

Kolektif Öğrenme 2023/1

Ödev No:2

Ödev Konusu: Küçük ve büyük Bert'lerle ensemble

Aşağıdaki 3 yöntem kullanılacaktır:

A) Bagging1: Resampling yapılan örneklerle aynı dil modelinden* finetune ile K adet dil modeli üretilip sonuçları birleştirilecek. Tekil modellerin ve ensemble in başarıları karşılaştırılacak. K=5. *Dil modeli olarak büyük ve küçük dil modeli kullanımının etkileri karşılaştırılacak.

B) Bagging2: Metinlerin dil modeliyle* temsilleri elde edildikten sonra temsil vektörleri üzerinde resampling yapılarak K adet sınıflandırıcı (SVM, RF vb.) eğitilip sonuçları birleştirilecek. Tekil modellerin ve ensemble in başarıları karşılaştırılacak. K=10. *Dil modeli olarak büyük ve küçük dil modeli kullanımının etkileri karşılaştırılacak.

C) Zero shot: Metnin temsili ile "bu metnin konusu X" ifadesinin temsili arasındaki kosinüs mesafesi hesaplanacak. X= sınıf isimleri. X'lerden hangisi en yakınsa tahmini sınıf olarak belirlenecek. K adet BERT modelinin bireysel ve ensemble sonuçları raporlanacak. K=5. Bu işlem test örnekleri üzerinde yapılacaktır.

Bonus1: Masking: Aynı model içinde aynı metin için rasgele kelimelerin maskelenerek farklı temsillerin elde edilmesi sonrası sonuçları birleştirilecek. Bu işlem test örnekleri üzerinde yapılacaktır. Maskelenecek kelime oranının(%5, %25, %50) etkisi incelenecek.

Bonus2: Dropout katmanını açıp maskeleye yapmadan aynı modelle aynı metnin farklı temsillerini elde ederek sonuçları birleştirmek. Bu işlem test örnekleri üzerinde yapılacaktır.

Seçtiğiniz en az 3 Türkçe metin veri kümesi (her biri en az 5 bin örnek içeren) üzerinde karşılaştırma yapılacaktır. Performans ölçütü olarak accuracy / F1 kullanılabilir. Her veri kümesi için test örneklerini tüm işlemlerden önce ayırıp tüm yöntemlerde aynı test kümelerini kullanınız.

Sonuç birleştirmede demokrasi usulü kullanımı yeterlidir.

Grafikler (tablolar) sizin bulgularınız. Bunlara dair yorum yapınız. Beklentilerinizle uyuşanlar uyuşmayanlar vb.

Sınıfta sunum yapmak %70 (yapmayanlar en fazla %30 alabilirler)

Önemli Not: Deney sonuçlarınızı, bulgularınızı iyi ifade eden tablolar ve grafiklerle sununuz.

Kaynak: Büyük BERT: <https://huggingface.co/dbmdz/bert-base-turkish-uncased>

Küçük BERTler: <https://huggingface.co/ytu-ce-cosmos>

Ödevin Son Teslim Tarihi: 28 Kasım 2023 saat 07:00

Ödevin Teslim Şekli: online.yildiz.edu.tr

Bu ödevde bireysel olarak çalışılacaktır.

28 Kasımdaki derste raporunuz üzerinden ödevinizi sunmanız beklenmektedir.

Ödevde Teslim Edilecekler: 1-Açıklama içeren kodlarınız (değişkenlerin ne için kullanıldıkları, algoritmanın adımları) 2-Ödev raporu