayan sinir ağının yapısı gösterilmiştir.



Şekil 2.4: Tekrarlayan sinir ağının yapısı [10]

2.6.1.1 Tekrarlayan Sinir Ağları Çalışma Prensibi

Geleneksel sinir ağlarında tüm giriş ve çıkışlar birbirinden bağımsız olarak davranır. Makine çevirisi gibi cümlede geçen bir kelimedenden sonra hangi kelimenin gelmesi

gerektiğinin tahmin edildiği problemlerde ise önceki kelimelerin hatırlanması gereklidir. Bu tarz problemlerin çözümünde geleneksel sinir ağları yetersiz kaldığından tekrarlayan sinir ağlarının önemi ön plana çıkmıştır.

Tekrarlayan sinir ağının her bir katmanında, bir önceki zaman adımındaki durum ve mevcut zaman adımındaki yeni girdi, şimdiki durumu oluşturmak için bir araya getirilir. Bu işlem dizideki tüm elemanlar için tekrarlanır. Böylece ağın dizideki sıralı yapıyı ve zaman içindeki bağlantılarını öğrenmesi sağlanır.

Tekrarlayan sinir ağları eğilirken kaybolan gradyan (vanishing gradient) ve patlayan gradyan (exploding gradient) gibi iki önemli sorunla karşılaşmaktadır [22]. Bu sorunların üstesinden gelmek amacıyla LSTM ve GRU gibi farklı türde tekrarlayan sinir ağları geliştirilmiştir.

2.6.1.2 Tekrarlayan Sinir Ağı Mimarileri

Tekrarlayan sinir ağlarının farklı varyasyonları belirli problemlere yönelik olarak geliştirilmiştir. Her bir tür, farklı yapısal özellikler ve öğrenme yetenekleri ile sıralı verilerin farklı yönlerini yakalamak için tasarlanmıştır. Aralarındaki temel fark döngünün bir diğer deyişle geri bildirim nasıl yapılacağıdır [10].

En çok bilinen mimariler;

- Tekrarlayan Sinir Ağları (Recurrent Neural Network - RNN)
- Uzun Kısa Süreli Bellek Ağları (Long Short-Term Memory - LSTM)
- Geçitli Tekrarlayan Birim (Gated Recurrent Unit - GRU) ağları olarak sıralanabilir.

RNN olarak da bilinen klasik Tekrarlayan Sinir Ağları önceki bölümlerde ele alındığından takip eden bölümlerde sırasıyla LSTM ve GRU'nun yapısı incelenmiştir.

2.6.2 Uzun Kısa Süreli Bellek Ağları (LSTM)

Uzun Kısa Süreli Bellek Ağları, Hochreiter ve Schmidhuber tarafından 1997'de önerilmiş ve dizilim biçimindeki verilerde bağımlılıkları modellemek için geliştirilmiş bir tekrarlayan sinir ağı mimarisidir [23]. Tipik bir LSTM birimi, bilginin uzun süreli

saklanıp gerektiğinde erişilmesini yöneten bir hücre ile giriş, çıkış ve unutma kapısı olmak üzere üç kapıdan oluşur [24]. Bunlardan giriş kapısı yeni bilginin hafızaya ne zaman gireceğini denetlerken, unutma kapısı gereksiz bir bilginin unutulmasını ve çıkış kapısı ise hücredeki bilginin hücrenin çıkışında ne zaman kullanılacağını yönetir [10]. Bu sayede LSTM tabanlı mimariler geleneksel tekrarlayan sinir ağlarının karşılaştığı kaybolan ve patlayan gradyan sorunu ile büyük ölçüde baş ederek, özellikle uzun dizilimlerden oluşan veriler üzerinde daha iyi öğrenme performansı sunar [24].

2.6.3 Geçitli Tekrarlayan Birim Ağları (GRU)

Uzun Kısa Süreli Bellek hücresinin daha az parametreye sahip bir versiyonu olan Geçitli Tekrarlayan Birim, Cho ve diğ. [3] tarafından 2014 yılında önerilmiştir. Her ne kadar Uzun Kısa Süreli Bellek hücresi ile benzer bir performansa sahip olmasının yanı sıra daha hafif bir yapıya sahip olduğundan daha hızlı çalıştırılıp eğitilirler.

Uzun Kısa Süreli Bellek hücresinden farklı olarak güncelleme ve sıfırlama kapısı olmak üzere iki kapı yapısına sahiptir [24]. Güncelleme kapısı hücrenin önceki içeriğinin ne kadarının korunacağını belirlerken, sıfırlama kapısı ise yeni girişin önceki hücre içeriğiyle nasıl birleştirileceğini kontrol eder [10]. Bu mekanizma GRU'nun bilgiyi etkili bir şekilde saklayarak gereksiz bilgileri unutulmasını sağlar.

Bu bölümde yapay sinir ağlarına giriş yapılarak tekrarlayan yapay sinir ağları ve yapılarına dair teorik bilgiler ele alınmış, çalışma mekanizmaları hakkında bilgi verilmiştir. Bir sonraki bölümde çalışmanın yürütülmesinde başvuru materyal ve yöntem ile ilgili detaylar sunulacaktır.

Bölüm 3

Materyal ve Yöntem

Projenin gerçekleştirilmesi sürecinde Google firması tarafından geliştirilmiş ve Google Colab adı verilen bulut tabanlı ve Jupyter Notebook benzeri bir notebook geliştirme ortamı kullanılmıştır. Modellerin eğitimi iyi ve hızlı bir donanım gerektirdiğinden Colab Pro üzerinde Linux işletim sistemine sahip, AMD x86_64 işlemcili, 51 GB CPU RAM ve NVIDIA Tesla V100 GPU 16 GB grafik kartına sahip bir donanım kullanılmıştır.

DeneySEL süreçte Python programlama dili ile veri manipülasyonu ve analizi için Pandas, sinir ağı oluşturmak için Tensorflow ve Keras gibi çeşitli kütüphane ve modüllerden faydalanılmıştır. Türkçe-İngilizce çeviri veri setinin orijinali Kaggle [25] üzerinden elde edilmiş olup veriye direkt olarak internet üzerinden ulaşılma kolaylığı bakımından Github [26] üzerine yüklenmiştir. Veri seti eğitim ve test olarak ayrılırken Scikit-learn kütüphanesinin train_test_split fonksiyonu kullanılmış olup %80 eğitim ve %20 test şeklinde bölünmüştür. Araştırmanın deneysel aşamaları Şekil 3.1’de sunulmuştur.



Şekil 3.1: Araştırmanın deneysel aşamaları

Takip eden alt başlıklarda araştırmanın deneysel aşamalarına ilişkin bilgi verilerek elde edilen bulgu ve sonuçlar paylaşılmıştır.

3.1 Veri Okuma ve Ön İşleme

3.1.1 Veri Seti

Türkçe-İngilizce çeviri veri setinin orijinali 473035 satırdan oluşmakta olup oluşturulacak modellerin eğitim süresini kısaltmak amacıyla ilk 100 bin satır baz alınmıştır. Tablo 3.1’de veri setinden örnek cümle kesitleri gösterilmiştir.

Tablo 3.1: İşlenmemiş veri setinden örnek cümleler

Türkçe	İngilizce
Bu yüzden geri döndüm.	That's why I came back.
Geri dönmemin nedeni bu.	That's why I came back.
Bu yüzden buraya geldim.	That's why I came here.
Geç gelmemin nedeni bu.	That's why I came late.
Bu yüzden seni işe aldım.	That's why I hired you.

3.1.2 Veri Okuma

İlgili veri seti HTTP istekleri göndermeye yarayan requests modülü yardımıyla okunarak Türkçe ve İngilizce dil çiftlerine bölünmüştür. Sonrasında pandas kütüphanesi yardımıyla bir veri çerçevesi (dataframe) oluşturulmuştur.

3.1.3 Veri Ön İşleme

Veri ön işleme, nöral makine çevirisi gibi doğal dil işleme görevlerinde önemli bir adımdır. Bu süreç veriyi modelin öğrenmesi için daha uygun hale getirerek modelin daha etkin bir şekilde genelleme yapmasını sağlar. Böylelikle çeviri kalitesi ve doğruluğu da artırılmış olur.

3.1.3.1 Veri Temizleme

Çalışma kapsamında aşağıdaki veri temizleme işlemleri uygulanmıştır;

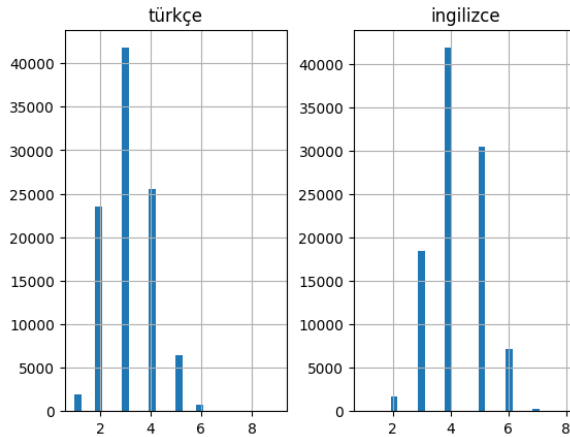
- **Kısaltmaların Açılması:** Veri setindeki İngilizce kısaltmalar contractions kütüphanesi yardımıyla uzun hale dönüştürülmüştür.
- **Noktalama İşaretlerinin Kaldırılması:** Modelin yalnızca kelime temsilleri üzerine odaklanmasını sağlamak amacıyla tüm noktalama işaretleri kaldırılmıştır.
- **Sayıların Kaldırılması:** Bu çalışma kapsamında sayıların çeviriye özel bir değer katmadığı düşünüldüğünden tüm sayılar kaldırılmıştır.
- **Tüm Harfleri Küçük Harfe Dönüştürme:** Metinlerdeki tüm harfler küçük harfe dönüştürülmüştür.

Belirtilen temizleme adımları sonrası veri setindeki cümlelerin görünümü Tablo 3.2'deki gibidir.

Tablo 3.2: İşlenmiş veri setinden örnek cümleler

Türkçe	İngilizce
bu yüzden geri döndüm	that is why i came back
geri dönmemin nedeni bu	that is why i came back
bu yüzden buraya geldim	that is why i came here
geç gelmemin nedeni bu	that is why i came late
bu yüzden seni işe aldım	that is why i hired you

Temizleme sonrası veri setindeki cümle uzunlukları dağılımı Şekil 3.2'de verilmiştir. Maksimum cümle uzunluğu Türkçe 9 iken İngilizce için 8'dir.



Şekil 3.2: Veri setindeki cümle uzunluklarının dağılımı

3.1.3.2 Tokenizasyon

Bu adımda metinlerde geçen kelimelerin her birine benzersiz bir sayısal değer (token) atanmıştır. Böylelikle metinsel veriler modelin anlayabileceği sayısal değerlere dönüştürülmüştür.

3.2 Eğitim ve Test Verisinin Hazırlanması

Ön işlemeden geçirilen veri seti scikit-learn kütüphanesinin `train_test_split` fonksiyonu kullanılarak %80'i eğitim ve %20'si test olmak üzere iki ayrı kümeye ayrılmıştır. Elde edilen eğitim ve test kümeleri ayrıca tanımlanan `encode_sequences` fonksiyonu ile tokenlara dönüştürülerek modelin anlayabileceği sayısal dizilere çevrilmiştir.

3.3 Modellerin Oluşturulması

Proje kapsamında RNN, LSTM ve GRU kullanarak toplamda üç farklı kodlayıcı-kod çözücü tabanlı model mimarisi oluşturulmuştur.

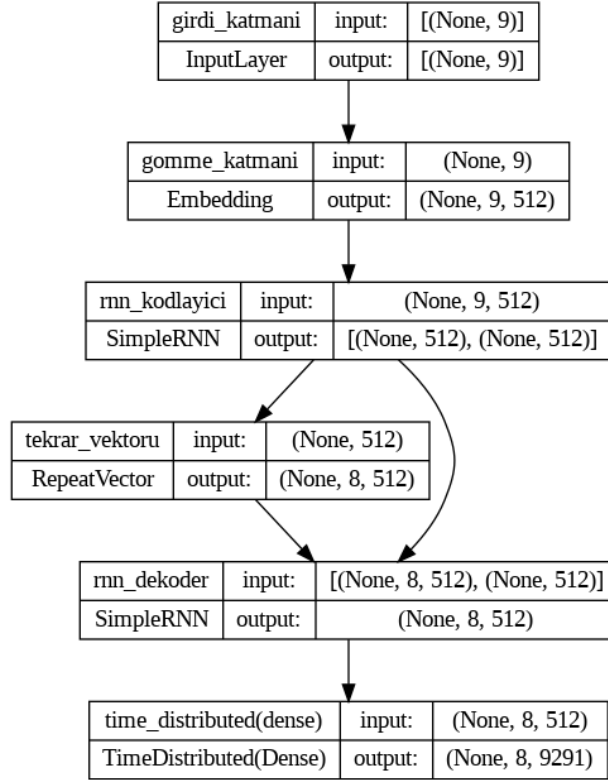
Kodlayıcı-kod çözücü tabanlı model mimarisi özellikle makine çevirisinde sıkça kullanılan bir derin öğrenme modelidir. Bu mimari türü girdi metni ile çıktı metni arasında bir eşleştirme yapmayı amaçlar. Bu mimari, metin verilerinin işlenmesinde kullanılır ve genellikle iki ana bileşen içerir:

- **Kodlayıcı (Encoder):** Bu kısımda giriş metni kodlayıcı ağı tarafından işlenerek daha yüksek seviyeli bir temsile dönüştürülür.
- **Kod Çözücü (Decoder):** Kod çözücü kodlayıcıdan gelen temsili alarak çıktı metnini oluşturur.

Buna ek olarak kodlayıcı-kod çözücü tabanlı model mimarilerinde öğrenme sürecini iyileştirmek için isteğe bağlı olarak çeşitli öğrenme mekanizmaları da kullanılabilir. Bu proje kapsamında dikkat mekanizması kullanılmamış olup klasik kodlayıcı-kod çözücü tabanlı model mimarileri üzerine odaklanılmıştır.

3.3.1 RNN ile Kodlayıcı-Kod Çözücü Tabanlı Model Oluşturma

RNN kullanılarak oluşturulan kodlayıcı-kod çözücü model Şekil 3.3'te gösterilmiştir.



Şekil 3.3: RNN tabanlı kodlayıcı-kod çözücü modeli

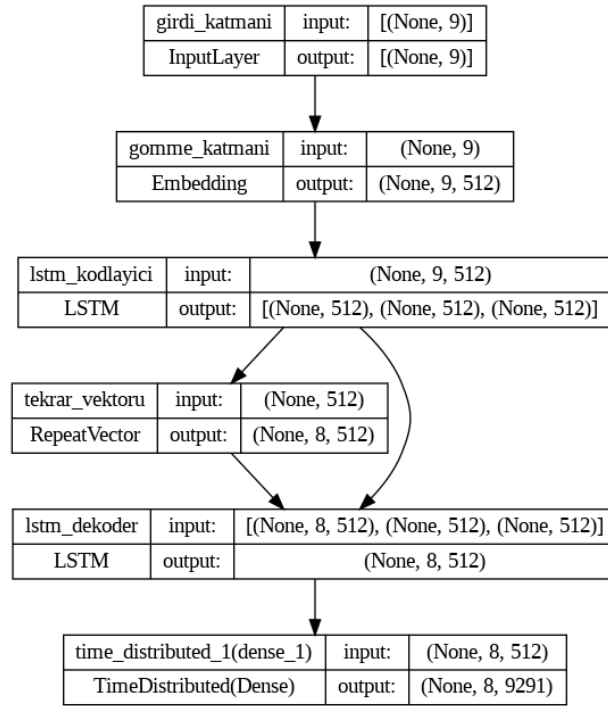
Modelde kullanılan katmanlar ve işlevleri aşağıdaki gibidir;

- **Girdi Katmanı (Input Layer):** Ağın ilk katmanı olan girdi katmanı girdi verilerinin alındığı katmandır. Girdi katmanındaki 9 sayısı, modelin her biri 9 elemandan (kelime ya da token) oluşan dizi şeklindeki verileri işleyebileceği anlamına gelir. Bu sayı aynı zamanda veri setindeki Türkçe cümlelerin maksimum uzunluğuna eşittir.
- **Gömme Katmanı (Embedding Layer):** Bu katman girdi verilerini sabit boyutlu yoğun vektörlere dönüştürmek için kullanılır.

- **RNN Kodlayıcı (RNN Encoder):** Bu katman gömme katmanından gelen vektörleri alarak zamana bağlı verileri işler.
- **Tekrar Vektörü (Repeat Vector):** Kodlayıcının son çıktısını alarak bu çıktıyı belirli bir sayı kadar tekrarlamaya yarar.
- **RNN Kod Çözücü (Decoder):** Tekrar vektörü katmanından gelen diziyi ve kodlayıcının son çıktısını alarak işler.
- **TimeDistributed Katmanı:** Bu katman, çözücü tarafından üretilen çıktıları zaman adımları boyunca işleyip her zaman adımı için ayrı sonuçlar üretmek için kullanılır. Bu sayede sıralı verilerle çalışırken modelin zaman bağımlılığını daha iyi yakalaması sağlanmış olur.

3.3.2 LSTM ile Kodlayıcı-Kod Çözücü Tabanlı Model Oluşturma

LSTM kullanılarak oluşturulan kodlayıcı-kod çözücü model Şekil 3.4'te gösterilmiştir.

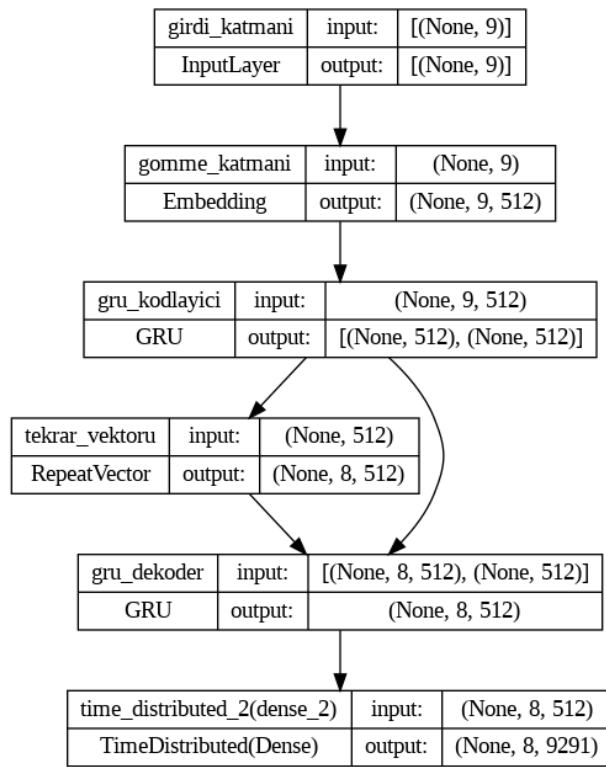


Şekil 3.4: LSTM tabanlı kodlayıcı-kod çözücü modeli

Modelin kodlayıcı ve kod çözücü katmanında LSTM kullanılmış olup diğer katmanlar aynı kalmıştır.

3.3.3 GRU ile Kodlayıcı-Kod Çözücü Tabanlı Model Oluşturma

GRU kullanılarak oluşturulan kodlayıcı-kod çözücü model Şekil 3.5'te gösterilmiştir.



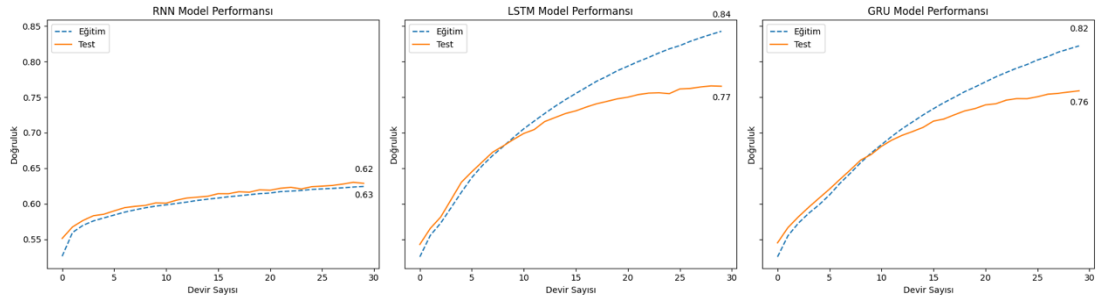
Şekil 3.5: GRU tabanlı kodlayıcı-kod çözücü modeli

Modelin kodlayıcı ve kod çözücü katmanında GRU kullanılmış olup diğer katmanlar aynı kalmıştır.

3.4 Model Performans ve Hatalarının Karşılaştırılması

Proje kapsamında RNN, LSTM ve GRU kullanılarak üç farklı kodlayıcı-kod çözücü tabanlı sıralı model oluşturulmuştur. Her bir devir için hata ve doğruluk değerleri hem eğitim hem de doğrulama kümesi için kaydedilerek grafikler oluşturulmuş ve modellerin performansı karşılaştırılmıştır.

RNN, LSTM ve GRU tabanlı kodlayıcı-kod çözücü tabanlı modellerin doğruluk grafiği Şekil 3.6'daki gibidir.

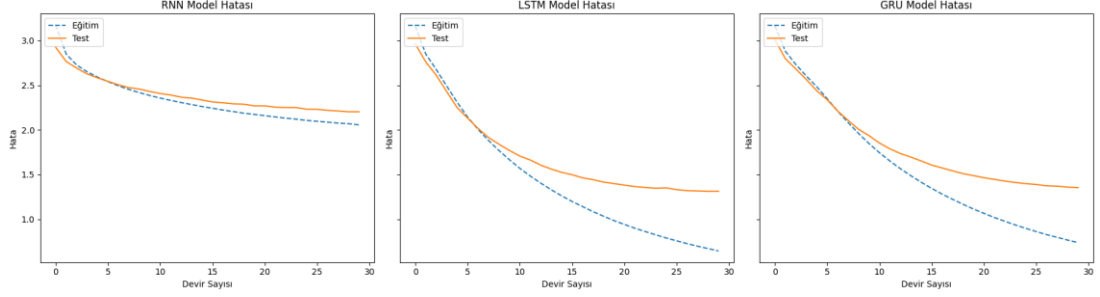


Şekil 3.6: RNN, LSTM ve GRU tabanlı kodlayıcı-kod çözücü tabanlı modellerin doğruluk grafiği

Doğruluk grafiklerine bakılarak aşağıdaki sonuçlara ulaşılmıştır;

- RNN modeli için doğruluk oranının diğer modellere kıyasla çok daha düşük kaldığı gözlemlenmiştir. Buradan RNN'in makine çevirisi gibi karmaşık görevler için yeterli olmadığı sonucuna varılabilir.
- LSTM modelinin doğruluk oranının diğer modellere kıyasla en yüksek olduğu gözlemlenmiştir. Buradan modelin genelleştirme yapabildiği ve RNN kadar overfitting (aşırı uyum) problemine maruz kalmadığı sonucuna varılmıştır.
- GRU modelinin doğruluk grafiği, test doğruluğu daha düşük olmasına rağmen LSTM modeline benzer bir eğilim göstermiştir. GRU modeli LSTM modeline göre daha az parametreye sahip olduğundan daha hızlı eğitilebilmektedir. Performanslarının yakın olduğu göz önüne alındığında özellikle donanım kaynaklarının sınırlı olduğu durumlarda tercih edilebilir.

Şekil 3.7’de RNN, LSTM ve GRU tabanlı kodlayıcı-kod çözücü tabanlı modellerin hata grafiği gösterilmiştir.



Şekil 3.7: RNN, LSTM ve GRU tabanlı kodlayıcı-kod çözücü tabanlı modellerin hata grafiği

Hata grafiklerine bakılarak aşağıdaki sonuçlara ulaşılmıştır;

- RNN modeli için eğitim hatası test hatasından daha hızlı azaldığından modelin overfitting (aşırı uyum) problemi yaşadığı sonucuna varılmıştır.
- LSTM modeli diğer modellere kıyasla daha düzgün bir azalma göstermiştir. Test hatası, devir sayısı arttıkça sürekli bir şekilde azalarak eğitim hatasına yaklaştığından modelin iyi bir şekilde genelleme yapabildiği belirlenmiştir.
- GRU modeli LSTM modeli ile benzer bir performans sergilemiştir. Bu modelin eğitim verilerine iyi uyum sağladığı ve test verileri üzerinde iyi bir genelleme yapabildiği sonucuna ulaşılmıştır.

3.5 Model Çeviri Performansı ve BLEU Puanı

Oluşturulan her bir modele ait veri setinden rasgele Türkçe cümleler için yapılmış olan çevirilere ait örnekler Tablo 3.3’te gösterilmiştir.

Tablo 3.3: Modeller kullanılarak yapılan çeviri örnekleri

Türkçe	İngilizce	RNN Tahmin	LSTM Tahmin	GRU Tahmin
şimdiden özür dilerim	i apologize in advance	i felt	i apologize	i am apologized
sorun bu	this is the problem	it is	the is the problem	this is issue
bu gerçekten sertti	it was really tough	this is really great	it was really hard	it was really hard
o oğluna baktı	she cared for her son	he is the	she looked at the	he looked at the
tom zamanında geldi	tom was on time	tom is always	tom arrived time	tom arrived on time
neden gitmiyorsun	why are not you going	why are you	why are not you going	why are not you going
çocuk mutlu oldu	the boy became happy	the floor is	it boy was happy	the man is busy
o sık sık soğuk algınlığına yakalandı	she caught colds often	he often to	he often caught and cold	he was got cold
sen düzenli misin	are you methodical	are you	are you tidy	are you tidy
bir taksiye binelim mi	shall we take a taxi	can you have a	can you get a cab	can you go a cab

Tabloya bakıldığında LSTM ve GRU modellerinin RNN'e kıyasla çok daha doğru çeviri yaptığı açıktır.

Makine çevirisi modellerinin performansını değerlendirmek için kullanılan bir diğer metrik BLEU (Bilingual Evaluation Understudy) puanıdır. Bu metrik özellikle tercüme sistemlerinin, dil modellerinin ve makine çevirisi sistemlerinin performansını objektif bir şekilde değerlendirmek için yaygın olarak kullanılmaktadır. Bu proje kapsamında her bir çeviri tahmini için 1-gram, 2-gram, 3-gram ve 4-gram BLEU puanları hesaplanarak bu puanların ortalaması bulunmuştur. 1-gram, 2-gram, 3-gram ve 4-gram puanları, BLEU metriğinin farklı n-gram ağırlıkları ile hesaplanan dört farklı değerini ifade eder. Açıklamaları aşağıdaki gibidir;

- **1-gram BLEU Puanı:** Yalnızca tek kelime seviyesinde eşleşmeleri değerlendirir. Yani çevirideki her kelimenin referans metinde de olup olmadığına bakar.
- **2-gram BLEU Puanı:** Çevirideki iki ardışık kelimenin referans metinde de aynı sırada bulunup bulunmadığına bakılır.
- **3-gram BLEU Puanı:** Çevirideki üç ardışık kelimenin referans metinde de aynı sırada bulunup bulunmadığına bakılır.
- **4-gram BLEU Puanı:** Çevirideki dört ardışık kelimenin referans metinde de aynı sırada bulunup bulunmadığına bakılır. Dört kelimenin ardışık olarak eşleşmesini dikkate alır.

Tablo 3.4’de RNN, LSTM ve GRU modellerinin n-gram BLEU puanları gösterilmiştir.

Tablo 3.4: Her bir model için n-gram BLEU puanı

BLEU Puanı	RNN Modeli	LSTM Modeli	GRU Modeli
1-gram	0.3350	0.6132	0.5952
2-gram	0.1434	0.4235	0.4010
3-gram	0.0943	0.3111	0.2884
4-gram	0.0698	0.2253	0.2041
Ortalama	0.1128	0.3228	0.3012

BLEU puanları tablosu dikkate alındığında LSTM ve GRU modellerinin RNN modeline göre daha yüksek performans gösterdiği görülmüştür. Bu durum LSTM ve GRU modellerinin RNN'e göre daha gelişmiş bir yapıya sahip olmaları ve dolayısıyla da daha iyi bir bellek yönetimine sahip olmaları ile açıklanabilir. LSTM modeli, her ne kadar GRU modeli ile benzer bir performans gösterse de diğer modellere kıyasla makine çevirisi için en iyi performansı sergileyen model olarak ön plana çıkmıştır.

Bölüm 4

Bulgular ve Tartışma

Proje kapsamında oluşturulan RNN, LSTM ve GRU tabanlı modellerin performansları çeşitli metrikler kullanılarak kıyaslanmıştır. Bu modeller arasında LSTM ve GRU'nun RNN'ye göre daha yüksek doğruluk ve daha düşük hata oranına sahip olduğu gözlemlenmiştir. Özellikle LSTM modeli genelleştirme yeteneği ve çeviri kalitesi açısından diğer iki modele göre üstün performans sergilemiştir. Bu sonuçlar LSTM ve GRU'nun, RNN'ye kıyasla daha gelişmiş bellek yönetim mekanizmalarına sahip olmalarından kaynaklanmaktadır.

Çeviri kalitesini değerlendirilmesinde BLEU puan sisteminden faydalanılmıştır. 1-gram'dan 4-gram'a kadar farklı n-gram seviyelerinde hesaplanan BLEU puanları, modellerin çeviri doğruluğunu kapsamlı bir şekilde ölçmek için kullanılmıştır. Yapılan analizler özellikle LSTM modelinin çeviri metinlerinde daha yüksek doğruluk oranı sağladığını göstermiştir.

Veri setinde yapılan ön işleme ve temizleme işlemlerinin model performansları üzerinde belirgin bir etkisi olduğu gözlemlenmiştir. Kısaltmaların açılması, noktalama işaretlerinin ve sayıların kaldırılması gibi işlemler modellerin çeviri metinlerini daha etkin bir şekilde işlemesine olanak sağlamıştır.

Bu çalışmada RNN, LSTM ve GRU gibi farklı sinir ağları ile oluşturulan çeşitli modellerin tercih edilmesinin temel nedeni bu model yapılarının Türkçe-İngilizce çeviri performansları üzerindeki etkilerini karşılaştırmaktır. Aynı zamanda literatürde Türkçe-İngilizce dil ikilisi için bu yönde fazla bir çalışma yapılmamış olması çalışmanın bu alana yönelmesine ön ayak olmuştur.

Çalışmayı sınırlayan bazı faktörler mevcuttur. Veri seti boyutu ve çeşitliliği bu sınırlayıcı faktörlerdendir. Gelecekteki çalışmalarda daha büyük ve çeşitli veri setleri kullanılarak modellerin genelleştirme kabiliyetleri daha iyi bir şekilde test edilebilir. Ayrıca mevcut kodlayıcı-kod çözücü tabanlı sıralı modellere dikkat mekanizması eklenerek veya Transformer mimarisi kullanılarak uzun bağımlılıkların daha etkin bir şekilde ele alınması sağlanarak modellerin öğrenme kapasitelerinde iyileştirmeler yapılabilir.

Bölüm 5

Sonuçlar

Bu çalışmada Türkçe-İngilizce dil ikilisi için çeviri yapabilen RNN, LSTM ve GRU kullanılarak üç farklı kodlayıcı-kod çözücü tabanlı sıralı model oluşturulmuş ve performansları karşılaştırmalı olarak incelenmiştir. Modellerin eğitilmesinde 100 bin satırdan oluşan çeviri veri seti kullanılmış olup %80'i eğitim, %20'si ise test kümesi olarak ayrılmıştır. Elde edilen bulgular test kümesi üzerinde LSTM modelinin 0.77, GRU modelinin 0.76 ve RNN modelinin 0.62 doğruluk oranına sahip olduğunu göstermiştir. N-gram BLEU puanları da hesaba katıldığında, LSTM modelinin genelleştirme yeteneği ve çeviri kalitesi açısından diğer iki modele göre en iyi çeviri performansını sergileyen model olarak ön plana çıktığı görülmüştür.

Bu çalışmanın temel motivasyonu, literatürde tekrarlayan sinir ağları kullanılarak Türkçe-İngilizce makine çevirisi için yeterince çalışma yapılmamış olmasıdır. Çalışmayı sınırlayan faktörler arasında veri setinin boyutu ve çeşitliliği bulunmaktadır. İleriki çalışmalarda daha büyük ve çeşitli veri setleri kullanılarak modellerin genelleştirme yetenekleri daha kapsamlı bir şekilde test edilebilir.

Sonuç olarak bu çalışma, Türkçe-İngilizce makine çevirisi için kullanılan tekrarlayan sinir ağları tabanlı modellerin etkinliğini ve bu modeller arasındaki performans farklılıklarını ortaya koymuştur. Literatürde Türkçe-İngilizce nöral makine çevirisi konusunda yürütülen çalışmalara katkıda bulunacağı düşünülmektedir. İleriki çalışmalarda, tanıtılan modellere dikkat mekanizması eklenerek veya Transformer gibi farklı bir mimari uygulanarak kelimeler arasındaki uzun bağımlılıkların daha etkin bir şekilde ele alınması sağlanıp modellerin öğrenme kapasitelerinde iyileştirmeler yapılması üzerine odaklanılabilir.

Kaynaklar

- [1] Özkan Y. Uygulamalı Derin Öğrenme, 2. baskı. İstanbul: Papatya Yayıncılık Eğitim; 2023.
- [2] Kalchbrenner N, Blunsom P. Recurrent continuous translation models. EMNLP 2013: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing; Seattle: Association for Computational Linguistics; 2013. 1700–1709.
- [3] Cho K, Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. EMNLP 2014: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing; Doha: Association for Computational Linguistics; 2014. 1724–1734.
- [4] Cho K, Merriënboer B, Bahdanau D, Bengio Y. Le On the Properties of Neural Machine Translation: Encoder–Decoder Approaches. Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation; Doha: Association for Computational Linguistics; 2014.
- [5] Sutskever I, Vinyals O, Quoc VL. Sequence to Sequence Learning with Neural Networks. NIPS 2014: Advances in Neural Information Processing Systems 27; Montreal; 2014. 3104–3112.
- [6] Chung J, Gulcehre C, Cho K, Bengio Y. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. NIPS 2014: Deep Learning and Representation Learning Workshop; Montreal; 2014.
- [7] Bahdanau D, Cho K, Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate. ICLR 2015: 3rd International Conference on Learning Representations; San Diego; 2015.

- [8] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN ve diğ. Attention is All you Need. NIPS 2017: Advances in Neural Information Processing Systems 30; Long Beach; 2017. 5998–6008.
- [9] Uhrig RE. Introduction to artificial neural networks. Proceedings of IECON '95-21st Annual Conference on IEEE Industrial Electronics; Orlando; 1995.
- [10] Elmas Ç. Yapay Zeka Uygulamaları, 5. baskı. Ankara: Seçkin Yayıncılık; 2021.
- [11] McCulloch W, Pitts W. A Logical Calculus of Ideas Immanent in Nervous Activity. Bulletin of Mathematical Biophysics; 1943; 5: 115–133. DOI: <https://doi.org/10.1007/BF02478259>.
- [12] Wiener N. Cybernetics: Or Control and Communication in Animal and the Machine. Paris; 1948.
- [13] Hebb DO. The Organization of Behavior. New York; 1949.
- [14] Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review. 1958; 65(6): 386–408. DOI: <https://doi.org/10.1037/h0042519>.
- [15] Arı A, Berberler ME. Yapay Sinir Ağları ile Tahmin ve Sınıflandırma Problemlerinin Çözümü İçin Arayüz Tasarımı. Acta Infologica; 2017; 1(2). <https://dergipark.org.tr/tr/download/article-file/398974>.
- [16] Minsky M, Papert S. Perceptrons: An Introduction to Computational Geometry. MIT Press; 1969
- [17] Kılıç O. Enerji iletim sistemlerinde yük tahmininin gerçekleştirilmesinde yeni bir yaklaşım (doktora tezi). İstanbul: Yıldız Teknik Üniversitesi; 1996.
- [18] Rumelhart DE, McClelland, JL. Parallel Distributed Processing: Explorations in the Microstructure of Cognition I and II. Cambridge: MIT Press; 1986. DOI: <https://doi.org/10.7551/mitpress/5236.001.0001>.
- [19] Ünal Z. Derin Öğrenme, 2. baskı. Ankara: Nobel Bilimsel Eserler; 2022.

- [20] Çelik Ş. Yapay Sinir Ağlarının Farklı Aktivasyon Fonksiyonlarında Uygulaması: Türkiye’de Ördek Popülasyonu Öngörüsü. OKU Fen Bilimleri Enstitüsü Dergisi; 2022; 5(2): 800–811. <https://dergipark.org.tr/tr/download/article-file/1956827>.
- [21] ResearchGate. Knowledge-Based Semantic Embedding for Machine Translation [İnternet]. [erişim tarihi 03.11.2023]. https://www.researchgate.net/figure/An-illustration-of-the-RNN-based-neural-network-model-for-Chinese-to-English-machine_fig1_306093825.
- [22] Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult. IEEE Transactions on Neural Networks. 1994; 5(2); 157–166.
- [23] Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Computation; 1997. 9(8); 1735–1780. DOI: <https://doi.org/10.1162/neco.1997.9.8.1735>
- [24] Onan A. Türkçe Metin Madenciliği için Çalışan Bellek Bağlantıları Tabanlı Uzun Kısa Süreli Bellek Mimarisi. Avrupa Bilim ve Teknoloji Dergisi 2022. Sayı 34; 239-246 . DOI: 10.31590/ejosat.1080239
- [25] Turkish to English Translation Dataset [İnternet]. 2020 [erişim tarihi 11.10.2023]. <https://www.kaggle.com/datasets/seymasa/turkish-to-english-translation-dataset>
- [26] TR2EN Translation Dataset [İnternet]. 2023 [erişim tarihi 11.10.2023]. <https://raw.githubusercontent.com/aozengi/TR2EN/main/data/TR2EN.txt>