

Morfem Düzeyinde Dikkat Mekanizması Kullanarak Türkçe Dizi Etiketleme

Using Morpheme-level Attention Mechanism for Turkish Sequence Labelling

Yasin Eşref
Hacettepe Üniversitesi
Bilgisayar Mühendisliği B.lümü
Ankara, Türkiye
yasinesref@gmail.com

Burcu Can
Hacettepe Üniversitesi
Bilgisayar Mühendisliği i B.lümü
Ankara, Türkiye
burcucan@cs.hacettepe.edu.tr

Özetçe—Derin öğrenmenin doğal dil işleme problemlerinde kullanılmaya başlaması ile bu alandaki birçok problemin çözümünde ciddi iyileşmeler olmuştur. Kelime dizilerini etiketleme problemlerinin de derin öğrenme yöntemleri ile sıkça çalışılan konulardan birisi olduğu görülmektedir. Bu çalışmada derin öğrenme yöntemleri kullanarak Türkçe üzerinde kelime dizisi etiketleme problemi için bir mimari önererek, karakter, morfem ya da sözcük gösterimlerinin bu problem kapsamındaki etkisini inceliyoruz. Türkçe gibi sondan eklemeli dillerde kelimenin bir bütün olarak modellenmesi seyreklik problemi de beraberinde getirmektedir. Bundan ötürü kelimeyi bir bütün olarak ele almaktansa karakterleri üzerinden bir kelimeyi ifade etmek ya da morfem ve morfemin etiket bilgisini hesaba katmak kelime hakkında daha detaylı bilgi vermekte ve seyreklik problemini de ortadan kaldırmaktadır. Bu çalışmada mevcut derin öğrenme modellerini Türkçe için farklı kelime veya alt-kelime düzeyinde gösterimler kullanarak uyguladık. Kelimelerin morfem bilgisinin kullanılmasının Türkçede dizi etiketleme sonuçlarını iyileştirdiği gözlenmiştir.

Anahtar Kelimeler — dizi etiketleme, sözcük türü işaretleme, varlık ismi tanıma, derin öğrenme, doğal dil işleme, dikkat ağıları

Abstract— With deep learning being used in natural language processing problems, there have been serious improvements in the solution of many problems in this area. Sequence labeling is one of the most frequently studied subject using deep learning methods. In this study, we examine the effects of character, morpheme, and word representations on sequence labelling problems by proposing a model for the Turkish language by using deep neural networks. Modeling the word as a whole in agglutinative languages such as Turkish causes sparsity problem. Therefore, rather than handling the word as a whole, expressing a word through its characters or considering the morpheme and morpheme label information gives more detailed information about the word and mitigates the sparsity problem. In this study, we applied the existing deep learning models using different word or sub-word representations for Turkish. The sequence labelling results in Turkish has improved with the addition of morpheme information of the words.

Keywords — sequence labeling; pos tagging; named entity recognition; deep learning; natural language processing, attention mechanism

I. GİRİŞ

Kelimelerin herhangi bir bağlamdaki kullanımlarına karşılık gelen isim, sıfat, eylem gibi sözdizimsel kategorilerinin bulunması (POS tagging), bir cümlede geçen varlık isimlerinin bulunması (NER), bir cümledeki isim ve fiil tamlamalarının bulunması (Chunking) ve cümledeki eksik kelimeyi bulma gibi doğal dil işleme (NLP) problemleri genel olarak dizi etiketleme problemleri olarak ele alınır. Dizi etiketleme problemleri birçok doğal dil işleme uygulamasının ilk adımı olduğundan bu problemlerin çözümlerinde yapılacak iyileştirmeler bu uygulamaların sonuçlarının da iyileşmesine katkı sağlayacaktır. Bu çalışma kapsamında Türkçe için dizi etiketleme problemlerinde mevcut en iyi sonuçların üzerine çıkacak bir model önerilmektedir.

Derin öğrenme yöntemleri yaygın olarak kullanılmaya başlamadan önce dizi etiketleme problemlerinin çözümünde istatistiksel yöntemler [1] daha çok kullanılıyordu. 2011’de Collobert ve diğ., [2] basit ama etkili bir sinir ağı önerdi. Kelimeler komşu kelimeleri de göz önünde bulundurularak bağımsız olarak etiketleniyordu. Sonraki yıllarda dizi etiketleme problemlerinde derin öğrenme yöntemleri kullanılmaya başlandı ve sonuçlarda ciddi iyileşmeler oldu. [3] ve [4] çalışmalarında dizi etiketlerini bulmak için girdi olarak kelime vektörlerini alan, iki yönlü, tekrarlayan sinir ağının (Bi-RNN) bir varyasyonu olan uzun-kısa vadeli bellek ağını (Bi-LSTM) kullanıldı. Varlık isimlerinin bulunması için tasarlanan benzer bir modelde [5] girdi olarak kelime vektörlerini kullanmanın yanı sıra evrişimli sinir ağı (CNN) ile çıkarılmış karakter özelliklerini de kullanıldı. [6] ve [7] çalışmalarında ise aynı modelin çıktısını Koşullu Rasgele Alandan (Conditional Random Field) geçirerek sonuçların daha da geliştirilmesi sağlandı. [8] çalışmasında ise benzer bir model için karakter özelliklerinin çıkarılmasında CNN yerine Bi-LSTM kullanıldı. Bir başka çalışmada [9] ise varlık isimlerini bulmak için kelime vektörleri yerine sadece karakter özellikleri LSTM ağında eğitildi. (Güngör ve diğ.,

2019) varlık isimlerini bulmada kelime ve karakter vektörlerine ilaveten morfem etiketlerine ait vektörel gösterimleri de LSTM ile eğitiler.

Bu çalışmalar üzerine yapılan değerlendirmeler kelime vektörüne ilaveten karakter ve morfem vektörleri kullanmanın, LSTM çıktısını CRF'ten geçirmenin dizi etiketlemede sonuçları iyileştirdiğini gösterdi. [10]

Son dönemde makine çevirisinden [11] doküman sınıflandırmaya [12] kadar dil üzerine yapılan birçok çalışmada modele eklenen dikkat mekanizmasının (Attention Networks) sonuçlardaki başarıyı artırdığı görülmüştür. Dikkat mekanizması ile verilen bir dizinin hangi kısımlarına yoğunlaşılması gerektiği ve böylece hangi sözcüklerin sonuç üzerinde daha fazla katkısı olduğu öğrenilebilmektedir. Böylece o anki çıktı oluşturulurken bütün sözcükler eşit olarak dikkate alınmamakta, bazılarının o anki çıktı üzerindeki önemine göre katkısı belirlenmektedir. Yang ve diğ. [13] çalışmasında doküman sınıflandırma için hiyerarşik bir dikkat ağı kullandı. Doküman vektörleri cümle vektörlerinin eğitilmesiyle, cümle vektörleri ise kelime vektörlerinin eğitilmesiyle oluşturuldu. Dikkat mekanizması sayesinde cümle vektörlerinin oluşmasında her kelimenin ağırlığı ve benzer şekilde doküman vektörlerinin oluşmasında da her cümlenin ağırlığı tahmin edilerek doküman vektörü oluşturuldu.

Bu birbiri ile çok benzer derin öğrenme modellerini kimi araştırmacılar sadece belirli bir dizi etiketleme problemini çözmek için geliştirirken kimi araştırmacılar ise modeli tüm problem tipleri için uygulamayı tercih etmişlerdir. Mesela Chiu ve diğ. [5] çalışmasında sadece isim varlıklarının bulunması problemini çözmeye çalışmış ve bu amaçla ayrıca bir isim sözlüğü kullanmıştır. Ma ve diğ. [6] ise hiçbir ek özellik kullanmadan oluşturdukları modeli 3 farklı problem (sözcük türü işaretleme, varlık isimlerini bulma, tamlamaları bulma) için kullanmışlardır.

Bu çalışma kapsamında biz de benzer bir derin ağ modelini Türkçe için uyguladık. Morfem seviyesindeki özellikleri de modele eklemek için dikkat mekanizması kullanan hiyerarşik bir model oluşturduk. Oluşturulan bu modeli sözcük türü işaretleme ve varlık isimlerini bulma problemi olmak üzere iki farklı dizi etiketleme problemi için test ettik. Sonuçlarımızın mevcut sonuçlardan daha iyi olduğunu gözlemledik. Bizim oluşturduğumuz modelin önceki modellerden en büyük farkı kelimelerin morfemlerini dikkat mekanizması kullanarak etiketleme sırasında hesaba katıyor olmasıdır. Böylece bütün morfemlerin eşit olarak modele dahil edildiği bir modelden ziyade, morfemlerin dikkat mekanizması ile etiketleme probleminin katkısı uyarınca modele dahil edilmesi bu çalışmanın katkılarının birisidir.

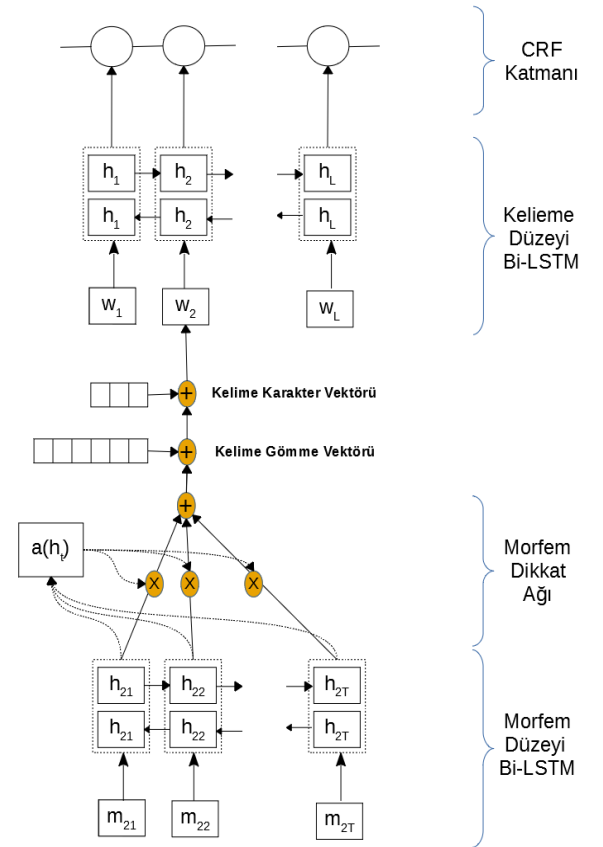
Türkçenin sondan eklemeli bir dil olmasından ötürü tek bir kökten onlarca farklı kelime türetilmektedir. Bu sebeple İngilizce gibi dillerde yapılan çalışmalar Türkçe gibi sondan eklemeli diller için bazen beklenenden daha düşük sonuçlar verebilmektedir. Daha önce yapılan çalışmalar kelimelerin morfemlerini hesaba katmanın model sonuçlarını daha yukarı çıktığını gösterdi [14] [15]

Bizim bu çalışmadaki asıl katkımızı iki başlık altında toplayabiliriz: 1) Başta İngilizce olmak üzere diğer dillerde

yapılan derin öğrenme kullanan dizi etiketleme modellerini Türkçe dili için uygulamak. 2) Morfem seviyesindeki dikkat ağı ile Türkçe dizi etiketleme için başarısı daha yüksek bir model elde etmek.

II. MODEL

Bu çalışmada Bi-LSTM ve CRF katmanlarından oluşan ve daha çok dizi etiketleme problemleri için kullanılan mimari baz alınmıştır. Kelime vektörleri uzun kısa vadeli bellek ağından geçirilerek bu katmanın çıktısı etiketleri tahmin etmek üzere CRF katmanına verilir. CRF katmanının çıktısı tahmin edilen etiketler olacaktır. Burada kullanılan tekrarlayan sinir ağ modeli çift yönlü olup her kelimenin hem sağındaki hem de solundaki kelimelerden gelecek özellikleri barındırmasını sağlamaktadır. Her kelime için ileri ve geri yönlü olan bu LSTM'lerin çıktıları birleştirilerek CRF katmanına aktarılmaktadır.



Şekil 1. Modelimizin temel mimarisini

Doğal dil işleme problemlerinde cümledeki kelimeleri matematiksel olarak ifade etmek için kelime vektörleri kullanılmaktadır. Kelimeleri sabit uzunluktaki bu vektörlerle ifade etmek birçok NLP probleminde başarıyı artırmıştır. [2] Daha önceki çalışmalarda dağıtımsal özellikler (distributional) kullanılarak istatistiksel veya frekans tabanlı modellerle kelimelerin vektörleri tahmin edilmeye çalışılsa da, son senelerde daha çok nöral modeller ön plana çıkmaya başlamıştır.

Bu nöral çalışmaların bir kısmı kelime vektörlerini model içerisinde öğrenmeye çalışırken diğer çalışmalarda ise daha önceden eğitilmiş hazır kelime vektörleri kullanılmaktadır.

Daha önce de bahsedildiği üzere bazı dizi etiketleme çalışmalarında ekstradan evrişimli sinir ağı ile elde edilmiş karakter vektörleri de bu modele eklenebilmektedir. Bizim modelimizde de benzer şekilde elde edilen karakter vektörleri, kelime vektörleri ve morfem vektörleri ile birleştirilerek kelime seviyesindeki tekrarlayan sinir ağı modeline girdi olarak verilmiştir. Morfem vektörlerini elde etmek için ise, her kelimeye ait morfem etiket bilgileri başka bir tekrarlayan sinir ağına girdi olarak verilmekte, bu ağın çıktısı da bir dikkat mekanizmasından geçirilmektedir. Burada dikkat mekanizması kullanılarak her kelime için hangi morfemin, kelimenin etiketinin bulunmasında daha anlamlı olduğu hesaba dâhil edilmiş olmaktadır. Modelin görünümü Şekil 1'de verilmiştir.

Bu çalışma kapsamında karakter ve morfem bilgisini kullanmadan sadece kelime vektörleri ile oluşturulan modeli temel model (Model 1) olarak ele alıp buna ilaveten 3 farklı model daha oluşturduk. Kelime vektörlerine sadece karakter vektörü eklenerek oluşturulan (Model 2), sadece morfem vektörü eklenerek oluşturulan (Model 3) ve hem karakter hem de morfem vektörü eklenerek oluşturulan (Model 4) modeller üzerinde farklı deneyler yaptık.

III. DENEYLER VE SONUÇLAR

A. Veri Kümesi

Türkçe için varlık isimlerini bulma problemi (NER) üzerine yapılan çalışmaların bir çoğu veri kümesi olarak [16] kullanılmaktadırlar. Biz de diğer çalışmalarla karşılaştırma yapabilmek için aynı veri kümesini kullandık. Bu veri kümesi gazete makalelerinden derlenmiş ve ENAMEX tipinde işaretlenmiş olarak yaklaşık olarak 500 bin kelime barındırmaktadır.

Üzerinde çalıştığımız bir diğer dizi etiketleme problemi olan sözcük türü işaretleme problemi içinse METU-Sabancı Türkçe Treebank [17] kullandık.

Her iki veri kümesi de 3 parçaya ayrılmış ve bu parçalar öğrenme, doğrulama ve test amaçlı kullanılmıştır. Tüm verinin %80'i öğrenmeye, %10'ar da doğrulama ve test için kullanılmıştır.

B. Model Parametreleri

Bu çalışmada kelime vektörü olarak daha önceden eğitilmiş olan Skipgram [18] vektörlerini kullandık. Bu vektörler her kelime için 300 boyutunda tutulmaktadır.

Kelimelerin karakter özelliklerini ifade etmek için CNN kullanılarak oluşturulan özellikler boyutu 30 olan bir vektör olarak belirlenmiştir. Modelde oluşturduğumuz morfem katmanının çıktısı olan morfem vektörleri ise 60 boyutunda vektörler olarak oluşturulmuştur.

Kelime vektörleri, karakter vektörleri ve morfem vektörleri uç uca eklenerek 390 uzunluğunda bir vektör oluşturulmuş ve oluşan bu vektör kelime düzeyindeki tekrarlayan sinir ağı modeline girdi olarak verilmiştir.

Makine öğrenmesi çalışmalarında oluşturulan modelin eğitim verisini aşırı öğrenmesi (overfitting) test verileri üzerinde yapılacak tahminlemeyi olumsuz etkileyeceği için unutmama parametreleri (dropout) kullanılmalıdır. Kullanılacak bu parametreler modelin başarısını doğrudan etkileyecektir. Biz de bu çalışmada hem karakter vektörlerinin çıkarımı sırasında CNN katmanında (0.5) hem morfem vektörlerinin çıkarımı sırasında Bi-LSTM katmanında (0.4) hem de kelime seviyesindeki Bi-LSTM katmanında (0.5) unutmama parametreleri kullandık.

Modellerin eğitimi esnasında her bir model için derlemin üzerinden 50 kere (epoch) geçtik.

C. Deney Sonuçları

Tablo I'de görüleceği üzere morfem veya karakter vektörleri de hesaba katıldığı zaman modelin doğruluğunu hem varlık isimlerinin bulunmasında (NER) hem de sözcük türlerinin işaretlenmesinde (POS) sadece kelime vektörleri kullanıldığı modele göre artış göstermiştir. Ancak morfem vektörlerinin etkisi karakter vektörlerinin etkisinden daha fazla olmuştur. Bu durum [14] çalışmasında olduğu gibi Türkçe gibi sondan eklemeli dillerde morfemlerin karakterlerden daha fazla anlam ifade etmesine örnek verilebilir. Modeli hem karakter hem morfem özellikleri ekleyerek oluşturduğumuzda ise sonuç daha da iyileşme göstermiştir. Sadece kelime vektörleri kullanıldığında varlık ismi tanımda elde edilen F skoru % 93.03 iken kelime vektörlerinin yanında karakter ve morfem vektörleri de kullanıldığında % 93.71'e yükselmiştir. Bu iyileşme, sözcük türü işaretlemeye daha fazla olmuştur. Sadece kelime vektörleri kullanıldığında elde edilen % 91.95 doğruluk oranı, morfem vektörleri kullanıldığında % 95.71'e yükselmiştir. Hem karakter, hem de morfem vektörleri kullanıldığında ise bu başarı % 95.68'de kalmıştır.

TABLO I. ÖNERİLEN DERİN ÖĞRENME MODELLERİNE AIT F-SKOR VE DOĞRULUK ORANLARI

Model	NER(F1)	POS(Doğruluk)
Model 1	93,03	91,95
Model 2 (Kelime vektörü + Karakter Vektörü)	93,23	93,85
Model 3 (Kelime vektörü + Morfem Vektörü)	93,68	95,71
Model 4 (Kelime vektörü + Karakter vektörü + Morfem Vektörü)	93,71	95,68

Varlık isimlerinin bulunması probleminde aynı veri kümesini kullanan önceki örneklerle karşılaştırdığımızda (bkz. Tablo II) bizim modelimizin sonuçları en iyi olarak gözükmektedir. Sonuçları bizim çalışmamıza en yakın olan [4] çalışmasında büyük/küçük harf, noktalama işareti gibi ilave yazım özellikleri de modele eklenmiştir. Bizim modelimizde ise manuel olarak hiçbir özellik modele eklenmemiştir. Bu çalışmalardan [19] bizim modelimize en benzer model olarak görünmektedir. Ancak bizim modelimiz morfem seviyesinde kullandığı dikkat ağı ile bu çalışmadan ayrılmaktadır.

TABLO II. DAHA ÖNCE YAPILAN NER ÇALIŞMALARI İLE KARŞILAŞTIRMA

Model	F1
Şeker ve Eryiğit [1]	91,94
Demir ve Özgür [20]	91,85
Kuru ve diğ. [9]	91,30
Güneş ve Tantuğ [4]	93,69*
Güngör ve diğ. [7]	93,37
Güngör ve diğ. [19]	92,93
Önerilen Model (Model 4)	93,71

* Yazım özellikleri kullanılmadan 91,59'a düştüğü ifade edilmiştir.

Sözcük türlerinin işaretlenmesi ile ilgili daha önce yapılmış çalışmalarla önerdiğimiz modelin karşılaştırması Tablo III'de verilmiştir. Bizim temel aldığımız model (Model 1) bu çalışmalardan biraz daha iyi olmasına rağmen önerdiğimiz model ile bu çalışmalar arasında ciddi bir başarı farkı gözükmemektedir. Bu da morfem bilgisinin sözcük türü işaretlemeye önemli bir katkısı olduğunu göstermektedir.

TABLO III. DAHA ÖNCE YAPILAN POS TAGGING ÇALIŞMALARI İLE KARŞILAŞTIRMA

Model	Doğruluk (Accuracy)
Dinçer ve diğ. [21]	88,90
Can ve diğ. [22]	91,05
Bahçevan ve diğ. [23]	89,00
Önerilen Model (Model 3)	95,71

IV. SONUÇ

Bu çalışmada Türkçede kelime dizilerini etiketleme problemleri üzerine çalışılmış ve mevcut çalışmalardan daha iyi sonuç verecek bir model geliştirilmiştir. Derin öğrenme ile oluşturulan bu model hem varlık isimlerinin bulunması hem de sözcük türü işaretleme problemlerinde kullanılmıştır.

Türkçenin sondan eklemeli bir dil olması ve bu gibi diller için morfemlerin daha fazla anlam içermesi sebebiyle oluşturulan modele morfem bilgisi de dikkat mekanizması kullanılarak dâhil edilmiş ve böylece mevcut çalışmalardan daha iyi sonuçlar elde edilmiştir.

KAYNAKLAR

- [1] G. A. Şeker and G. Eryiğit, "Initial explorations on using CRFs for Turkish named entity recognition," *Proceedings of COLING 2012*, pp. 2459--2474, 2012.
- [2] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu and P. Kuksa, "Natural language processing (almost) from scratch," *Journal of Machine Learning Research*, vol. 12, pp. 2493--2537, 2011.
- [3] Z. Huang, W. Xu and K. Yu, "Bidirectional LSTM-CRF models for sequence tagging," *arXiv preprint arXiv:1508.01991*, 2015.
- [4] A. Güneş and A. C. Tantuğ, "Turkish named entity recognition with deep learning," in *2018 26th Signal Processing and Communications Applications Conference (SIU)*, 2018, pp. 1--4.
- [5] J. Chiu and E. Nichols, "Named entity recognition with bidirectional LSTM-CNNs," *Transactions of the Association for Computational Linguistics*, vol. 4, pp. 357--370, 2016.
- [6] X. Ma and E. Hovy, "End-to-end sequence labeling via bi-directional lstm-cnns-crf," *arXiv preprint arXiv:1603.01354*, 2016.
- [7] O. Güngör, S. Üsküdarlı and T. Güngör, "Recurrent neural networks for Turkish named entity recognition," in *2018 26th Signal Processing and Communications Applications Conference (SIU)*, 2018, pp. 1--4.

- [8] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami and C. Dyer, "Neural architectures for named entity recognition," *arXiv preprint arXiv:1603.01360*, 2016.
- [9] O. Kuru, O. A. Can and D. Yuret, "Charner: Character-level named entity recognition," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 2016, pp. 911--921.
- [10] J. Yang, S. Liang and Y. Zhang, "Design Challenges and Misconceptions in Neural Sequence Labeling," *arXiv preprint arXiv:1806.04470*, 2018.
- [11] D. Bahdanau, K. Cho and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.
- [12] P. Zhou, W. Shi, J. Tian, Z. Qi, B. Li, H. Hao and B. Xu, "Attention-based bidirectional long short-term memory networks for relation classification," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2016, pp. 207--212.
- [13] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola and E. Hovy, "Hierarchical attention networks for document classification," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2016, pp. 1480--1489.
- [14] A. Üstün, M. Kurfalı and B. Can, "Characters or Morphemes: How to Represent Words?," in *Proceedings of The Third Workshop on Representation Learning for NLP*, 2018, pp. 144--153.
- [15] R. Yeniterzi, "Exploiting morphology in Turkish named entity recognition system.," in *Proceedings of the ACL 2011 Student Session*, Association for Computational Linguistics, 2011, pp. 105--110.
- [16] G. Tür, D. Hakkani-Tür and K. Oflazer, "A statistical information extraction system for Turkish," *Natural Language Engineering*, vol. 9, pp. 181--210, 2003.
- [17] K. Oflazer, B. Say, D. Z. Hakkani-Tür and G. Tür, "Building a Turkish treebank," in *Treebanks*, Springer, 261--277, p. 2003.
- [18] O. Güngör and E. Yıldız, "Linguistic Features in Turkish Word Representations," in *Signal Processing and Communications Applications Conference (SIU), 2017 25th*, IEEE, 2017, pp. 1--4.
- [19] O. Güngör, T. Güngör and S. Üsküdarlı, "The effect of morphology in named entity recognition with sequence tagging," *Natural Language Engineering*, vol. 25, pp. 147--169, 2019.
- [20] H. Demir and A. Özgür, "Improving Named Entity Recognition for Morphologically Rich Languages Using Word Embeddings," in *ICMLA*, 2014, pp. 117--122.
- [21] T. Dincer, B. Karaoglan and T. Kisla, "A suffix based part-of-speech tagger for Turkish," in *Information Technology: New Generations, 2008. ITNG 2008. Fifth International Conference on*, IEEE, 2008, pp. 680--685.
- [22] B. Can, A. Üstün and M. Kurfalı, "Turkish POS tagging by reducing sparsity with morpheme tags in small datasets," in *International Conference on Intelligent Text Processing and Computational Linguistics*, Springer, 2016, pp. 320--331.
- [23] C. A. Bahcevan, E. Kutlu and T. Yildiz, "Deep Neural Network Architecture for Part-of-Speech Tagging for Turkish Language," in *2018 3rd International Conference on Computer Science and Engineering (UBMK)*, IEEE, 2018, pp. 235--238.
- [24] T. Mikolov, I. Sutskever, K. Chen, G. Corrado and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in neural information processing systems*, 2013, pp. 3111--3119.
- [25] J. Pennington, R. Socher and C. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532--1543.