

ex2 - part 3

Eitan kerzhner 205697139
itamar Shachen tov 207497769

Model's Params:

POS:

- o Dropout: $p=0.5$ (default), applied after Tanh activation.
- o Optimizer: Adam.
- o Learning rate: $1e-3$ (Adam's default).
- o Hidden layer size: 128 units.
- o Word embedding size: 50 dimensions (initialized with pre-trained vectors).
- o Context window: 5 words.
- o Batch size (train & dev): 64.
- o Number of epochs: 9.

NER:

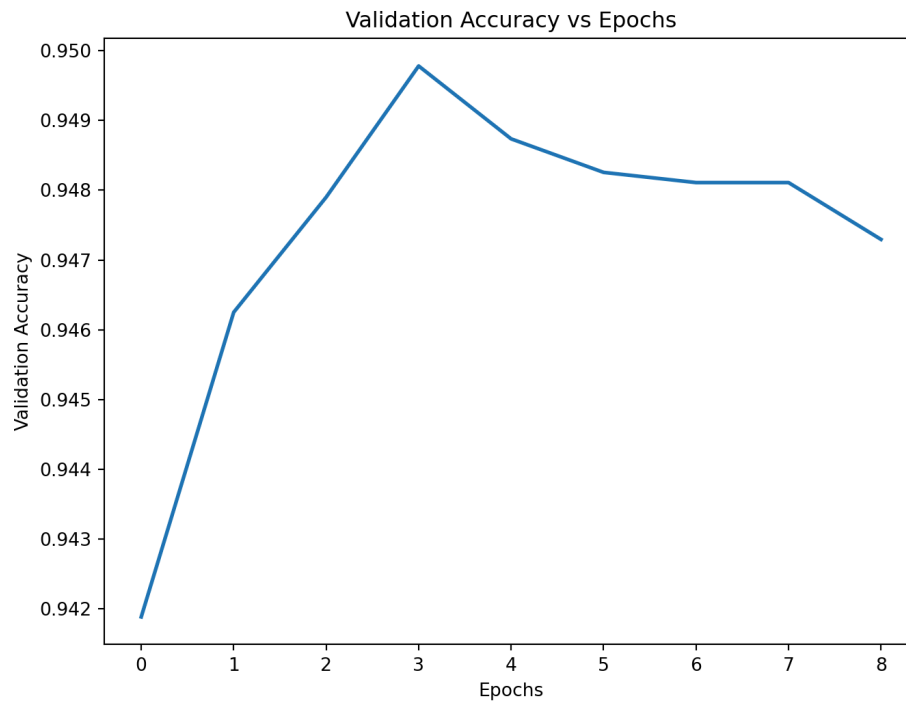
- o Dropout: $p=0.5$ (default), applied after Tanh activation.
- o Optimizer: Adam.
- o Learning rate: $1e-3$ (Adam's default).
- o Hidden layer size: 128 units.
- o Word embedding size: 50 dimensions (initialized with pre-trained vectors).
- o Context window: 5 words.
- o Batch size (train & dev): 64.
- o Number of epochs: 10.

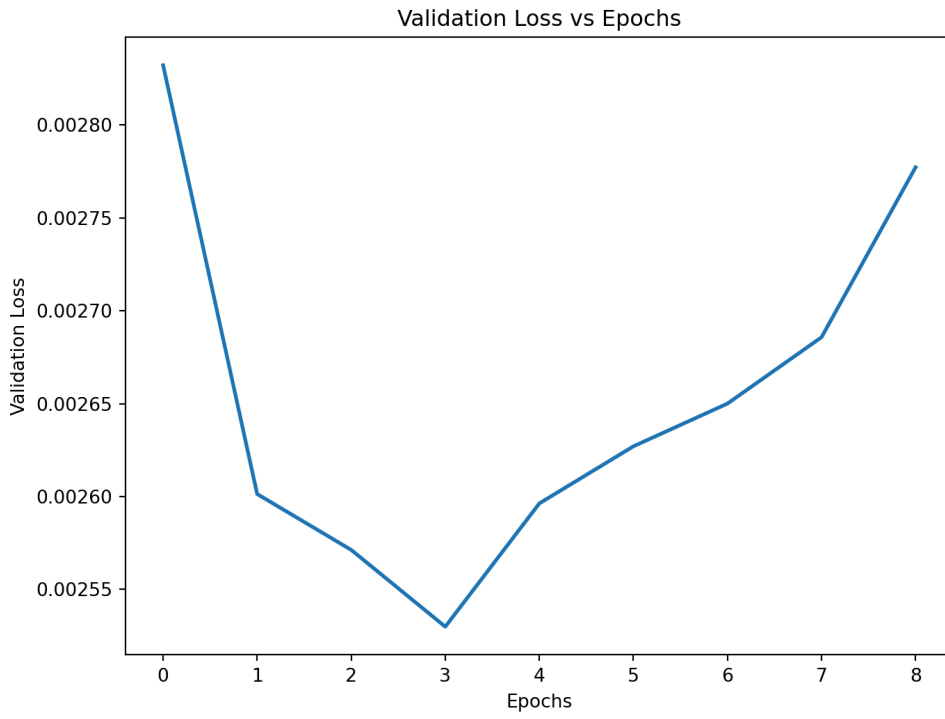
Considerations:

Embedding vocabulary being lower-case: If an input word doesn't match a pre-trained embedding in its original case, the system attempts to match its lowercase version. If the lowercase form exists in the pre-trained vocabulary, its embedding is used.

Words in training file but not embeddings file: These words are mapped to an '<unknown>' token. The embedding for this '<unknown>' token is used, which is sourced from the pre-trained set if available for the token itself, or initialized (e.g., to zeros) and learned otherwise.

POS:





NER:

