

学生・若手研究者のためのBERTワークショップ2

固有表現抽出で試したこと

名前: しまじろう よう

職業: エンジニア・データ解析

最終結果: 88.1

パラメータの設定

- MaxLength: 512
- 学習率: 4e-5
- BatchSize: 32
- Epoch: 50

正則化

- AdamW
 - WeightDecay: 0.005
- Dropout
 - 識別層 3 つの間に導入
 - 割合は約 0.5
- Label Smoothing: 0.0005

学習の調整

- 複数スケジューラの導入
 - LinearLR
 - CosineAnnealingLR

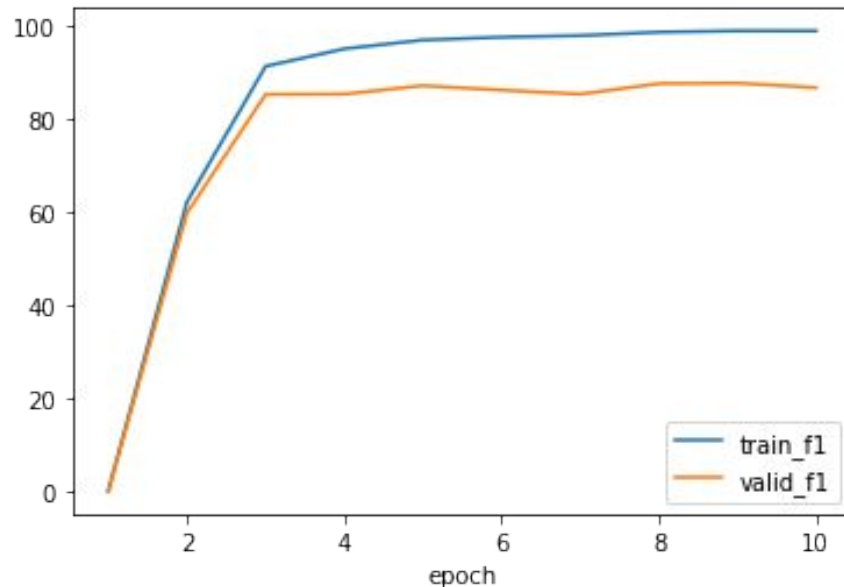
その他

- Train / Valid で評価できるようにする

Train / Valid の評価

オリジナルコードに改良を入れて学習データでのスコアも算出し、学習・テストそれぞれのスコアを比較できるようにしたところ過学習の傾向(スコアで 10 pt 程の乖離)

→ 正則化を検討



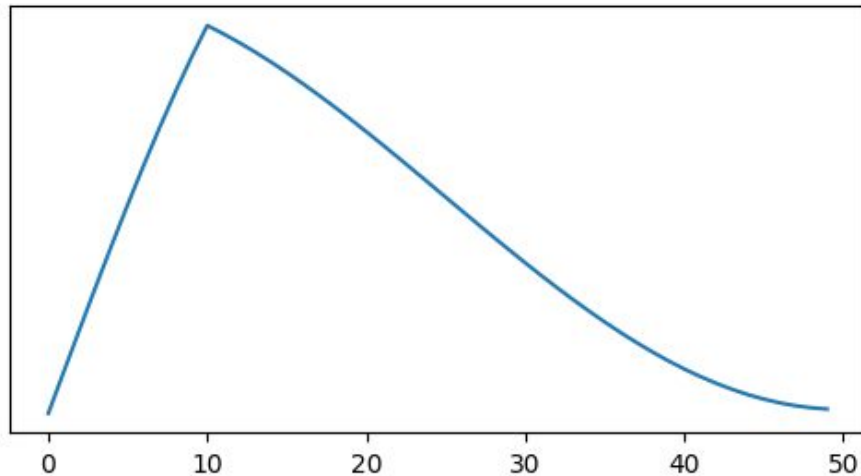
工夫したこと: 正則化

- AdamW の導入
 - Adam だと WeightDecay(荷重減衰) がうまく効きづらいため
- 識別層に DropOut を導入
 - BERT の後の全結合層を 3 層に
 - BERT 直後および全結合層の各層の間に DropOut を追加
- Label Smoothing(ラベル正則化)

工夫したこと: スケジューラの組み合わせ

最初の数エポックで学習率を上昇させ、その後に減少させると良いらしい事を知ったので試してみた

- LinearLR: 最初の 10 エポックで線形に増加
- CosineAnnealingLR: 残りの 40 エポックで緩やかに減少
 - 線形にしなかったのは学習率の最大値付近を長く保ちたかったため



失敗したこと: BERT の学習対象レイヤの指定

- fine-tuning の際に BERT の最終層に近いところ(最終 4 層辺り)のみを学習対象として指定
 - 過学習の抑制を期待

→ **スコア低下**

失敗したこと: Overfitting が解消できず

最終的に学習・テストデータによるスコアの差は解消できなかった

- 今回試した正則化では解消しきれなかった
 - 正則化の導入には意味はあった(スコアに寄与)
- そもそも BERT だとあるあるなのか？
- BERT での一般的なやり方を知らないから？

もう少し工夫できたかも

- RoBERTa large の導入
 - 1, 2 位のスコアを見ると圧倒的
- 正則化をもっと強めてみる