Motivation:

- Classifier only as good as its features
- Improving word embeddings will offer higher quality features
- Because changes to features are architecture-agnostic, embedding improvements have broad potential

Improving LST
Classifiers by O
Embedding

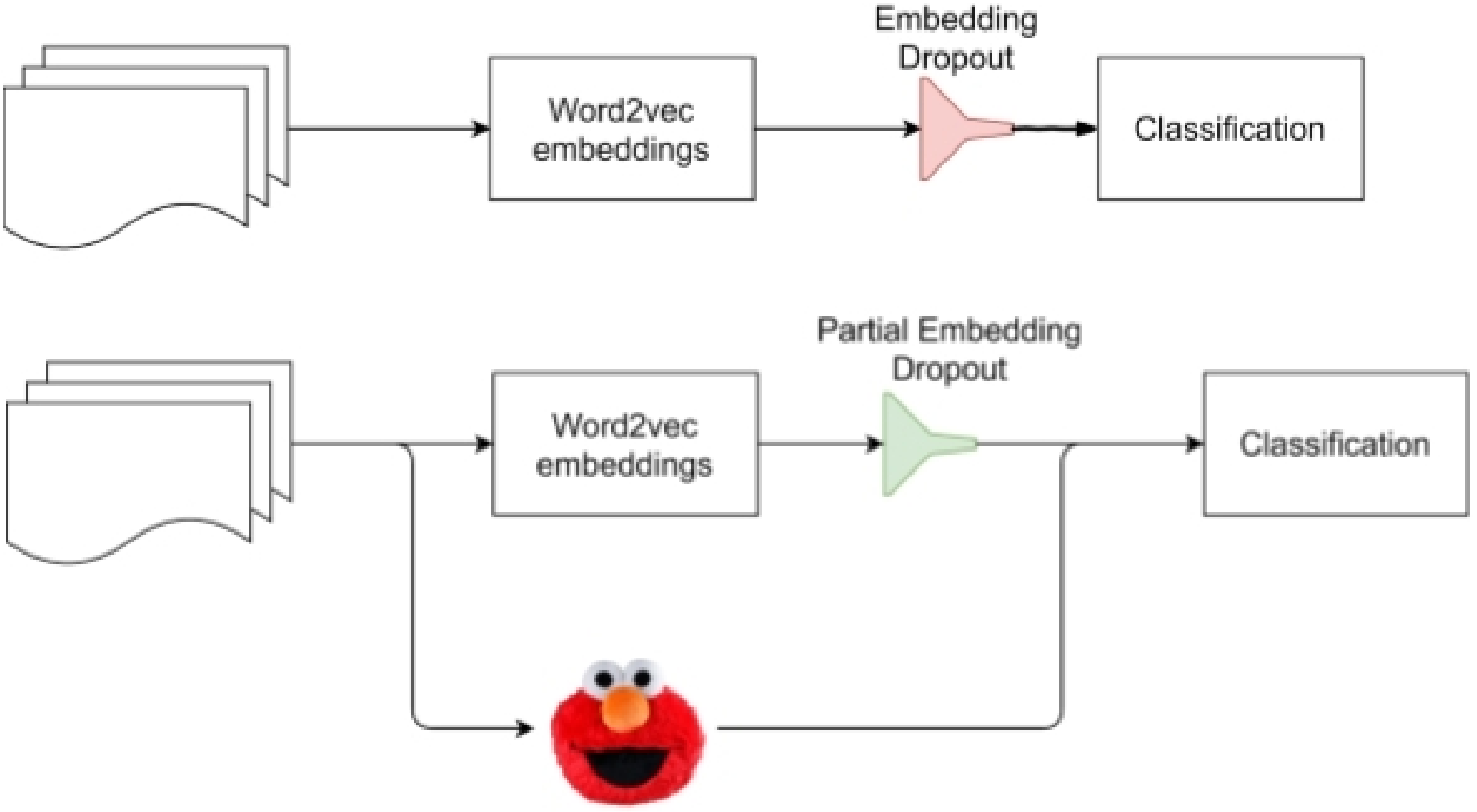
Kevin Liu, William S

TM Document ptimizing Word Techniques

weeny, Justin Tran

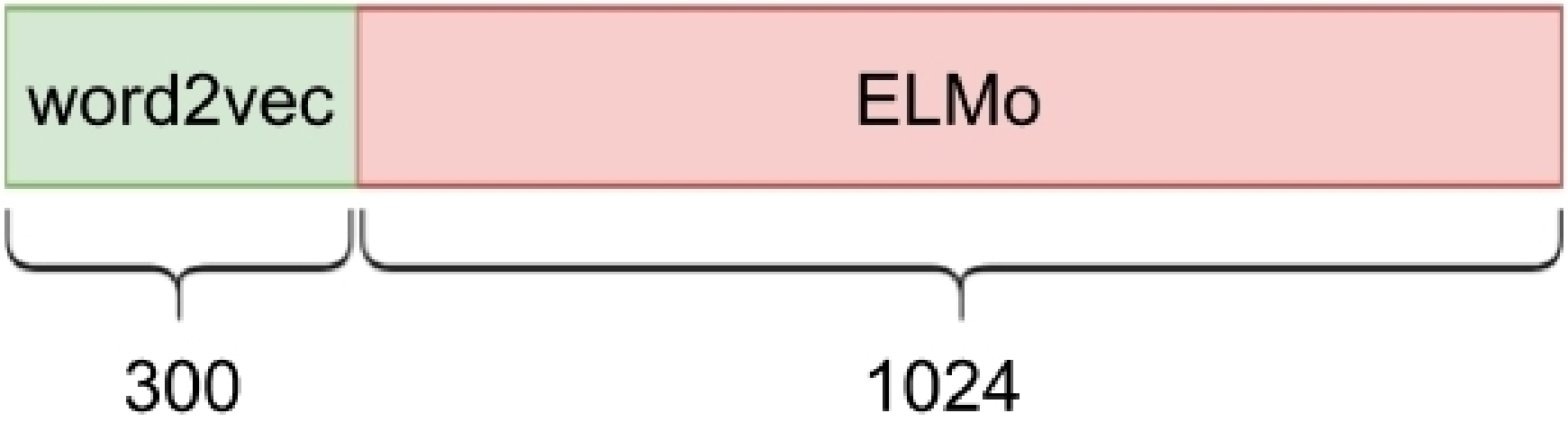
PRINCETON
UNIVERSITY

Approach:


Embedding
Dropout
Word2vec
embeddings
Classification
Partial Embedding
Dropout
Word2vec
embeddings
Classification

Modification 1: ELMo Embeddings

- Added ELMo embeddings to existing word2vec word embeddings



Modification 2: Partial Embedding Dropout

- Replaced word-level embedding dropout with “partial embedding dropout”

Conclusions:

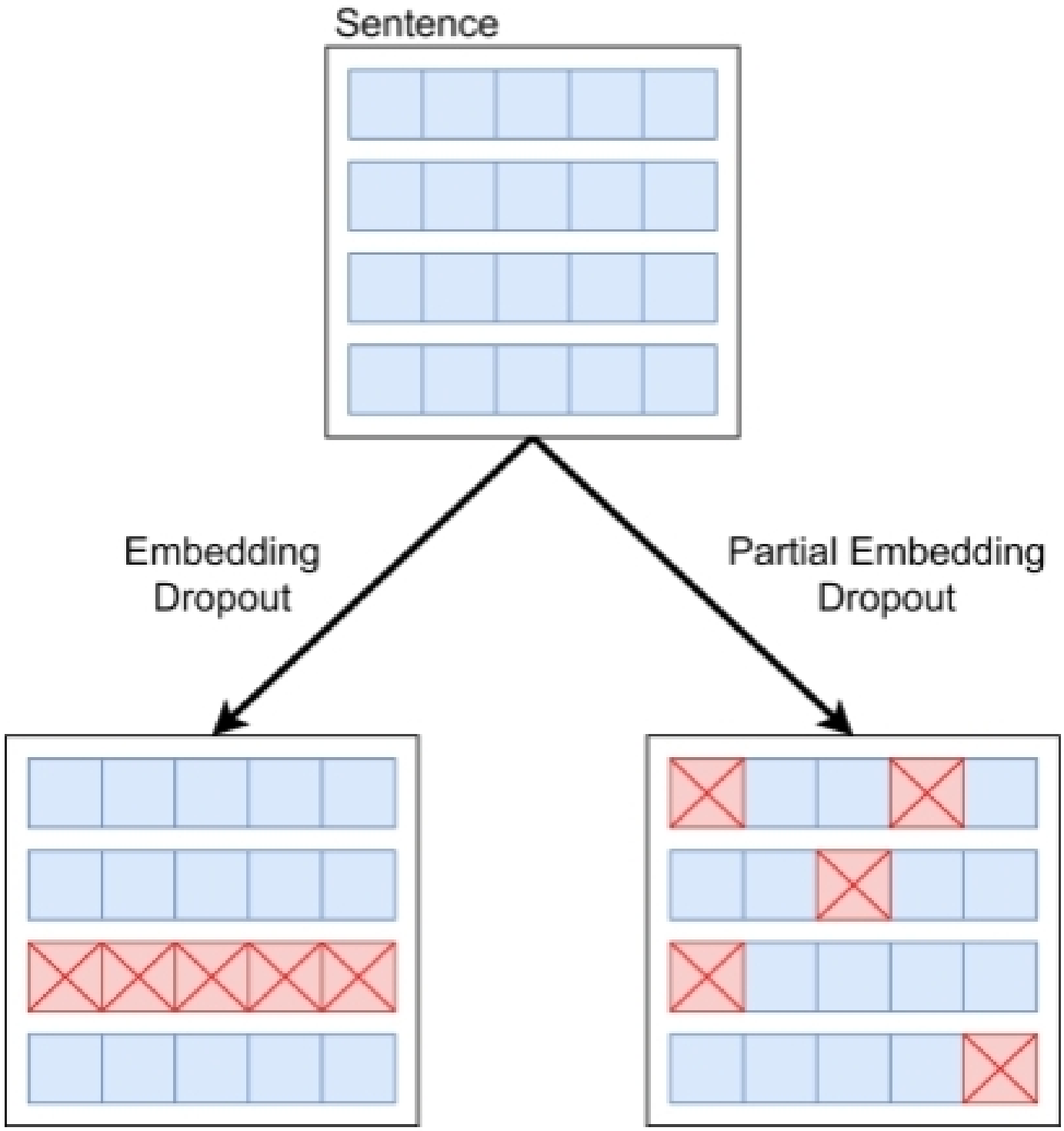
- Partial embedding dropout gives results slightly above baseline, with an optimal dropout rate of 0.3
- Introducing ELMo embeddings did not produce the expected performance gains
    - Perhaps due to limited train time (30 epochs)

Prior Work:

- Attention-based LSTMs perform competitively at cost of increased complexity
- Modifying word embeddings can improve results without complicating architecture substantially

Training Complications:

- ELMo embeddings dramatically increase training times
    - (4x - 6x per epoch)
- Using size 1024 ELMo embeddings quadruples feature size, massively increasing computational overhead


Sentence
Embedding
Dropout
Partial Embedding
Dropout

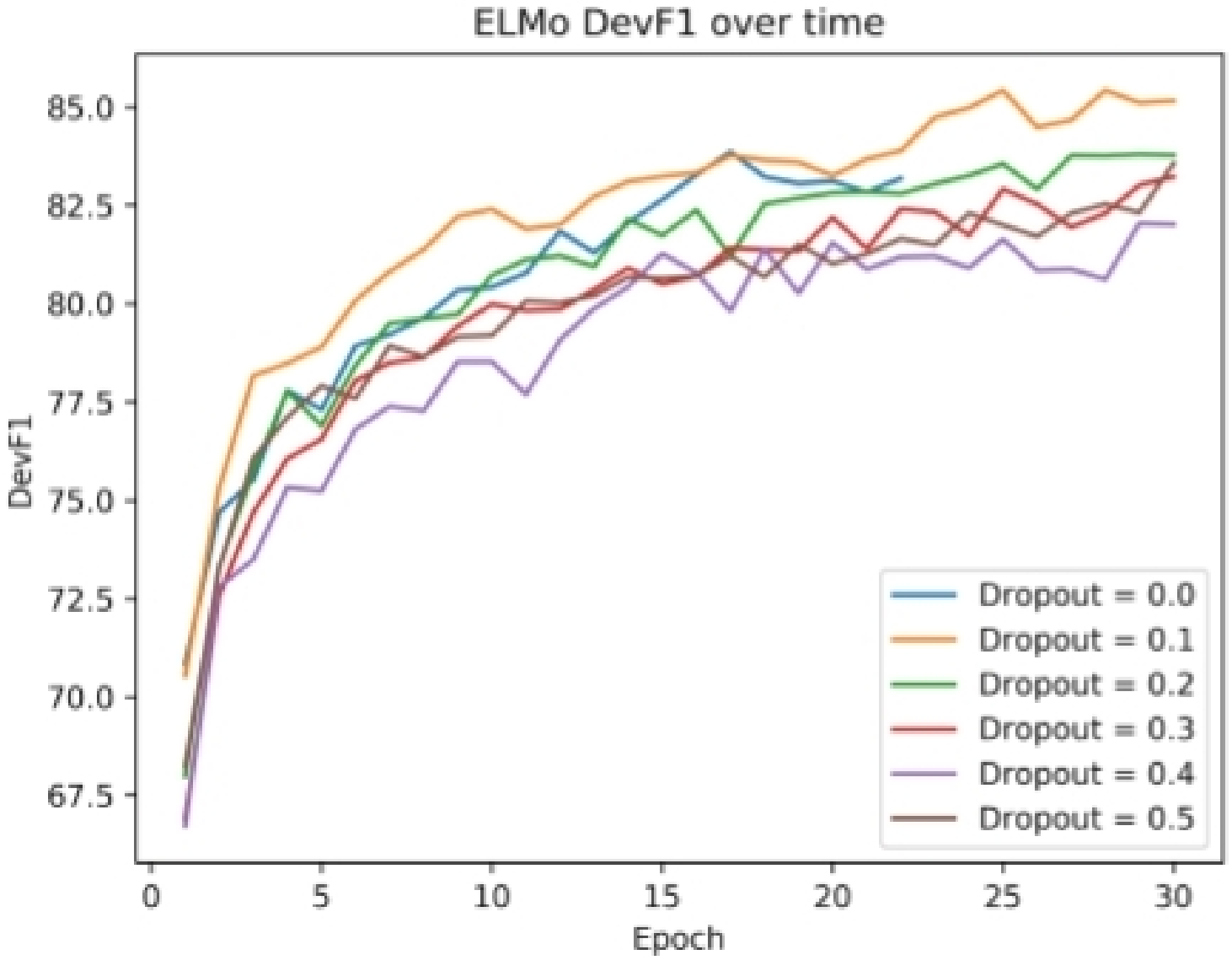
Future Steps:

- Obtain more funding to run ELMo with more epochs
- Combine multiple word embedding dropout schemes
- Confirm results with more trials

Acknowledgements:

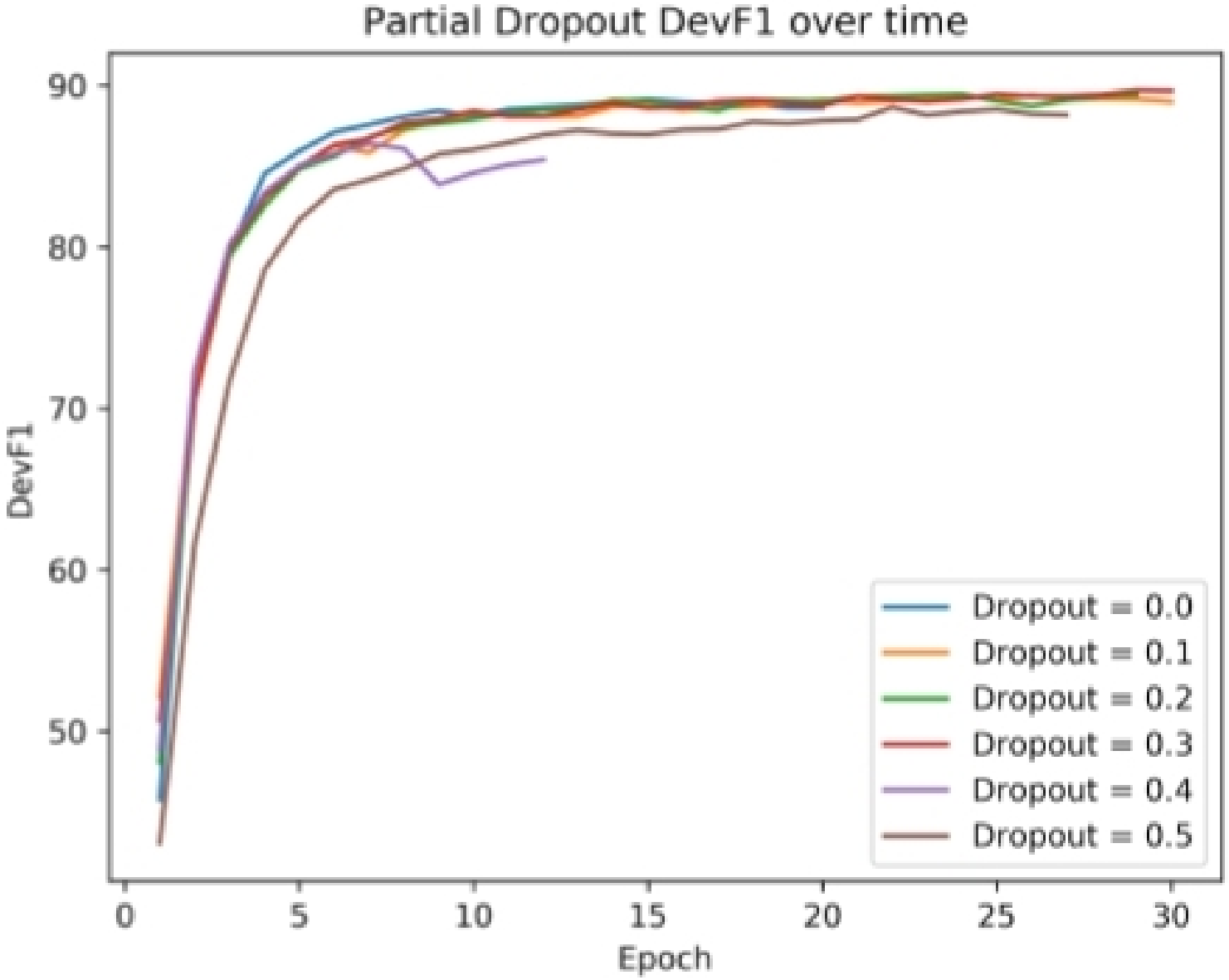
- Danqi Chen for her support and advice, plus her suggestion to evaluate benefits of ELMo
- HackPrinceton for providing the Google Cloud Credit necessary to train our model

Results:



(Using Reuters Dataset)

Results:



(Using Reuters Dataset)

References:

- Ashutosh Adhikari, Achyudh Ram, Raphael Tang, and Jimmy Lin. 2019. Rethinking complex neural network architectures for document classification.
- Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations.