



ANATOLIAN JOURNAL OF SOCIAL SCIENCES AND EDUCATION

DOĞAL DİL İŞLEME ALANINDA DERİN ÖĞRENME YÖNTEMLERİNİN
SÖZDİZİMSEL ANALİZ UYGULAMALARINDAKİ ETKİNLİĞİ VE GELECEĞİ

EFFECTIVENESS AND FUTURE OF DEEP LEARNING METHODS IN SYNTAX
ANALYSIS APPLICATIONS IN NATURAL LANGUAGE PROCESSING

Meerim Ryspakova¹

Aygül Tursunova²

Öz

Makale, cümlelerin otomatik olarak ayrıştırılması alanındaki yenilikçi teknolojilere kapsamlı bir genel bakış sunmaktadır. Derin öğrenmenin (RNN, BiLSTM, transformatörler) kullanımı da dahil olmak üzere, geleneksel yaklaşımlardan modern sinir ağı mimarilerine kadar yöntemlerin evrimi ele alınmaktadır. Dikkat mekanizmasına (BERT, XLNet, RoBERTa) dayalı modellere ve bunların sözdizimsel analize uyarlanmasına özellikle dikkat edilir. Araştırma aynı zamanda grafik yöntemlerini (GNN, grafik transformatörleri), çok dilli ve diller arası yaklaşımları da kapsamaktadır. Anlamsal bilginin entegrasyonunu, aktif öğrenme yöntemlerini ve önceden eğitilmiş dil modellerinin kullanımını tartışır. Ayrıştırmanın kalitesini değerlendirmek için yeni ölçütler ve Evrensel Bağımlılıklar projesinin bu alandaki standardizasyon üzerindeki etkisi dikkate alınmaktadır. Çalışma, yöntemlerin kronolojik ve karşılaştırmalı analizini, sınıflandırılmasını ve eleştirel değerlendirmesini içeren literatürün sistematik bir incelemesine dayanmaktadır. Çeşitli kaynaklardan bilgilerin sentezi ve gelecekteki eğilimlerin tahmini gerçekleştirilir. Çalışma, NLP, dil bilimi ve yapay zekâ alanındaki araştırmacıların ve uygulayıcıların ilgisini çekmektedir ve otomatik ayrıştırmanın geliştirilmesine yönelik mevcut durum ve beklentiler hakkında kapsamlı bir genel bakış sunmaktadır.

Anahtar kelimeler: Hesaplamalı Dilbilim, Sözdizimsel Analiz, Doğal Dil İşleme, Derin Öğrenme, Sinir Ağları.

Abstract

The article presents a comprehensive review of innovative technologies in the field of automated syntactic sentence analysis. The evolution of methods from traditional approaches to modern neural network architectures, including the application of deep learning (RNN, BiLSTM, Transformers) is considered. Special attention is given to attention mechanism based models (BERT, XLNet, RoBERTa) and their adaptation for syntactic analysis. The study also covers graph-based methods (GNN, graph transformers), multilingual and cross-linguistic approaches. Semantic information integration, active learning techniques and the application of pre-trained language models are discussed. New metrics for assessing the quality of parsing and the impact of the Universal Dependencies project on standardisation in this area are discussed. The study is based on a systematic literature review including chronological and comparative analysis of methods, their categorisation and critical evaluation. Information from various sources is synthesised and future trends are predicted. The paper is of interest to researchers and practitioners in the fields of NLP, linguistics and artificial intelligence, providing a comprehensive overview of the current state-of-the-art and future prospects of automated parsing.

Keywords: Computational Linguistics, Syntactic Analysis, Natural Language Processing, Deep Learning, Neural Networks.

¹ Uluslararası Kırgızistan Üniversitesi , Öğretim Görevlisi, meerim.ryspakova@gmail.com, ORCID İD: 0000-0001-5689-1365

² İşenali Arabayev Kırgız Devlet Üniversitesi, Öğretim Görevlisi, aygulya2380@gmail.com, ORCID İD: 0000-0003-4483-2581

GİRİŞ

Bilinmeyen bir dilde bir kitap okuduğumuz zaman tek tek kelimeleri görebiliyoruz ama birbirleriyle nasıl ilişkili olduklarını ve bir bütün olarak cümlenin anlamının ne olduğunu anlamıyoruz. Ayırıştırmanın kurtarmaya geldiği yer burasıdır. Ayırıştırma olarak da bilinen cümle ayırıştırma, doğal dil işleme (NLP) alanında temel bir görevdir. Makine çevirisi, bilgi çıkarma ve duygu analizi gibi birçok uygulama için kritik olan metnin yapısını ve anlamını anlamada temel bir rol oynar.

Sözdizimsel analiz, bir cümlenin dilbilgisel yapısını belirleme sürecidir. Bilgisayarların bir cümledeki kelimelerin birbirleriyle nasıl ilişkili olduğunu, her kelimenin oynadığı rolü ve bir ifadenin genel anlamının nasıl oluştuğunu "anlamasına" yardımcı olur. Bu, okuldaki üyelere göre cümleleri ve konuşma bölümlerini nasıl analiz ettiğimize benzer; ancak artık bilgisayarlar bu işi yapıyor. Bu neden gerekli? Ayırıştırma, doğal dil işlemenin birçok alanının merkezinde yer alır:

1. Makine çevirisi: Bir cümleyi bir dilden diğerine doğru bir şekilde çevirmek için bilgisayarın cümlenin yapısını anlaması gerekir.
2. Sesli asistanlar: Kullanıcı komutlarının doğru yorumlanması için.
3. Duygu Analizi: Bir metnin duygusal tonunu belirlemek için cümle içindeki kelimelerin nasıl ilişkili olduğunu anlamak önemlidir.
4. Bilgi çıkarma: metinlerden otomatik veri toplamak için.

Bu yeni teknolojilerin anlaşılması, yalnızca hesaplamalı dilbilim araştırmaları için değil, aynı zamanda daha gelişmiş dil işleme sistemleri geliştirmeye çalışan yazılım mühendisleri için de büyük önem taşımaktadır. Daha karmaşık ve hassas ayırıştırma tekniklerine yönelmek, makinelerin insan dilini daha iyi anlaması ve doğru bir şekilde üretmesi için önemli bir ilerleme sağlamaktadır. Son yıllarda yapay zekâ ve makine öğrenimi alanındaki gelişmeler, dil ayırıştırma süreçlerinde önemli atılımlara olanak tanımıştır. Bu yazıda, cümleleri otomatik olarak ayırıştırmaya yönelik en yenilikçi ve etkili yaklaşımlar ele alınacaktır.

2. Sinir ağlarının ve derin öğrenmenin sentaktik çözümleme için uygulanması

Derin öğrenmedeki modern ilerlemeler, metinden özellikleri otomatik olarak çıkarabilen ve karmaşık sözdizimsel yapılar oluşturabilen sinir ağlarının geliştirilmesine yol açmıştır. Sinir ağları, büyük miktarda veriden öğrenme, kelimeler arasındaki karmaşık bağımlılıkları belirleme ve insan beyninin işleyişini modellemeye dayalı yapay zeka alanında ileri bir teknolojiyi temsil etme yetenekleri nedeniyle görevlerin ayırıştırılmasında yüksek verimlilik göstermiştir. Tarihleri, yapay sinir ağlarına ilişkin ilk kavramsal fikirlerin ortaya çıktığı 20. yüzyılın ortalarında başladı. İlk önemli ilerlemelerden biri, 1957'de Frank Rosenblatt tarafından önerilen ve sonraki yıllarda daha karmaşık sinir ağı modellerinin geliştirilmesinin temelini oluşturan algılayıcıydı. Çoğunlukla Google, Facebook, Microsoft, Amazon gibi büyük teknoloji şirketlerinin yanı sıra akademik ve bilimsel kurumlardaki makine öğrenimi uzmanları, araştırmacıları ve geliştiricileri tarafından kullanılırlar. Gelecekte, daha karmaşık ve akıllı sistemler oluşturmak için grafik modelleri ve olasılıksal yöntemler gibi diğer yapay zeka alanlarıyla etkileşime girmeyi bekleyebiliriz.

Sinir ağları, insan beyninin işleyişinden ilham alan bilgisayar modelleridir. Derin öğrenme, sinir ağlarını büyük miktarda veri üzerinde eğitmek için kullanılan bir yöntemdir.

2.1. Tekrarlayan sinir ağları (RNN)

Diyelim ki cümleyi kelime kelime okudunuz. Her yeni kelimeyi öncekileri hesaba katarak algıyorsunuz. Tekrarlayan sinir ağları tam olarak bu şekilde çalışır; bilgileri sırayla işler ve önceki kelimelerin "hafızasını" korurlar. LSTM (Uzun Kısa Süreli Bellek) ve GRU (Gated Recurrent Unit) gibi RNN çeşitlerinin özellikle etkili olduğu kanıtlanmıştır. Uzun kelime dizileri üzerinden önemli bilgileri "hatırlamada" daha iyidirler. 2014 yılında araştırmacılar Chen ve Manning, sinir ağlarına dayalı ilk başarılı ayırıştırıcılardan birini yarattılar (Chen & Manning, 2014). Sistemleri, her bir kelimeyi öncekiler bağlamında analiz ederek bir cümledeki kelimelerin birbirleriyle nasıl ilişkili olduğunu tahmin etti.

Çift Yönlü LSTM'ler (BiLSTM'ler)

Normal RNN'ler, metni soldan sağa doğru bir insan gibi sıralı bir şekilde işler. Ancak bir cümlenin yapısını anlamak için genellikle her iki yönden gelen bağlamın dikkate alınması gereklidir. Bu sorunu çözmek amacıyla BiLSTM'ler, metni hem ileri hem de geri yönde analiz eder. 2016 yılında Kiperwasser ve Goldberg, bağlama duyarlı kelime temsilleri oluşturmak için BiLSTM'yi kullanmayı önererek ayırıştırma doğruluğunu önemli

ölçüde artırmıştır. Bu çalışma, ayrıştırma görevlerinde bağlamı çok yönlü bir şekilde değerlendiren kelime temsillerinin etkinliğini ortaya koymuştur. Nispeten basit BiLSTM tabanlı mimariler, bir kelimenin geniş bağlamı dikkate alındığında, karmaşık sözdizimsel bağımlılıkları etkili bir şekilde yakalayabilmiştir. Model, karmaşık işlemlere veya büyük miktarda verilere ihtiyaç duymadan birden fazla dilde tutarlı performans göstermiştir. Bu araştırma aynı zamanda dikkat mekanizmalarının ayrıştırma süreçlerine entegrasyonuna zemin hazırlayarak, daha gelişmiş dönüştürücü tabanlı modellerin ortaya çıkmasına katkı sağlamıştır (Kiperwasser & Goldberg, 2016).

Derin sinir ağları

Bilgi işlem gücünün artmasıyla birlikte daha karmaşık, birçok katmandan oluşan "derin" sinir ağlarının kullanılması mümkün hale geldi. Bu, bir uzmanın bir sorunu genelden özele doğru çeşitli düzeylerde analiz etmesine benzer. 2018 yılında Ma liderliğindeki bir araştırmacı ekibi, Stack-Pointer Networks (Ma et al., 2018) adı verilen bir sistem geliştirdi. Bir kelimenin bağlamını her iki tarafta da dikkate alabilen karmaşık sözdizimsel yapılar oluşturmak için bir "dikkat" mekanizmasıyla birlikte derin BiLSTM ağlarını kullanır.

Dönüştürücüleri ve dikkat temelli modelleri kullanmak

2017'de yeni bir sinir ağı mimarisi olan - transformatörler ortaya çıkmıştır. Ayrıştırma da dahil olmak üzere doğal dil işlemede gerçek bir devrim yaptılar.

Dikkat mekanizması

Transformatörlerin ana fikri dikkat mekanizmasıdır. Karmaşık bir cümle okuduğunuzu hayal edin. Anlamını daha iyi anlamak için farklı kelimelere odaklanarak bazı kısımlarını birkaç kez tekrar okuyabilirsiniz. Dikkat mekanizması da benzer şekilde çalışır; çıktı verilerinin her bir ögesini oluştururken modelin girdi verilerinin farklı bölümlerine "odaklanmasını" sağlar.

BERT ile sözdizimsel analiz

BERT (Bidirectional Encoder Representations from Transformers), transformatörlere dayalı bir model olup NLP dünyasında büyük bir çığır açmıştır. Bu model, cümlelerdeki kelimelerin bağlamını "anlayabilmek" için devasa miktarda metin üzerinde önceden eğitilir. 2018 yılında Kitaev ve Klein, BERT'in self-attention (kendine dikkat) mekanizmasını kullanarak son derece hassas bir sözdizimsel ayrıştırıcı (syntax parser) geliştirmişlerdir (Kitaev & Klein, 2018). Bu model, cümlelerdeki kelimeler arasındaki karmaşık bağımlılıkları daha iyi yakalayabilmiştir.

XLNet ve RoBERTa

BERT'in başarısından sonra XLNet (Yang et al., 2019) ve RoBERTa (Liu et al., 2019) gibi daha gelişmiş modeller ortaya çıktı. Ayrıştırmada daha da yüksek doğruluk elde etmek için gelişmiş eğitim ve veri işleme tekniklerini kullanırlar.

Transformatörlerin sözdizimsel görevlere uyarlanması

Araştırmacılar, transformatörleri özellikle sözdizimsel analiz görevlerine uyarlamak için de çalışıyorlar. Örneğin, 2020 yılında Zhou ve meslektaşları, tekrarlayan sinir ağlarının (ardışık verileri işleme yeteneği) ve dikkat mekanizmasının (girdi verilerinin önemli kısımlarına odaklanma yeteneği) avantajlarını birleştiren bir model önerdiler (Zhou et al., 2020).

Grafik tabanlı ayrıştırma yöntemleri

Grafik yöntemleri ayrıştırmaya yönelik başka bir yenilikçi yaklaşımı temsil eder. Bu bağlamda bir cümle, kelimelerin köşeleri ve sözdizimsel bağlantıların kenarları olduğu bir grafik olarak görülür.

Grafik Sinir Ağları (GNN)

Grafik sinir ağları, grafikler biçimindeki verilerle çalışabilen özel bir sinir ağları türüdür. Yalnızca tek tek kelimelerin özelliklerini değil aynı zamanda aralarındaki bağlantıların yapısını da dikkate alabilirler. 2019 yılında Ji, Cao ve Huang, Grafik Tabanlı Sinirsel Bağımlılık Ayrıştırıcısı (Ji, Cao, & Huang, 2019) adlı bir model sundular. Bu model, bir cümlenin bağımlılık yapısını öğrenmek için GNN'yi kullanarak karmaşık sözdizimsel yapılarla daha iyi başa çıkabilmesini sağlar.

Bu tür yöntemler, kelimeler arasındaki bağımlılıkları temsil etmek ve analiz etmek için bir grafik yapısı kullanan bağımlılık yöntemlerinin geliştirilmiş halidir. Bu tür modellerde her kelime bir düğümü temsil eder

ve düğümler arasındaki yaylar sözdizimsel bağlantıları yansıtır. Grafik yöntemlerinin kullanılması, karmaşık sözdizimsel yapıların verimli bir şekilde işlenmesini ve analizin doğruluğunun iyileştirilmesini mümkün kılar. Ayrıştırmadaki grafik yöntemleri, bir cümledeki kelimeler arasındaki sözdizimsel bağımlılıkları temsil etmek ve analiz etmek için grafik yapısını kullanan gelişmiş bir yaklaşımdır. Bu yöntemlerde, her kelime grafikteki bir düğümü temsil eder ve düğümler arasındaki yönlendirilmiş kenarlar sözdizimsel ilişkileri ve bağımlılıkları yansıtır.

Grafik yöntemlerinin yönleri:

- Bağımlılık Görünümü

Kelimeler arasındaki sözdizimsel ilişkiler (örneğin, konu-yüklem, bağımlı-temel) bir grafiğin yönlendirilmiş kenarları olarak modellenir. Bu, cümle yapısında hangi kelimelerin anahtar rol oynadığını doğru bir şekilde belirlemenizi sağlar.

- Derin yapı analizi

Grafik yöntemleri, cümlelerin karmaşık ve çok düzeyli sözdizimsel yapılarını işleme yeteneğine sahiptir. Yalnızca kelimeler arasındaki doğrudan bağımlılıkları değil, aynı zamanda daha karmaşık ilişkileri de tanımlamanıza ve analiz etmenize olanak tanır.

- Analiz algoritmaları

Grafik yapılarıyla çalışmak için, grafik geçiş algoritmaları (örneğin, DFS ve BFS), en kısa yol algoritmaları (örneğin, Dijkstra), minimum yayılan ağaç algoritmaları (örneğin, Prim'in veya Kruskal'ın algoritması) gibi çeşitli algoritmalar kullanılır. Bu algoritmalar bir cümledeki yapısal bağımlılıkları etkili bir şekilde tanımlamaya ve analiz etmeye yardımcı olur.

- NLP görevlerinde uygulama

Grafik yöntemleri, makine çevirisi, bilgi çıkarma, metin özetleme ve anlamsal ilişki analizi gibi çeşitli doğal dil işleme görevlerinde aktif olarak kullanılmaktadır. Sözdizimsel yapıların daha derin anlaşılması yoluyla analizin kalitesini ve doğruluğunu artırır.

Grafik transformatörleri

Grafik transformatörleri, grafik sinir ağlarının ve transformatör mimarisinin avantajlarını birleştirir. 2020 yılında Wu liderliğindeki bir grup araştırmacı, Grafik Transformatör Ağları (GTN) (Wu et al., 2020) adı verilen bir model önerdi. GTN'ler, bir cümledeki kelimeler arasındaki karmaşık ilişkileri otomatik olarak tespit etme ve kullanma yeteneğine sahiptir; bu, özellikle uzun ve karmaşık cümleleri analiz etmek için kullanışlıdır.

Ayrıştırmada çok dilli yaklaşımlar

Küreselleşmenin artmasıyla birlikte farklı dillerdeki metinleri analiz edebilen sistemlere olan ihtiyaç da artıyor. Çok dilli ayrıştırma modelleri bu ihtiyacı karşılamayı amaçlamaktadır.

Diller arası bilgi aktarımı

Bu yaklaşım, farklı dillerdeki sözdizimsel yapıların pek çok ortak noktaya sahip olduğu fikrine dayanmaktadır. Bir dilde eğitilen bir model diğer dillerle çalışacak şekilde uyarlanabilir. 2019'da Ahmad ve arkadaşları, bağlamsallaştırılmış kelime yerleştirmeleri kullanarak diller arası ayrıştırma üzerine bir çalışma sundular (Ahmad et al., 2019). Yaklaşımları, her dil için ayrı bir model eğitmeye gerek kalmadan, birden fazla dildeki cümleleri ayrıştırabilen bir ayrıştırıcı oluşturmalarına olanak sağladı.

Transformatörleri temel alan çok dilli modeller

Birçok dildeki metinler üzerinde önceden eğitilmiş büyük dil modelleri, çok dilli ayrıştırma görevlerinde de mükemmel sonuçlar verir. Conneau ve diğerleri tarafından 2020'de sunulan XLM-RoBERTa modeli (Conneau et al., 2020) 100 farklı dildeki metinler üzerinde eğitilmiştir. Sınırlı kaynaklara sahip diller de dahil olmak üzere çok çeşitli diller için görevlerin ayrıştırılmasında yüksek verimlilik gösterir.

Anlamsal bilgiyi ayrıştırmaya entegre etme

Geleneksel olarak sözdizimsel ve anlamsal analiz ayrı görevler olarak görülmüştür. Ancak modern yaklaşımlar, dilin daha derin anlaşılması için bu yönleri birleştirmeyi amaçlamaktadır.

Sözdizimi ve anlambilimin işbirlikçi öğretimi

Bu yaklaşım, hem sözdizimsel hem de anlamsal görevler için bir modelin eş zamanlı olarak eğitilmesini içerir. 2018 yılında Swayamdipta ve diğerleri, Sözdizimsel Yapı İskeleleri (Swayamdipta et al., 2018) adlı bir model

sundular. Bu model, anlamsal analizi geliřtirmek için sözdizimsel bilgileri kullanır ve bunun tersi de cümlelerin yapısının ve anlamının daha doğru anlaşılmasına yol açar.

Anlamsal olarak zenginleştirilmiş kelime gösterimleri

Kelimelerin anlamsal olarak zenginleştirilmiş vektör temsillerinin kullanılması ayrıştırma kalitesini artırabilir. Kasai ve arkadaşlarının 2019'da yaptığı bir araştırma (Kasai et al., 2019), anlamsal bağlamı dikkate alan bağlamsallaştırılmış sözcük yerleřtirmelerin kullanılmasının sözdizimsel ayrıştırma doğruluğunu önemli ölçüde artırabildiğini buldu.

Ayrıştırma doğruluğunu artırmak için aktif öğrenme yöntemlerinin uygulanması

Aktif öğrenme, modelin kendisinin eğitim için en bilgilendirici örnekleri seçtiği ve sınırlı açıklama kaynaklarının verimli kullanımına olanak tanıyan bir stratejidir.

Belirsizliğe dayalı örnekleme

Bu yöntem, modelin en az güvenilir olduğu örneklerin açıklama için seçilmesini içerir. 2020 yılında Li, ve diğeri, belirsizliğe dayalı sözdizimsel ayrıştırma için aktif bir öğrenme yaklaşımı önerdi (Li et al., 2020). Yöntemleri, daha az açıklamalı veri kullanırken ayrıştırma doğruluğunu önemli ölçüde geliřtirdi.

Örnek çeşitliliği

Bu yaklaşım, farklı dilsel olguları kapsamak amacıyla açıklama için çeşitli örnekler seçmeyi amaçlamaktadır. Joshi ve arkadaşları tarafından 2021'de yapılan bir çalışma (Joshi et al., 2020), eğitim verilerindeki sözdizimsel yapıların çeşitliliğini en üst düzeye çıkarmaya dayanan bir örnekleme stratejisinin etkinliğini gösterdi.

Ayrıştırma için önceden eğitilmiş dil modellerini kullanma

GPT ve BERT gibi önceden eğitilmiş dil modelleri NLP alanında devrim yarattı ve ayrıştırma da bir istisna değil.

Önceden eğitilmiş modellerde ince ayar yapma

Bu yaklaşım, önceden eğitilmiş bir modelin ayrıştırma görevine ilişkin ek eğitimini içerir. 2019 yılında Kondratyuk ve Straka, birden çok dille çalışabilen genel amaçlı bir sözdizimi ayrıştırıcısı oluşturmak için BERT'in çok dilli bir versiyonunu kullanan UDify modelini (Kondratyuk & Straka, 2019) tanıttı.

Önceden eğitilmiş modellerden sözdizimsel bilgilerin çıkarılması

Bu yöntem, modelin eğitim öncesi süreçte edindiği gizli sözdizimsel bilginin ortaya çıkarılmasına odaklanır. Hewitt ve Manning tarafından 2019'da yapılan bir çalışma (Hewitt & Manning, 2019), BERT temsillerinin zaten basit doğrusal dönüşümler kullanılarak çıkarılabilecek önemli miktarda sözdizimsel bilgi içerdiğini buldu.

Ayrıştırmanın kalitesini değerlendirmek için yeni ölçümler

Ayrıştırma yöntemleri geliřtikçe, bunların kalitesini değerlendirmek için daha gelişmiş yöntemlere ihtiyaç duyulur.

Yapı Tabanlı Metrikler

Bu ölçümler yalnızca bireysel sözdizimsel bağlantıların doğruluğunu değil aynı zamanda sözdizimsel ağacın genel yapısını da dikkate alır. Evrensel Bağımlılıklar (UD): Evrensel Bağımlılıklar, diller arası tutarlı sözdizimsel bağımlılık ek açıklamaları geliřtirmeyi amaçlayan bir projedir. UD yalnızca bir açıklama şeması sağlamakla kalmaz, aynı zamanda ayrıştırma kalitesini değerlendirmek için bir dizi ölçüm de sağlar.

- Etiketli Ek Puanı (LAS): Bu ölçüm, modelin hem ana kelimeyi (sözdizimsel ebeveyn) hem de bağımlılık türünü doğru şekilde tahmin ettiği kelimelerin yüzdesini değerlendirir.
- Etiketlenmemiş Ek Puanı (UAS): Ekin türü ne olursa olsun, doğru tahmin edilen ana kelimeye sahip kelimelerin yüzdesini değerlendirir.
- Etiket Doğruluğu (LA): Doğru tahmin edilen bağımlılık türüne sahip kelimelerin yüzdesini ölçer.

Tek bir UD şeması kullanmak, farklı diller için ve diller arasında sözdizimsel ayrıştırıcıların kalitesinin karşılaştırılmasına olanak tanır. 2018 yılında Zeman ve diğeri (de Marneffe et al., 2021), UD verilerini ve ölçümlerini kullanarak çok dilli sözdizimsel ayrıştırma üzerine CoNLL 2018 Paylaşılan Görev yarışmasını düzenledi. Bu yarışma, ayrıştırmaya yönelik evrensel yaklaşımların geliřtirilmesinde önemli bir dönüm noktasıydı.

Evrensel Bağımlılıklar, birçok dil için birleşik bir açıklama şeması sağlayan önemli bir ayrıştırma projesidir.

2020'de Mille ve arkadaşları, ayrıştırma kalitesini değerlendirmek için SID (Sözdizimsel Bilgilendirici Bağımlılıklar) adı verilen yeni bir ölçüm önerdiler (Mille et al., 2018). Bu metrik, standart UD metriklerini tamamlayan çeşitli sözdizimsel bağımlılık türlerinin bilgi içeriğini dikkate alır.

Anlamsal olarak motive edilmiş metrikler

Bu ölçümler, ayrıştırmanın kalitesini bir cümlenin anlamının anlaşılması üzerindeki etkisi açısından değerlendirir. Slobodkin ve arkadaşları tarafından 2021'de yapılan bir araştırma (Slobodkin et al., 2021), doğal dil anlama görevlerindeki performansı iyileştirme becerisine dayalı olarak ayrıştırma kalitesini değerlendiren bir ölçüm önerdi.

Sonuç

Modern bilgi teknolojisi dünyasında, dilbilimsel araştırmalar giderek otomasyona ve bilgisayar yöntemlerinin kullanımına odaklanmaktadır. Sözdiziminin temel yönlerinden biri olan cümle analizi bu alanda en önemli rolü oynamaktadır. Cümle ayrıştırma alanı son yıllarda makine öğrenimi, sinir ağları ve derin öğrenmedeki ilerlemelere dayanan yeni teknolojilerin ortaya çıkmasıyla önemli ölçüde ilerleme kaydetti. Bu teknolojiler, ayrıştırma sorunlarını geleneksel yöntemlere göre daha fazla doğruluk, verimlilik ve esneklikle çözebilir. Yeni teknolojiler, karmaşık ve belirsiz cümlelerde bile kelimeler arasındaki sözdizimsel ilişkilerin daha doğru bir şekilde belirlenmesini mümkün kılmaktadır. Verilerden öğrenme yeteneği sayesinde cümle ayrıştırmanın gücünü ve verimliliğini büyük ölçüde artıran yeni teknolojiler, farklı dillere ve görevlere uyarlanabiliyor. Cümlelerin sözdizimsel yapısını daha iyi anlayarak daha doğru ve doğal çeviriler oluşturabilirler ve ayrıca arama motorlarının kullanıcı sorgularını daha iyi anlamasına ve daha alakalı arama sonuçları döndürmesine olanak tanır. Sinir ağları ve derin öğrenme modelleri, büyük miktarda veriyi geleneksel yöntemlere göre çok daha hızlı işleyebilir ve yalnızca sözdizimsel ilişkileri tanımlamakla kalmaz, aynı zamanda bir cümlenin anlamsal ve faydacı yönlerini de anlayabilir. Otomatik olarak kısa ve bilgilendirici metin pasajları oluşturur, önemli fikirleri vurgulayın ve mantıksal bağlantılarını koruyun. Yeni yöntemler, çeşitli dillerin sözdizimini incelemek ve karmaşık kalıplarını belirlemek için yeni fırsatlar yaratıyor. Genel olarak, yeni teknolojiler cümle ayrıştırmada devrim yaratarak dili daha iyi anlamamıza ve daha verimli doğal dil işleme sistemleri geliştirmemize olanak tanıyor.

Ayrıştırma, basit kural tabanlı algoritmalarından, dil yapılarının ince nüanslarını yakalayabilen karmaşık sinir modellerine kadar uzun bir yol kat etti. Derin öğrenmeyi, dönüştürücüleri ve grafik yöntemlerini kullanan modern yaklaşımlar, çeşitli dillerin sözdizimini analiz etmede etkileyici bir doğruluk elde edebilir. Anlamsal bilgilerin entegrasyonu, aktif öğrenme yöntemlerinin kullanılması ve önceden eğitilmiş dil modellerinin kullanılması bu alanda yeni ufuklar açmaktadır. Ek olarak, daha iyi değerlendirme ölçütlerinin geliştirilmesi, farklı yaklaşımların güçlü ve zayıf yönlerinin daha iyi anlaşılmasına olanak tanır.

Ayrıştırmanın geleceği muhtemelen dilin farklı yönlerinin daha fazla entegrasyonunu içerecek ve birden fazla dili ve etki alanını işleyebilecek daha güçlü ve genel amaçlı modeller yaratacaktır. Bu ilerlemeler şüphesiz doğal dil işleme sistemlerinin ve genel olarak yapay zekanın gelişimine önemli bir katkı sağlayacak ve bizi insan konuşmasını gerçekten anlayabilen ve üretebilen makineler yaratmaya daha da yaklaştıracak. Kaynakların kısıtlı olduğu diller için özellikle önemli olan, daha az etiketli veri gerektiren yöntemlerin geliştirilmesine büyük olasılıkla özel önem verilecektir. Ayrıca, ayrıştırmanın diğer doğal dil işleme görevleriyle daha fazla entegre olmasını bekleyebiliriz, bu da daha bütünsel dil anlama sistemlerine yol açar.

Kaynakça (References)

- Ahmad, W. U., Zhang, Z., Ma, X., Hovy, E., Chang, K. W., & Peng, N. (2019). On difficulties of cross-lingual transfer with order differences: A case study on dependency parsing. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.
- Chen, D., & Manning, C. D. (2014). A fast and accurate dependency parser using neural networks. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*.
- Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., ... & Stoyanov, V. (2020). Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*.
- de Marneffe, M. C., Manning, C. D., Nivre, J., & Zeman, D. (2021). Universal dependencies. *Computational Linguistics*, 47(2), 255-308.

- Hewitt, J., & Manning, C. D. (2019). A structural probe for finding syntax in word representations. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.
- Ji, T., Cao, Y., & Huang, L. (2019). Graph-based neural dependency parsing with latent structural predictions. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*.
- Joshi, M., Chen, D., Liu, Y., Weld, D. S., Zettlemoyer, L., & Levy, O. (2020). SpanBERT: Improving pre-training by representing and predicting spans. *Transactions of the Association for Computational Linguistics*.
- Kasai, J., Cahn, J., Tanaka, R., & Choi, Y. (2019). Syntax-aware neural semantic role labeling. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.
- Kiperwasser, E., & Goldberg, Y. (2016). Simple and accurate dependency parsing using bidirectional LSTM feature representations. *Transactions of the Association for Computational Linguistics*, 4, 313-327.
- Kitaev, N., & Klein, D. (2018). Constituency parsing with a self-attentive encoder. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*.
- Kondratyuk, D., & Straka, M. (2019). 75 languages, 1 model: Parsing universal dependencies universally. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*.
- Li, Z., Li, J., Tang, D., Chen, B., & Li, J. (2020). Deep active learning for dependency parsing. In *Proceedings of the 28th International Conference on Computational Linguistics*.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692*.
- Ma, X., Hu, Z., Liu, J., Peng, N., Neubig, G., & Hovy, E. (2018). Stack-pointer networks for dependency parsing. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*.
- Mille, S., Belz, A., Bohnet, B., Graham, Y., Pitler, E., & Wanner, L. (2018). The first multilingual surface realisation shared task (SR'18): Overview and evaluation results. In *Proceedings of the First Workshop on Multilingual Surface Realisation*.
- Slobodkin, A., Bhattacharyya, P., & Ungar, L. (2021). A unified feature attribution framework for interpreting dependency parsers. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*.
- Swayamdipta, S., Thomson, S., Lee, K., Zettlemoyer, L., Eisenstein, J., & Smith, N. A. (2018). Syntactic scaffolds for semantic structures. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*.
- Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2020). A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*.
- Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. (2019). XLNet: Generalized autoregressive pretraining for language understanding. In *Advances in Neural Information Processing Systems*.
- Zhou, J., Zhang, Z., Zhao, H., & Zhang, S. (2020). LIMIT-BERT: Linguistic informed multi-task BERT. In *Findings of the Association for Computational Linguistics: EMNLP 2020*.