

Teksir Takımı

Bağlamsal Modeller Kullanarak Metinsel Veri Çoğaltma Kütüphanesi



Ekibimiz

Sinan ÇALIŞIR

- Kocaeli Üniversitesi | Bilgisayar Mühendisliği
- Makine Öğrenmesi Mühendisi | Newmind (GSI Göksu Safi Işık Avukatlık Ortaklığı)

Muhammed Emir KOÇAK

- Yeditepe Üniversitesi | Bilgisayar ve Öğretim Teknolojileri Eğitimi
- Makine Öğrenmesi Mühendisi | Newmind (GSI Göksu Safi Işık Avukatlık Ortaklığı)

Muhammed Furkan ÇANKAYA

- Fatih Sultan Mehmet Üniversitesi | Bilgisayar Mühendisliği
- Kodiks Bilişim



Problem

- Kısıtlı metinsel veri kümelerinin model geliştirmeleri için yetersiz kalması
- Sınıf dağılımlarının eşit olmadığı veri kümeleri üzerinde geliştirilen modellerin baskın sınıfa bir yöneliminin bulunması



Çözüm ve İş Akışı

Bağlamsal modellerden faydalanarak veri artırımı yapabilen doğal dil işleme kütüphanesi

Uygulama Biçimi:

- Veri kümesinde yer alan cümlelere ait sözcüklerin rastgele maskelenerek yerine bağlamsal model (BERT) aracılığı ile olası sözcüklerin yerleştirilmesi
- Artırılmış veri kümesi ile başarımın ölçülmesi ve sonuçların karşılaştırılması



DEMO

LİNK: <https://www.youtube.com/watch?v=5XFSSyKoFVU>



Sorumluluklar

Sinan ÇALIŞIR

- Bağlamsal modellerin uygulanması ve veri kümelerinin hazırlanması

Muhammed Emir KOÇAK

- Geliştirilen modellerin tasarımı ve eğitimi

Muhammed Furkan ÇANKAYA

- Geliştirilen modellerin tasarımı ve eğitimi



Sonuç ve İleriki Çalışmalar

Bağlamsal modeller kullanılarak veri artırım kütüphanesi geliştirildi.

Diğer bağlamsal modellerin (RoBERTa, ELECTRA vb.) desteklenmesi planlanmaktadır.

