

# Derin Öğrenme Kullanılarak Kural Tabanlı Varlık Tanıma



Derman Akgol, Necva Bolucu, Salih Tuc  
{dermanakgol,necvaa,salihtuc0}@gmail.com

## Özetçe

Kural Tabanlı Varlık Tanıma (Named Entity Recognition – NER) Doğal Dil İşleme (Natural Language Processing – NLP) uygulamalarında çok önemli bir ön işleme adımıdır. Son zamanlarda, Derin Öğrenme (Deep Learning) yöntemleri elle uygulanan (hand-crafted) özellikler ve bilgi kaynakları (knowledge resources) yerine veri kümesinden öğrenilen özellikler kullanıldığı için NLP uygulamalarında kullanılmaktadır. Çift Yönlü Uzun Kısa Süreli Bellek (Bidirectional Long Short Term Memory – BiLSTM) önceki ve sonraki birkaç nörondan gelen bilgiler ile girdi sayısını arttırmayı sağlayan yaklaşımdır. Bu çalışmada Chiu ve Nichols [1]’in çalışmasındaki BiLSTM-CNN modeline ekstra sözcük özellikleri ve ilgi (attention) mekanizması eklenilerek genişletilmiştir.

## Giriş

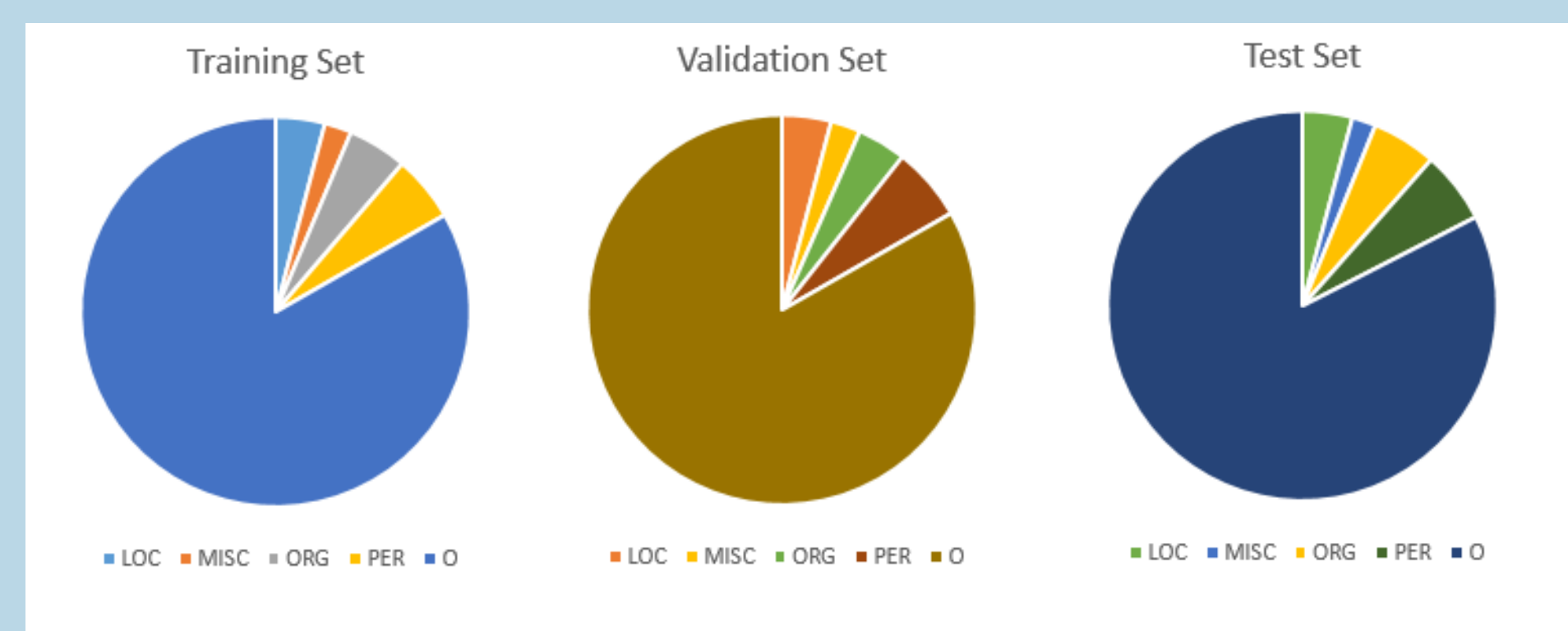
Kural Tabanlı Varlık Tanıma (NER) yer, kişi, zaman, kurum gibi benzer özelliklerine göre sınıflandırılan kelime gruplarını tanımlayan ve dile bağımlı veya dilden bağımsız olarak sınıflandıran bir sistemdir [2]. NER sistemi çok büyük çalışma kümelerinde uygulanabilmektedir. Elle uygulanan kurallar (handcrafted rules) ve yer adları dizini (gazetteers), sözcük türlerindeki (Part-of-Speech – PoS) etiketler (tagger) gibi kurallara göre veri kaynakları NER sistemindeki başarıyı etkileyen faktörlerdir [3]. Aşağıdaki cümle NER’e göre bir cümlenin nasıl etiketlendiğinin bir örneğidir:

**Örnek:** “ **Vancouver** {konum} is a coastal seaport city on the mainland of **British Columbia** {konum}. The city’s mayor is **Gregor Robertson** {kişi}.”

Bu çalışmada Chiu ve Nichols [1]’in çalışmasındaki BiLSTM-CNN modeline ekstra sözcük özellikleri ve ilgi (attention) mekanizması eklenilerek genişletilmiş ve daha yüksek bir başarı hedeflenmiştir.

## Veri Seti

Veri Seti olarak ConLL-2003 NER veri seti kullanılmıştır. [4]



## İlgi (Attention)

İlgi (Attention) mekanizması NLP alanında çok fazla kullanılmaya başlamıştır.

$$\alpha_{ts} = \frac{\exp(\text{score}(h_t, \hat{h}_s))}{\sum_{s'} \exp(\text{score}(h_t, \hat{h}_{s'}))} \quad (1)$$

$$c_t = \sum_s \alpha_{ts} \hat{h}_s \quad (2)$$

$$a_t = f(c_t, h_t) = \tanh(W_c [c_t; h_t]) \quad (3)$$

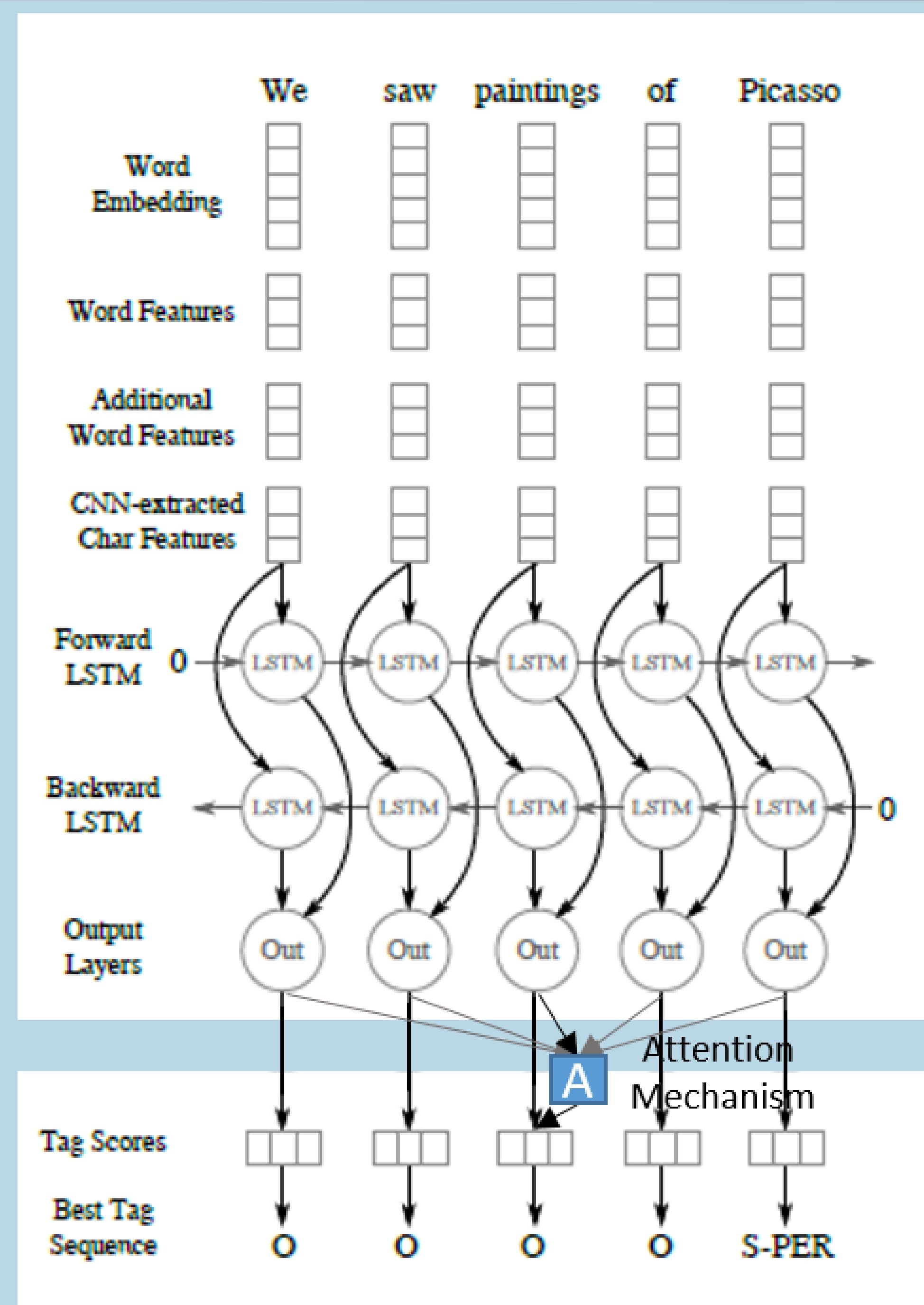
$$\text{score}(h_t, \hat{h}_s) = h_t^T W \hat{h}_s \quad (4)$$

öyle ki  $\alpha$  ilgi (attention) vektörü,  $c$  içerik vektörü (context vector) ve skor fonksiyonu (score function) ise gizli durumların (hidden states) 'kareler ortalaması (mean-squared)' versiyonudur. BiLSTM'in amaç gizli durumu ilgi ağırlıklarını üretmek için tüm durumlarla karşılaştırılmıştır. Final ilgi vektörünü elde etmek için ilgi ağırlıklarına dayanan kaynak vektörleri (context vector) hesaplanmıştır ve amaç gizli durumla birleştirilmiştir.

## Deneyler

Models	F1-scores
BiLSTM-CNN + emb	89.40
Att-BiLSTM-CNN + emb	89.90
BiLSTM-CNN + emb - with PoS tag	89.80
Att-BiLSTM-CNN + emb - with PoS tag	90.10
BiLSTM-CNN + emb - with suffix - attention	89.70
Att-BiLSTM-CNN + emb - with suffix - attention	89.30
BiLSTM-CNN + emb - with PoS tag+suffix	90.00
Att-BiLSTM-CNN + emb - with PoS tag+suffix	89.90
BiLSTM + emb + lex [1]	91.62
CNN-BiLSTM-CRF [5]	91.21
JERL	91.20
Baseline + Gaz + LexEmb	90.90
W500 + P125 + P64	90.90
Word-class Model	90.80
BiLSTM-CRF	90.10

## Model



## Referanslar

- [1] Jason PC Chiu, Eric Nichols: *Named entity recognition with bidirectional LSTM-CNNs*, arXiv preprint arXiv:1511.08308, (2015)
- [2] Daniele Bonadiman, Aliaksei Severyn, Alessandro Moschitti: *Deep neural networks for named entity recognition in Italian*, CLiC it, (2015)
- [3] V Rudra Murthy, PushpakLazslo Bhattacharyya: *A Deep Learning Solution to Named Entity Recognition*, International Conference on Intelligent Text Processing and Computational Linguistics (2016)
- [4] Tjong Kim Sang, Erik F and De Meulder, Fien: *ntroduction to the CoNLL-2003 shared task: Language-independent named entity recognition*, Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4 (2003)
- [5] Xuezhe Ma, Eduard Hovy: *End-to-end sequence labeling via bi-directional lstm-cnns-crf*, arXiv preprint arXiv:1603.01354, (2016)
- [6] Gang Luo, Xiaojiang Huang, Chin-Yew Lin, Zaiqing Nie: *Joint entity recognition and disambiguation*, Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (2015)
- [7] Alexandre Passos, Vineet Kumar, Andrew McCallum: *Lexicon infused phrase embeddings for named entity resolution*. arXiv preprint arXiv:1404.5367 (2014)
- [8] Dekang Lin, Xiaoyun Wu: *Phrase clustering for discriminative learning*. Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2 (2009)