

Few-Shot Learning For Text Classification

Master's Thesis by Shaour Haider

First Referee : Prof. Dr. Benno Stein

Second Referee : Prof. Dr. Volker Rodehorst

Overview

- Introduction
- Approaches And Results
- Related Work
- Future Work

What is text classification?

- For given input:
 - a paragraph A
 - a fixed set of classes $C = \{c_1, c_2, \dots, c_n\}$
- Output: a predicted class $c \in C$

Why Text Classification?

- **Sentiment Analysis**

- **Spam Detection**

- **Topic Classification**

[Image: Sentiment Analysis](#)
[Image: Spam Detection](#)
[Image: Topic Classification](#)

Situation

- Limited data

Few-Shot Learning

Few-shot learning aims to learning a classifier with limited amount of labeled examples (<10)

Few-shot task

Train Set → 4-way 1-shot task

Class: Paragraph

Video: Video provides a powerful way to help you prove your point.

Document: You can also type a keyword to search online for the video that best fits your document.

Themes: Themes and styles also help keep your document coordinated.

Design: When you click Design and choose a new Theme, the pictures, charts, and SmartArt graphics change to match your new theme.

Test Set

When you apply styles, your headings change to match the new theme.

Save time in Word with new buttons that show up where you need them.

To change the way a picture fits in your document, click it and a button for layout options appears next to it.

When you work on a table, click where you want to add a row or a column, and then click the plus sign.

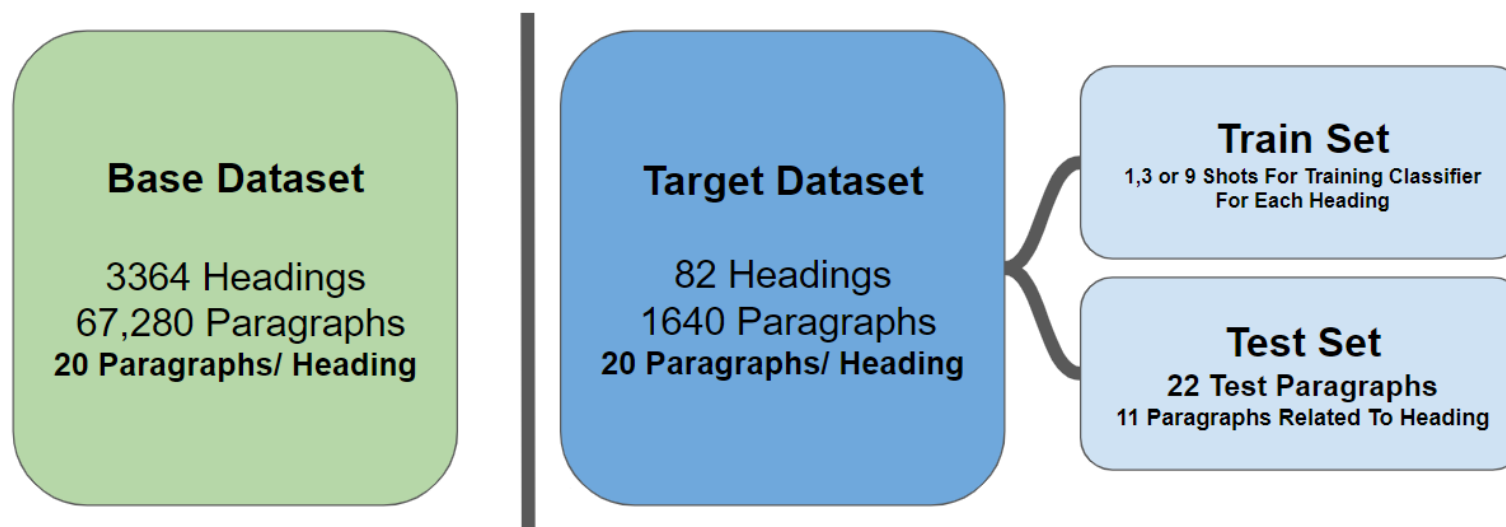
Terminologies

Datasets

Target Dataset:

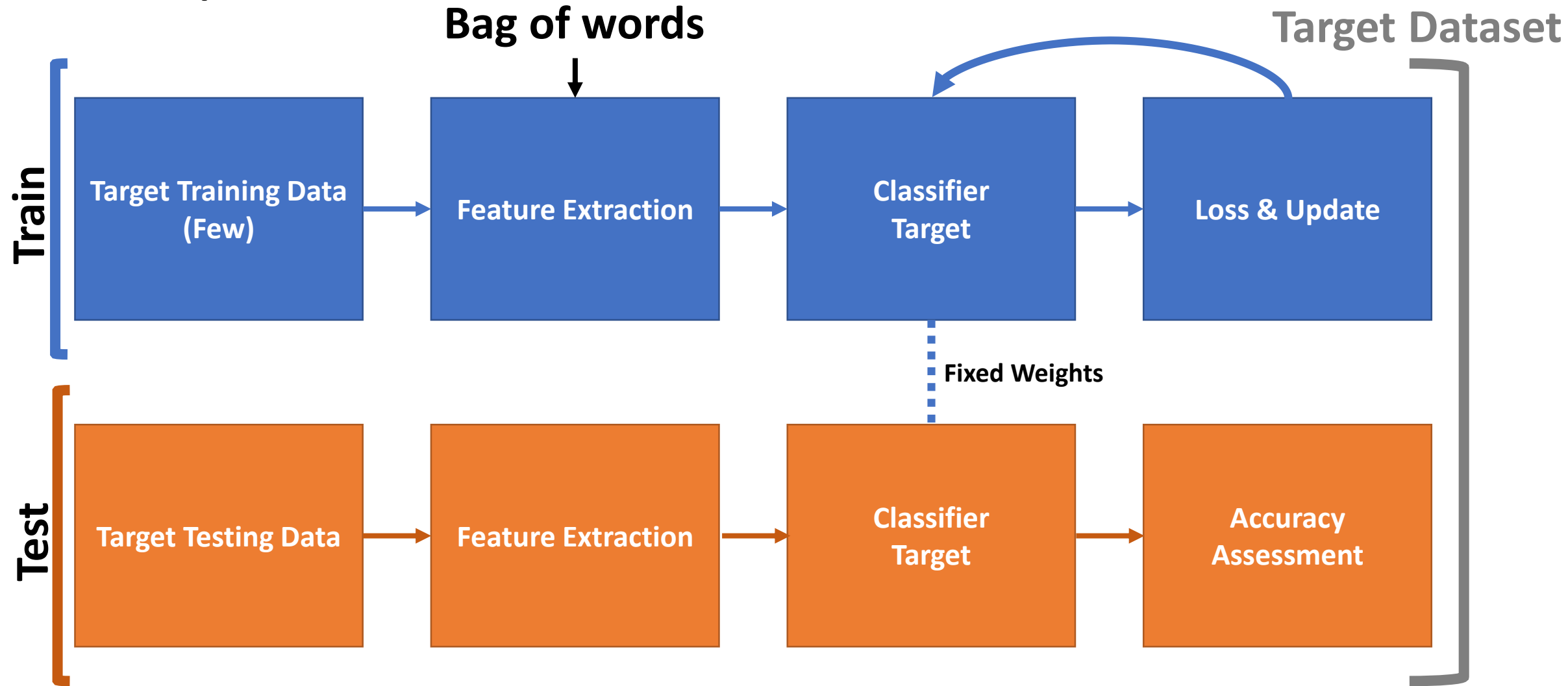
- Train Set (few-shot training set)
- Test Set (testing set)

Base Dataset: Additional dataset that is disjoint to train and test set of target dataset



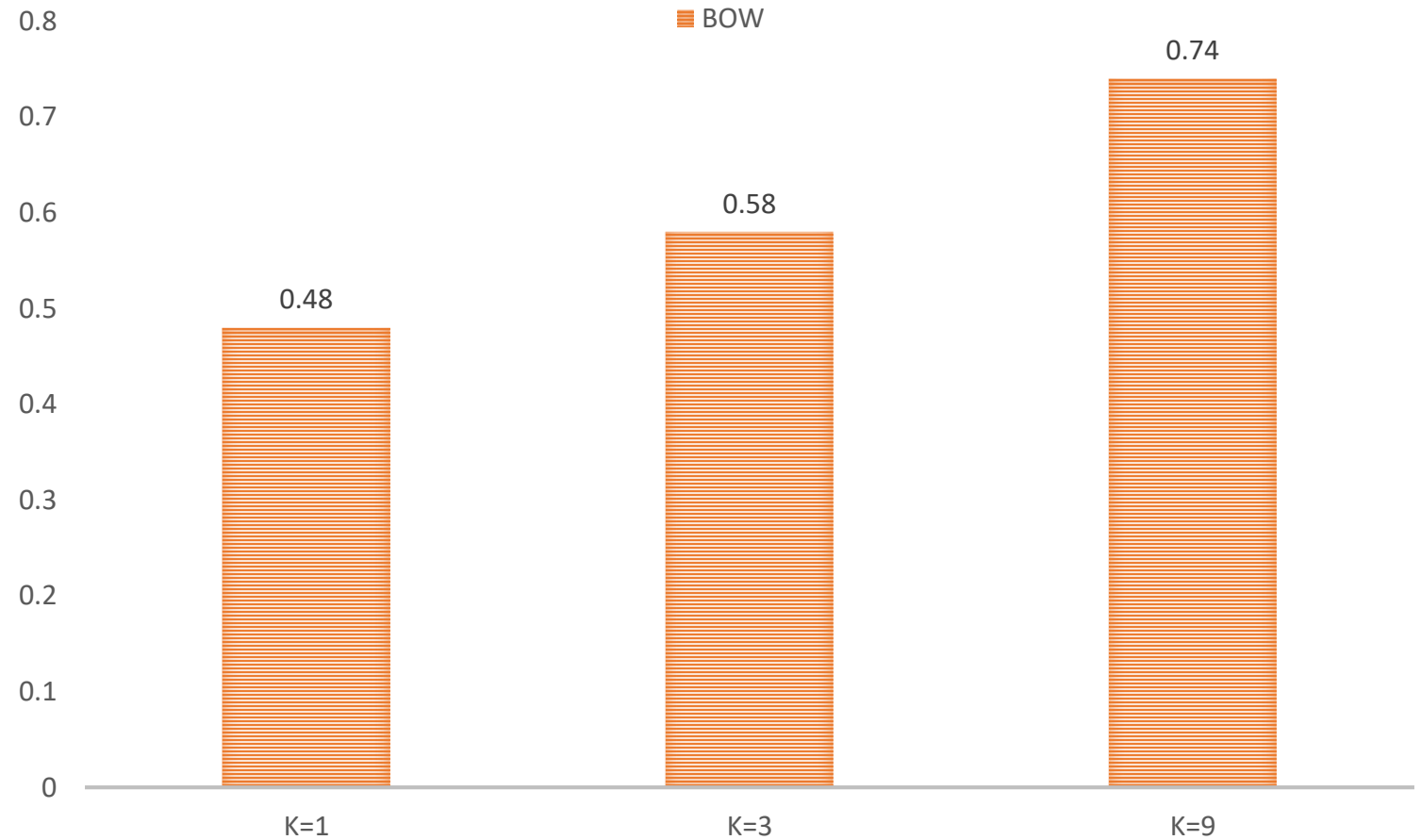
Let's Implement

Approaches And Results



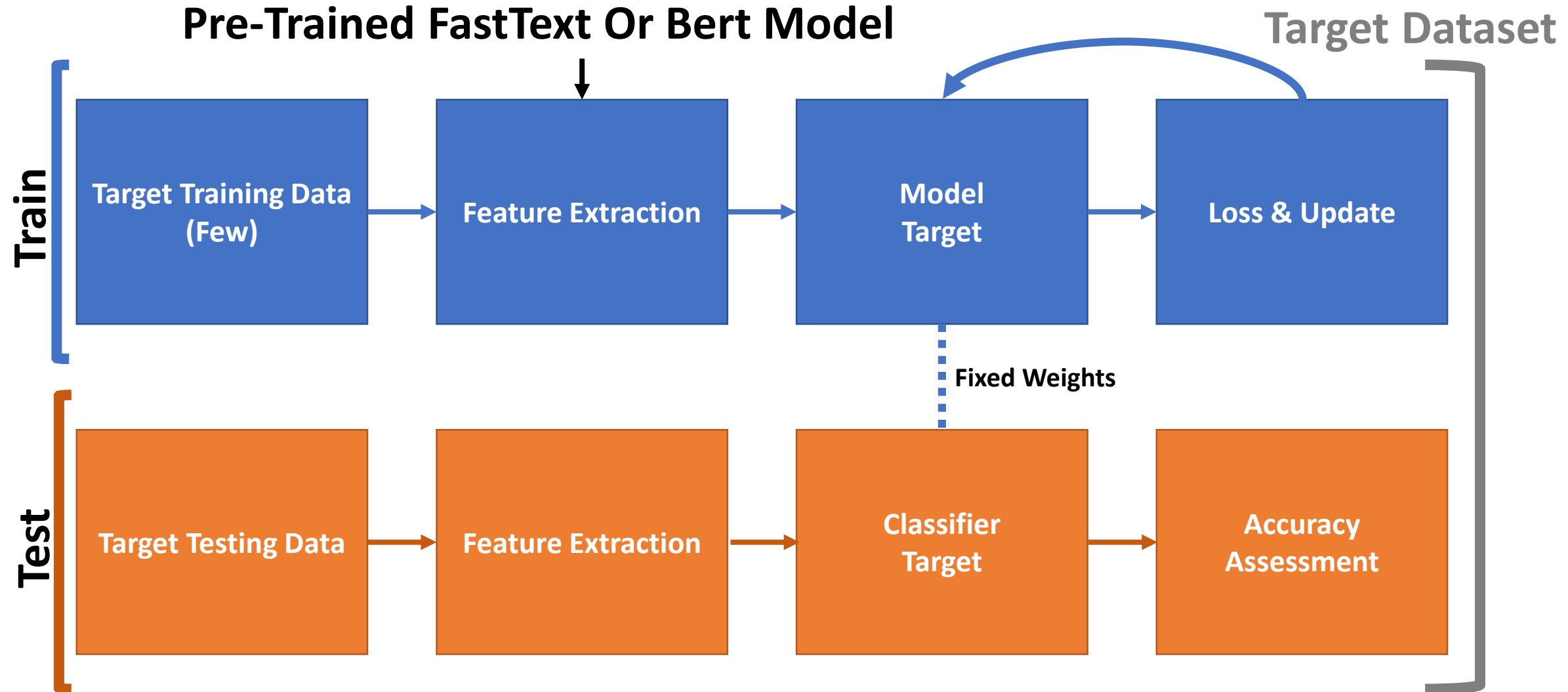
	K=1	K=3	K=9
BOW	0.48	0.58	0.74

Baseline Bag of words



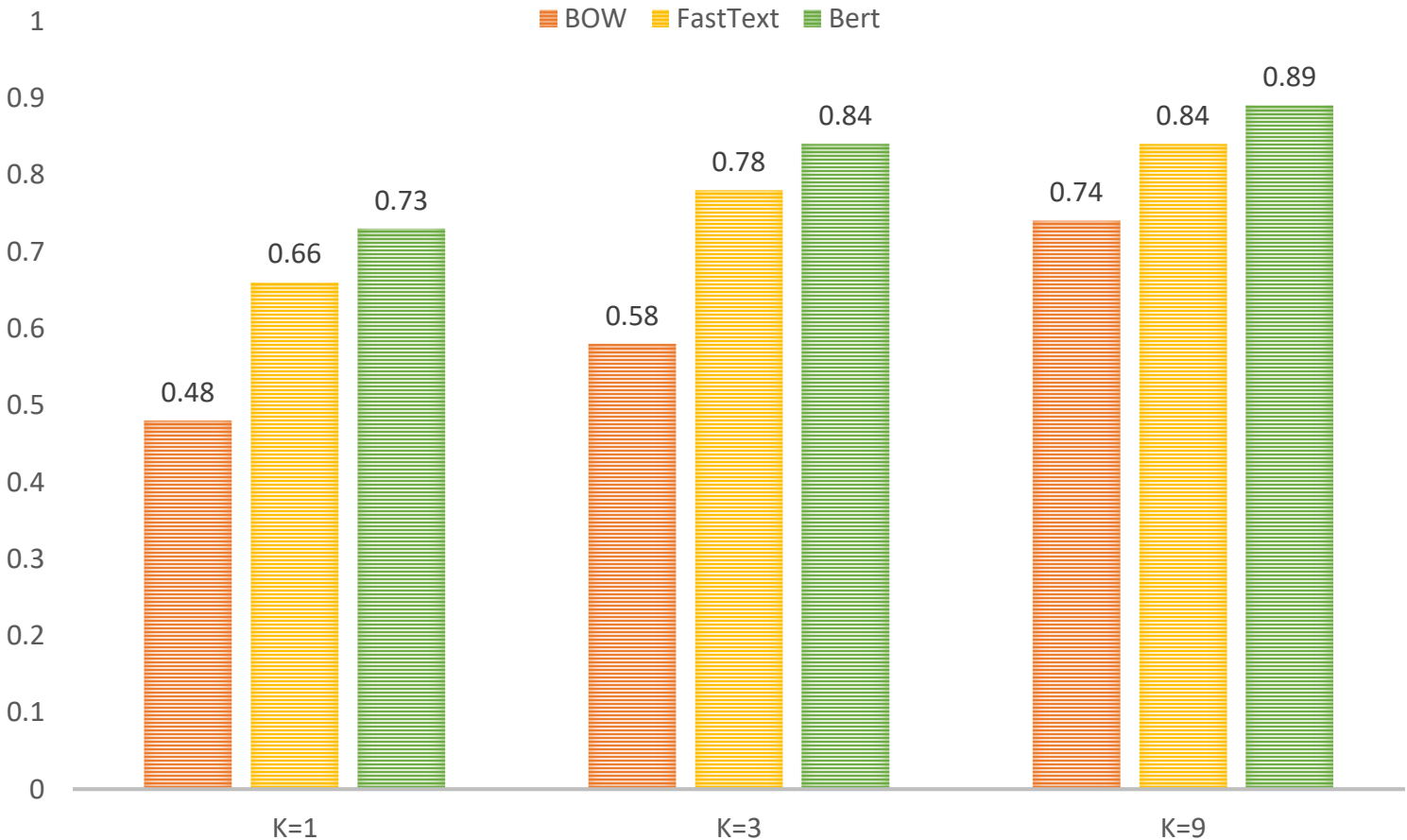
Better representations

Approaches And Results

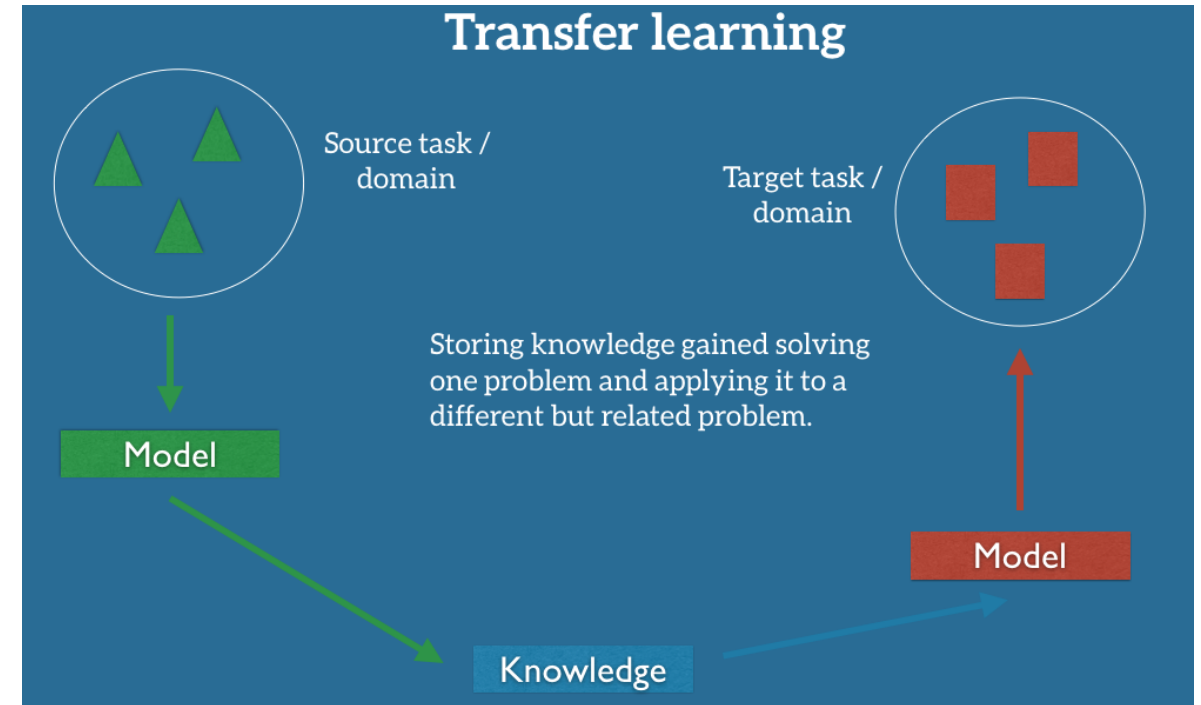
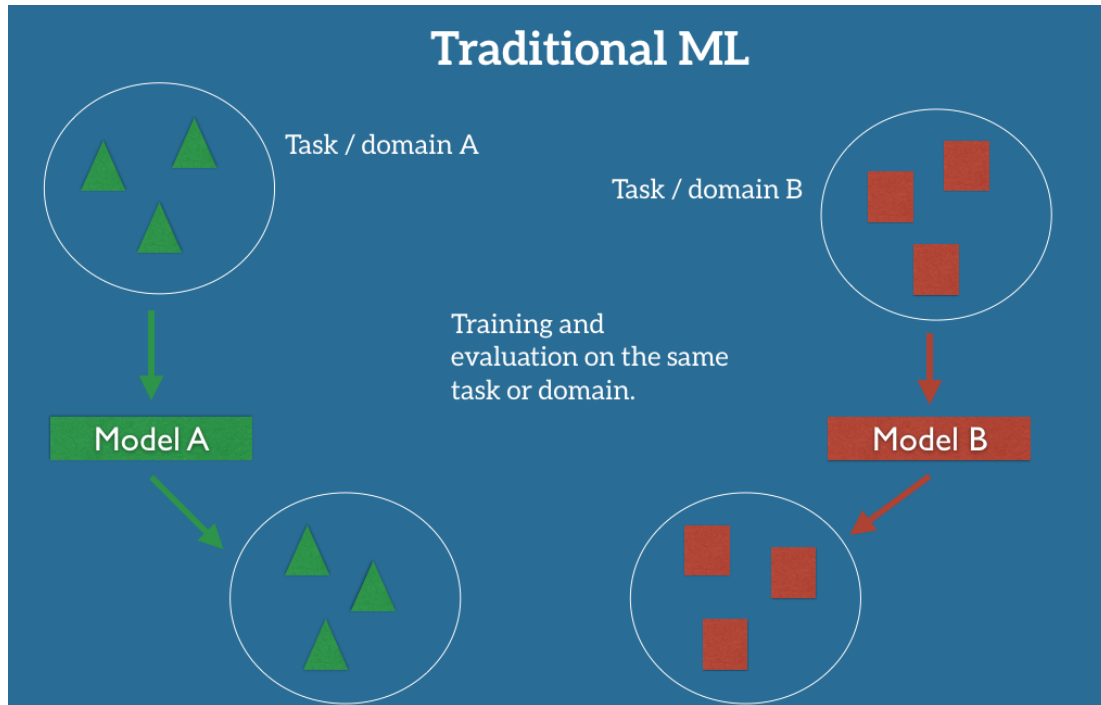


	K=1	K=3	K=9
BOW	0.48	0.58	0.74
FastText	0.66 (+ 0.18)	0.78 (+ 0.20)	0.84 (+ 0.10)
Bert	0.73 (+ 0.25)	0.84 (+ 0.26)	0.89 (+ 0.15)

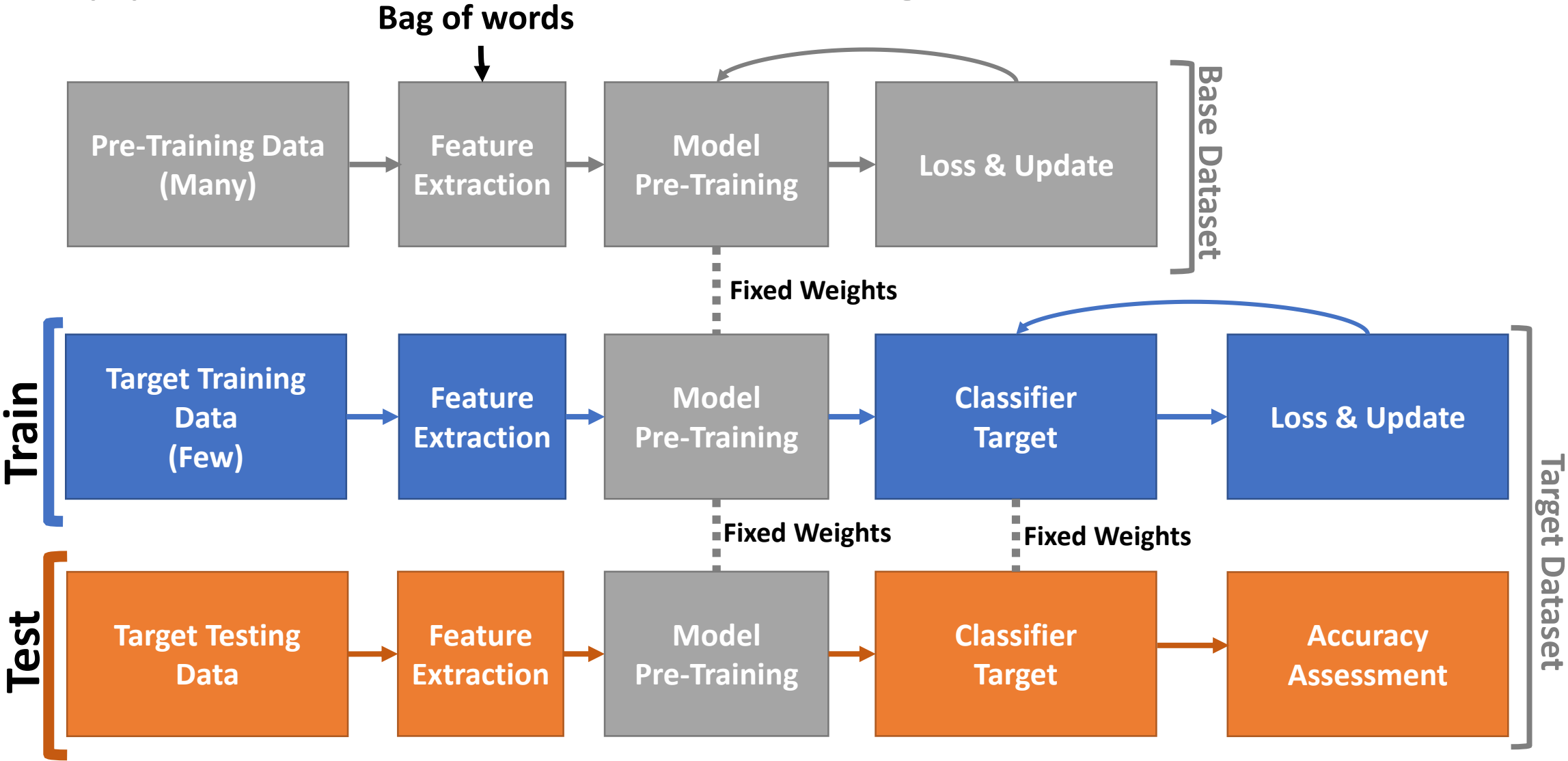
Baseline FastText & Bert



Can we improve any further?

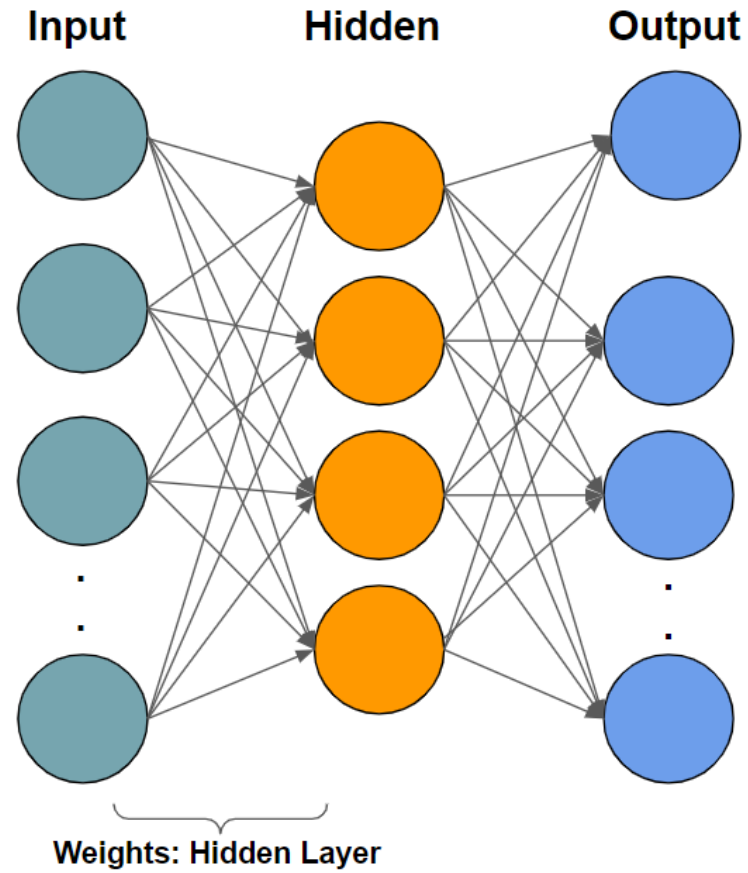


Approach: Transfer Learning



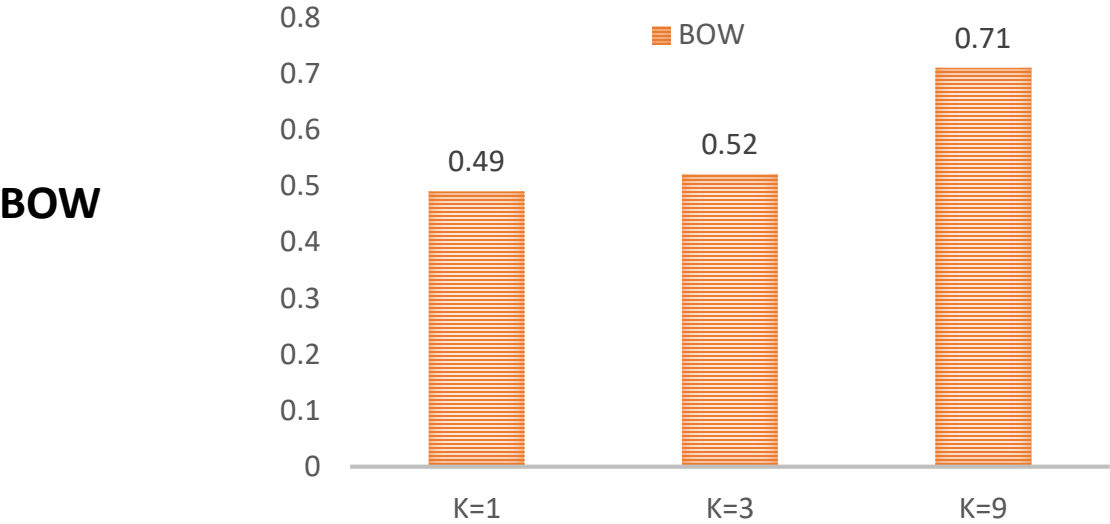
Model

Standard



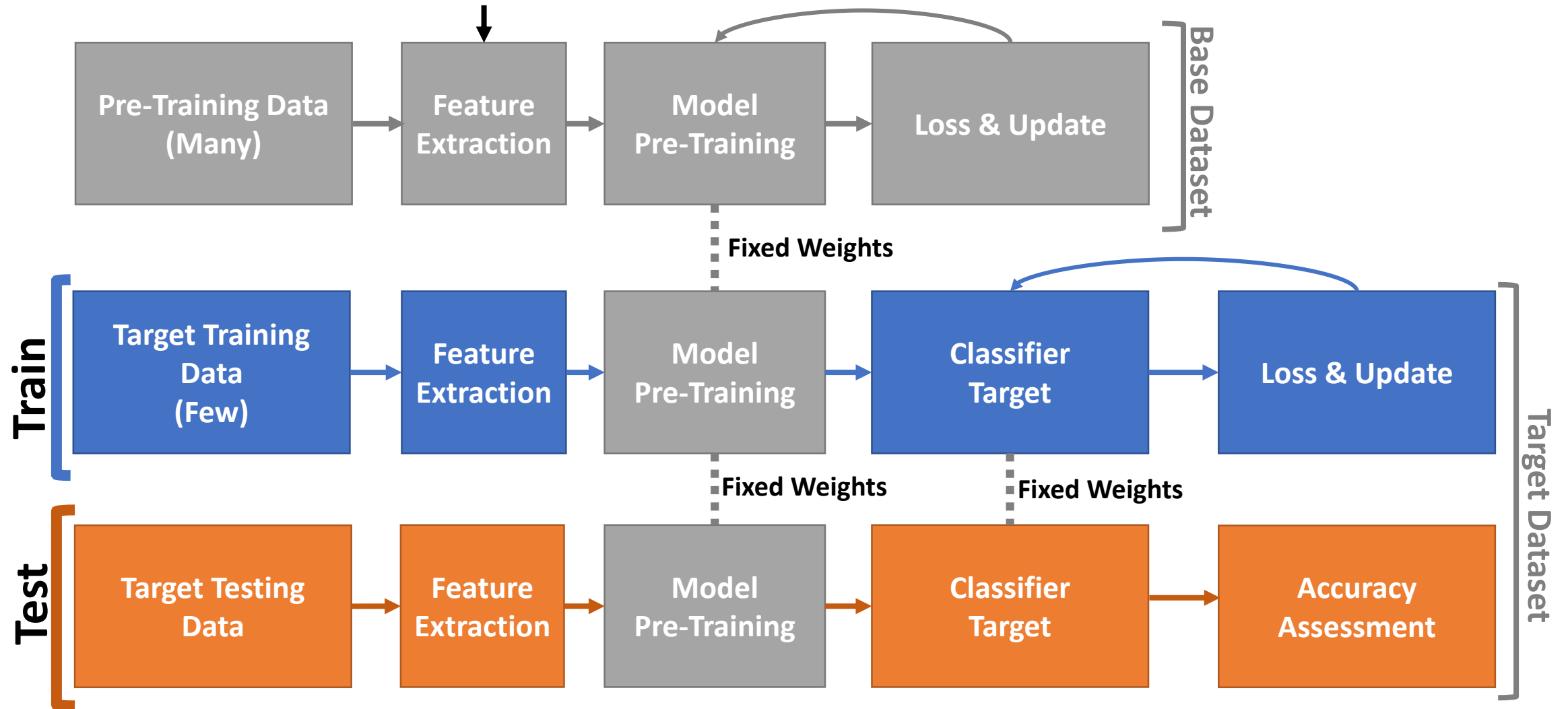
Results

Transfer Learning - Standard Model



Transfer Learning (Standard)	K=1	K=3	K=9
BOW	0.49 (+ 0.01)	0.52 (- 0.06)	0.71 (- 0.03)

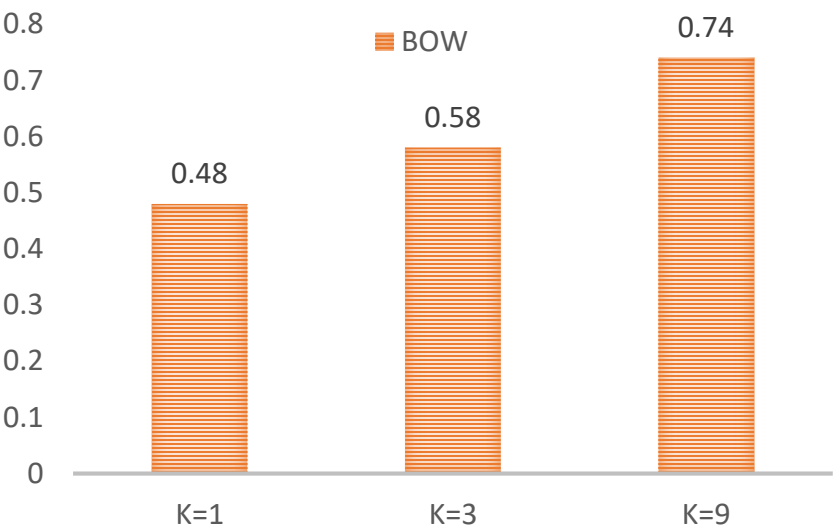
Pretrained FastText & Bert Model



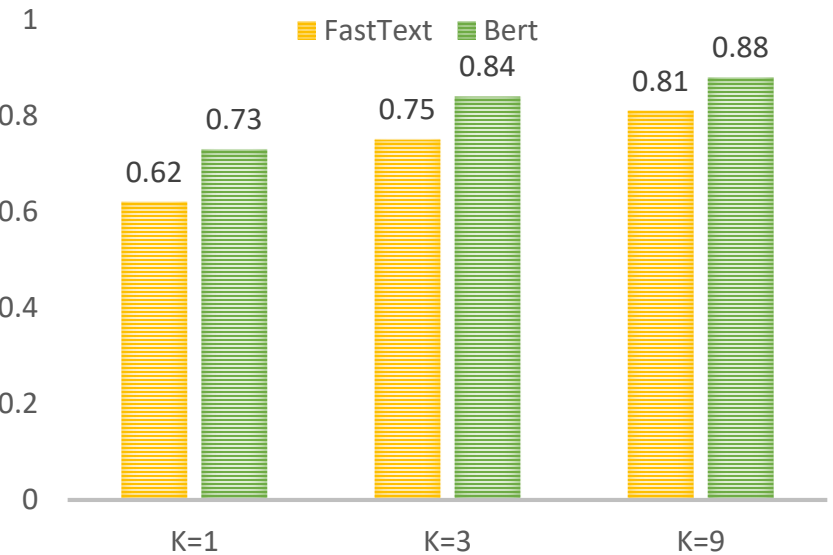
Results

Transfer Learning - Standard Model

BOW



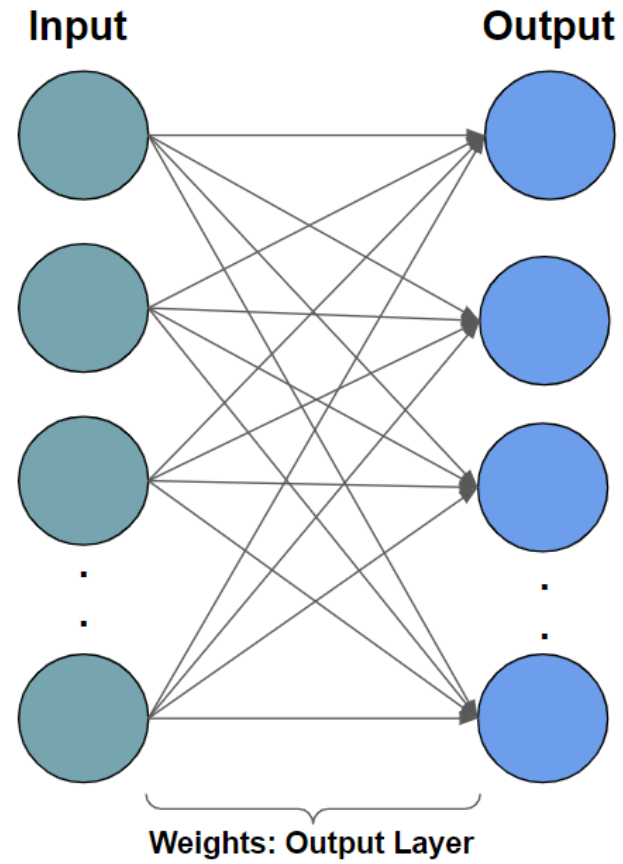
FastText And Bert



Transfer Learning (Standard)	K=1	K=3	K=9
BOW	0.49 (+ 0.01)	0.52 (- 0.06)	0.71 (- 0.03)
FastText	0.62 (- 0.04)	0.75 (- 0.03)	0.81 (- 0.03)
Bert	0.73 (0.00)	0.84 (0.00)	0.88 (-0.01)

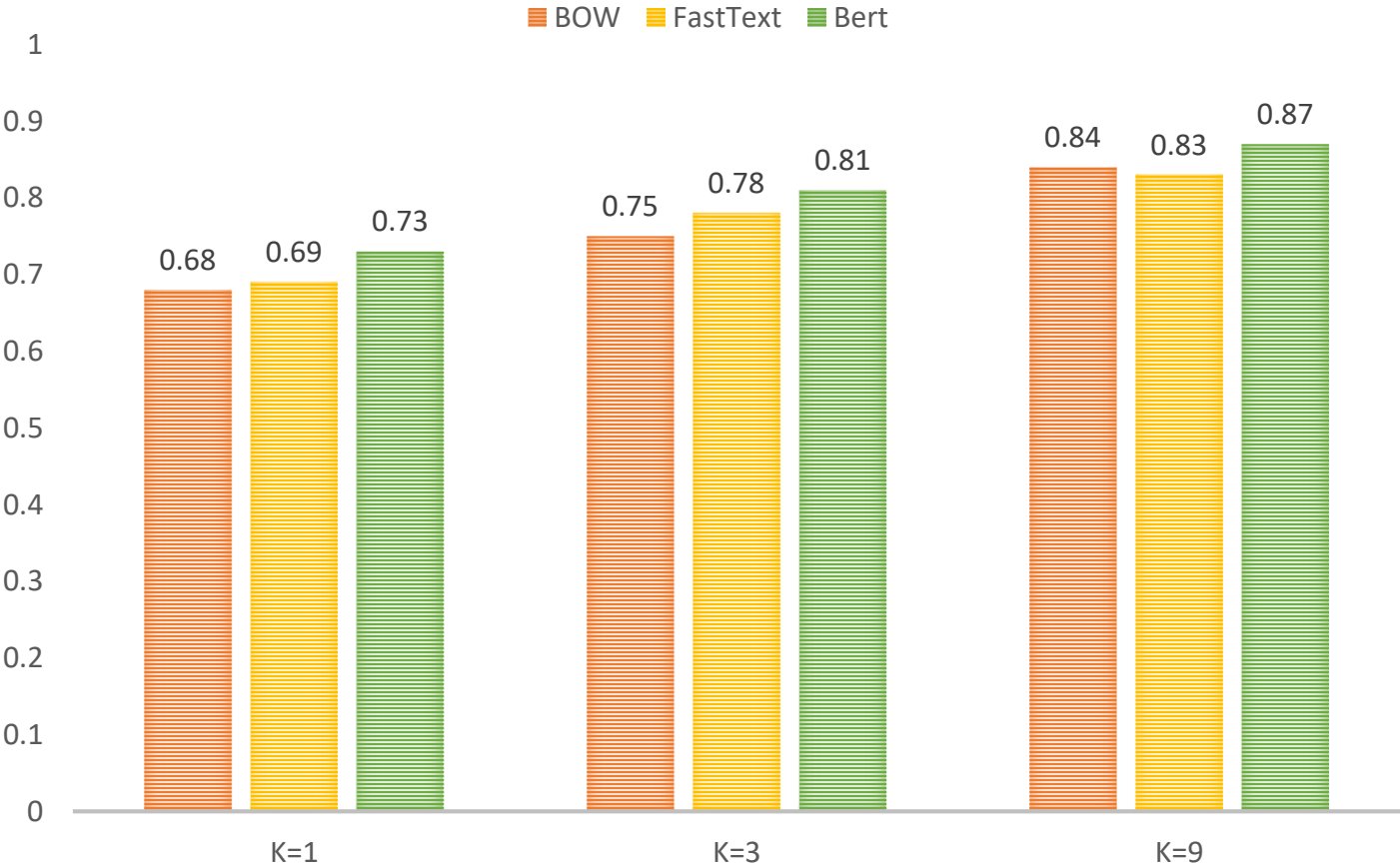
Model

Modified



Results

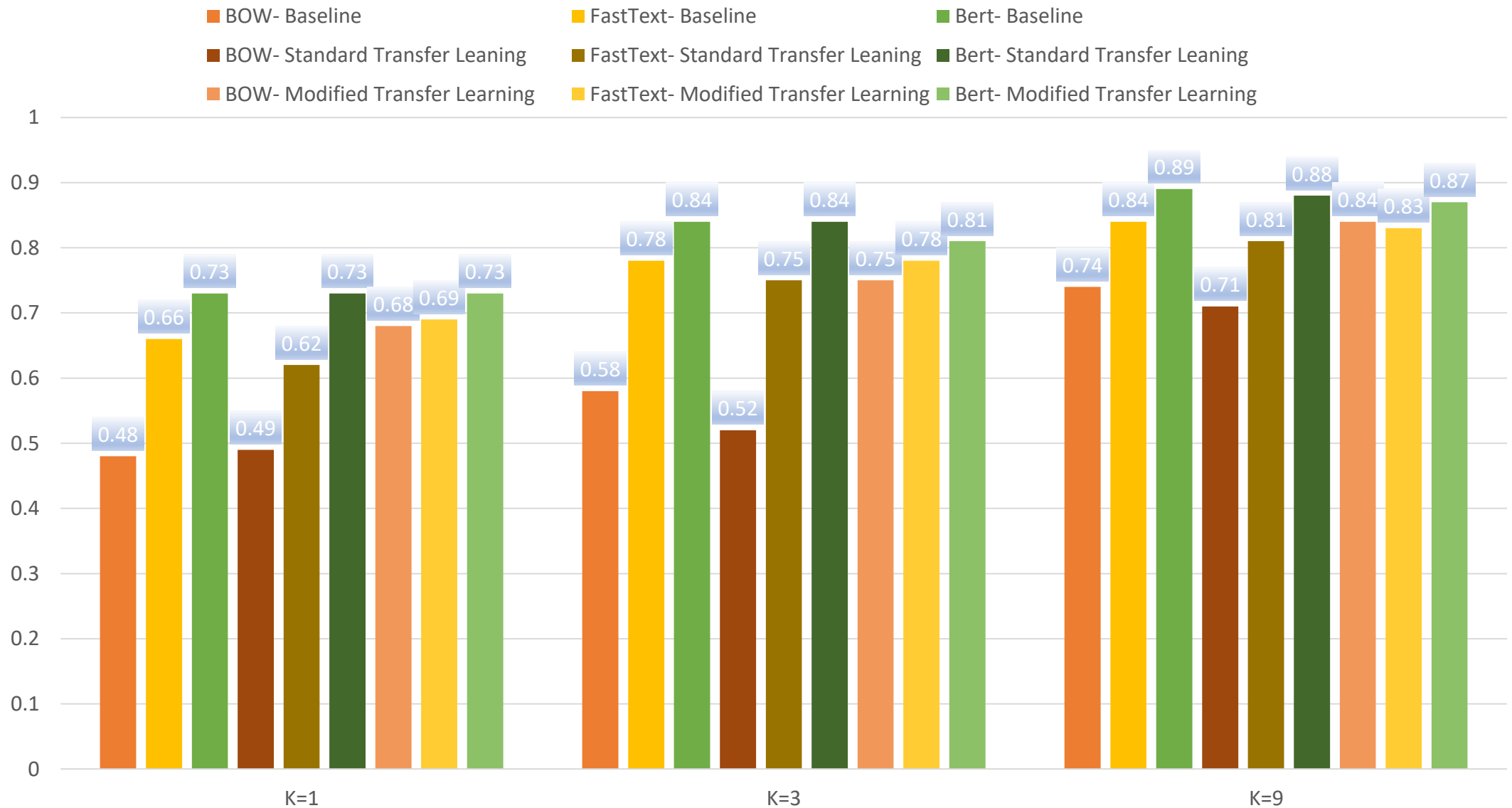
Transfer Learning - Modified Model



Transfer Learning (Modified)	K=1	K=3	K=9
BOW	0.68 (+ 0.20)	0.75 (+ 0.17)	0.84 (+ 0.10)
FastText	0.69 (+ 0.03)	0.78 (0.00)	0.83 (- 0.01)
Bert	0.73 (0.00)	0.81 (- 0.03)	0.87 (- 0.02)

Complete Results

Approaches And Results



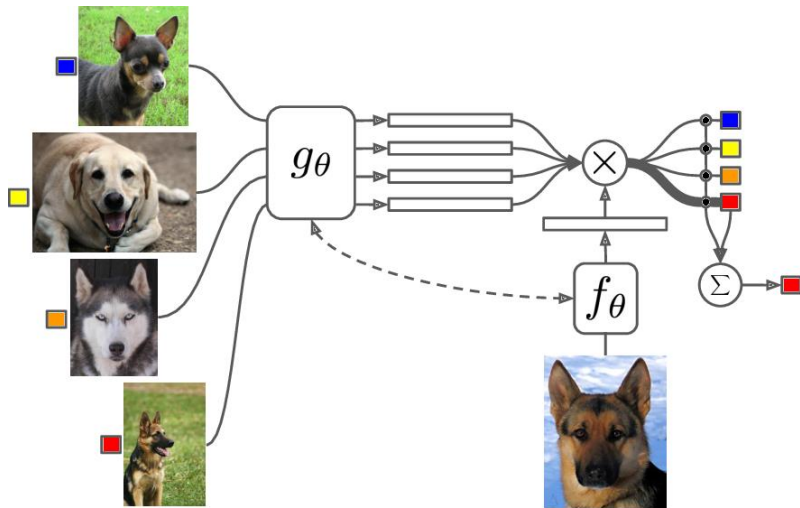
Results Summary

- An average improvement of 10-20% in the modified transfer learning using bow representations as compared to the baseline scores of the bow model.
- A general increase in the accuracy with the increase in the size of training task.
- No real improvements when fine-tuning the representations from both the advanced pre-trained models fasttext and bert.
- Bow representation can be improved by pre-training on Wikipedia section heading classification task.

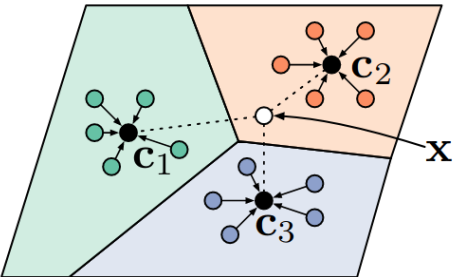
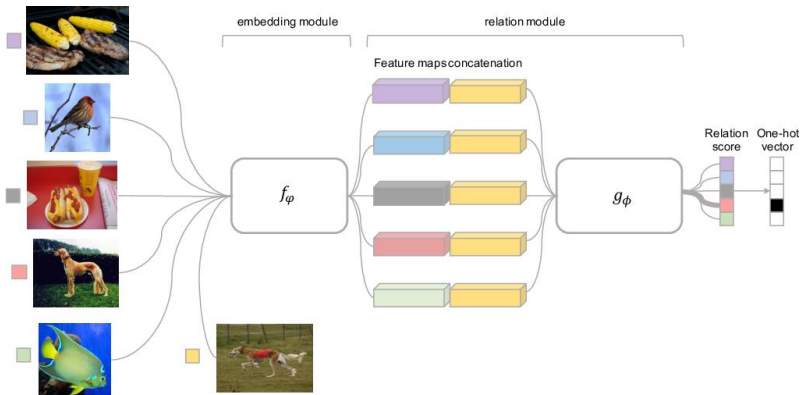
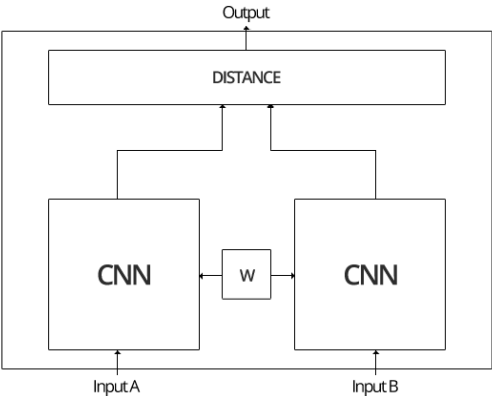
Few-shot learning approaches:

- Metric Learning
- Meta Learning

Metric Learning

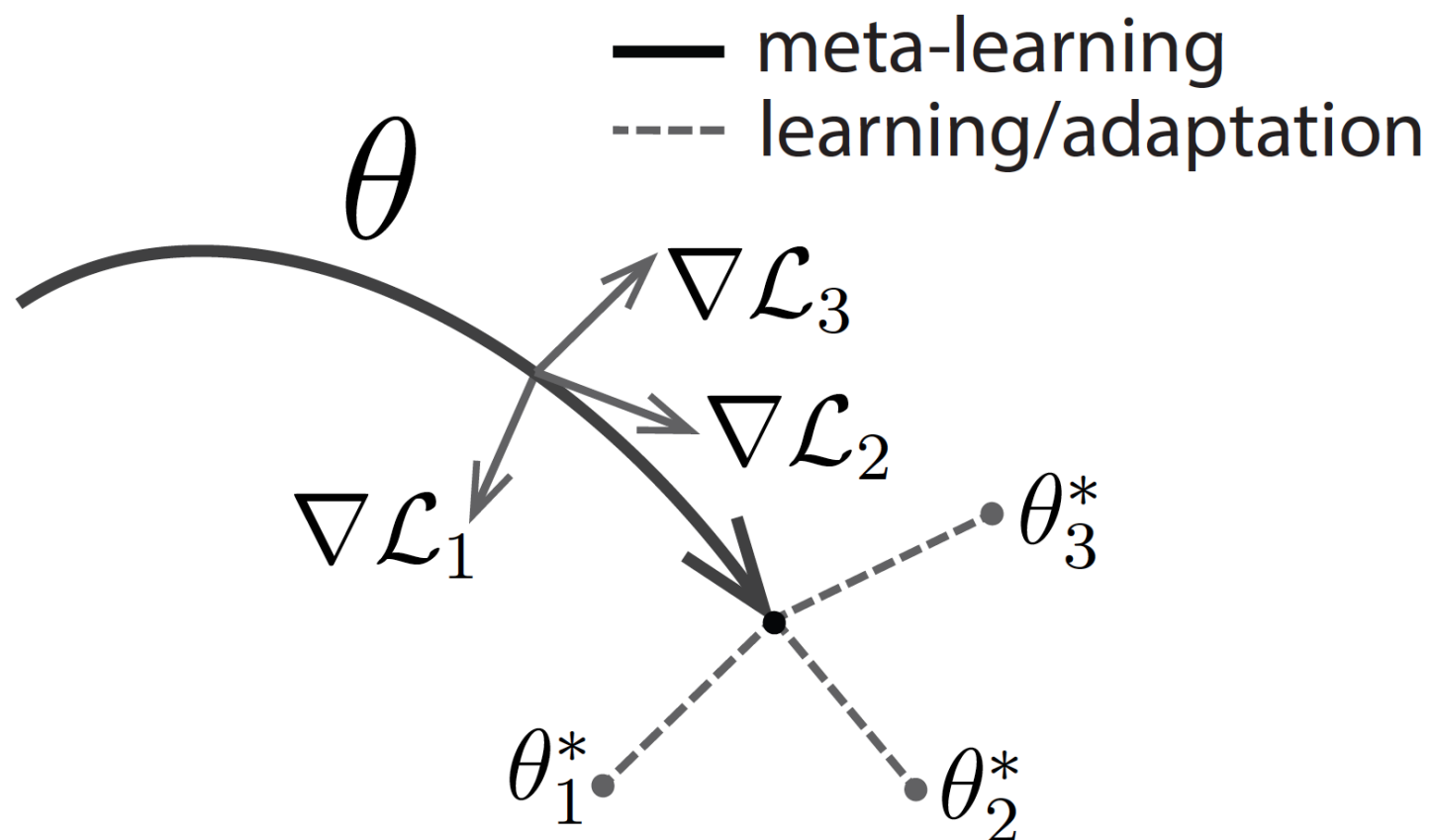


Related Work



[Relation Network](#)
[Advances in few-shot learning](#)
[Siamese](#)

Meta Learning



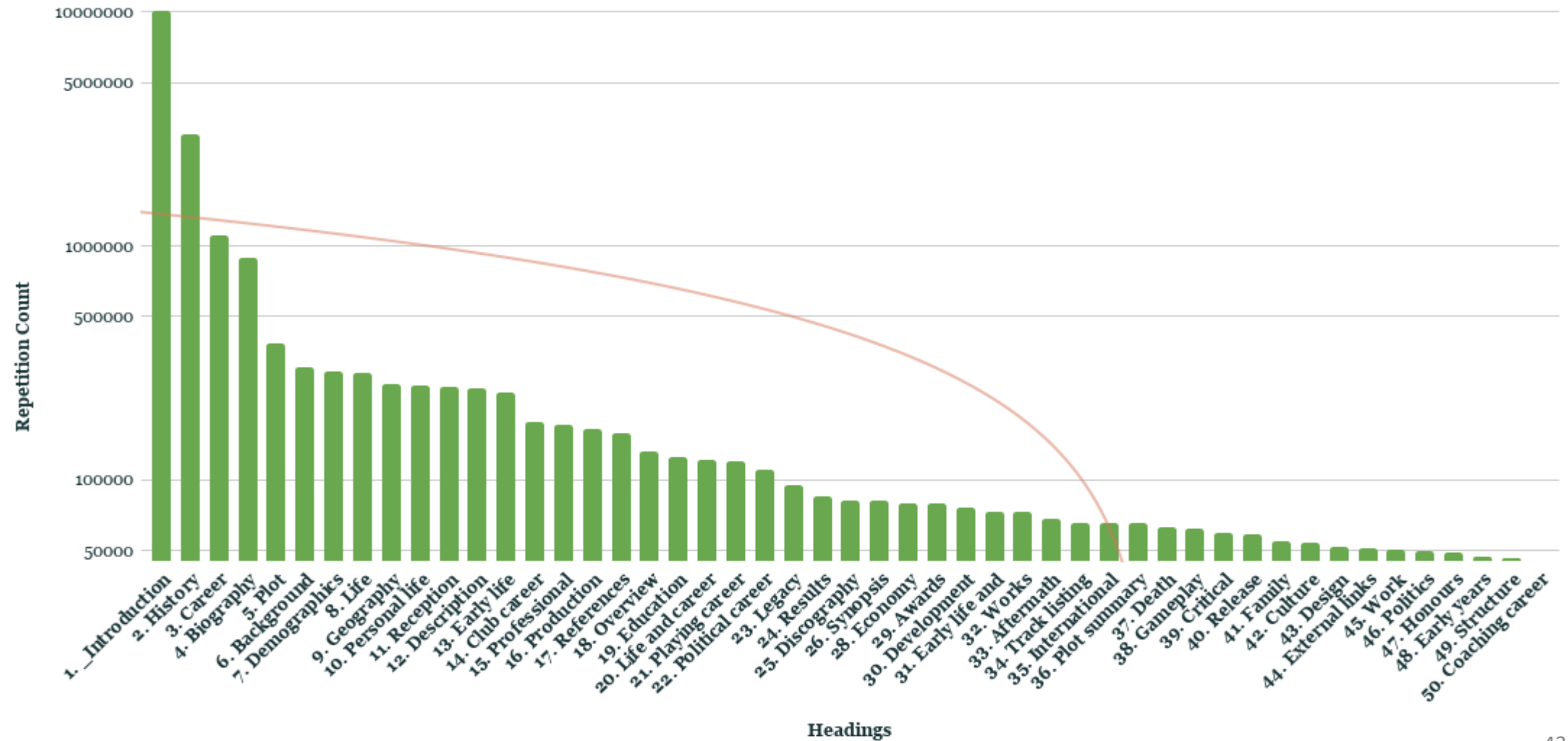
- Using other few-shot learning approaches such as meta learning and metric learning.
- Increasing the dataset by not just limiting to the level 2 section heading- Would require having increased computation resources.
- Using bert-large model instead of bert-base.
- Finding peak accuracy score for bert model.
- Testing the trained classifier on topic classification data other than Wikipedia.

Thank you

Additional Slides

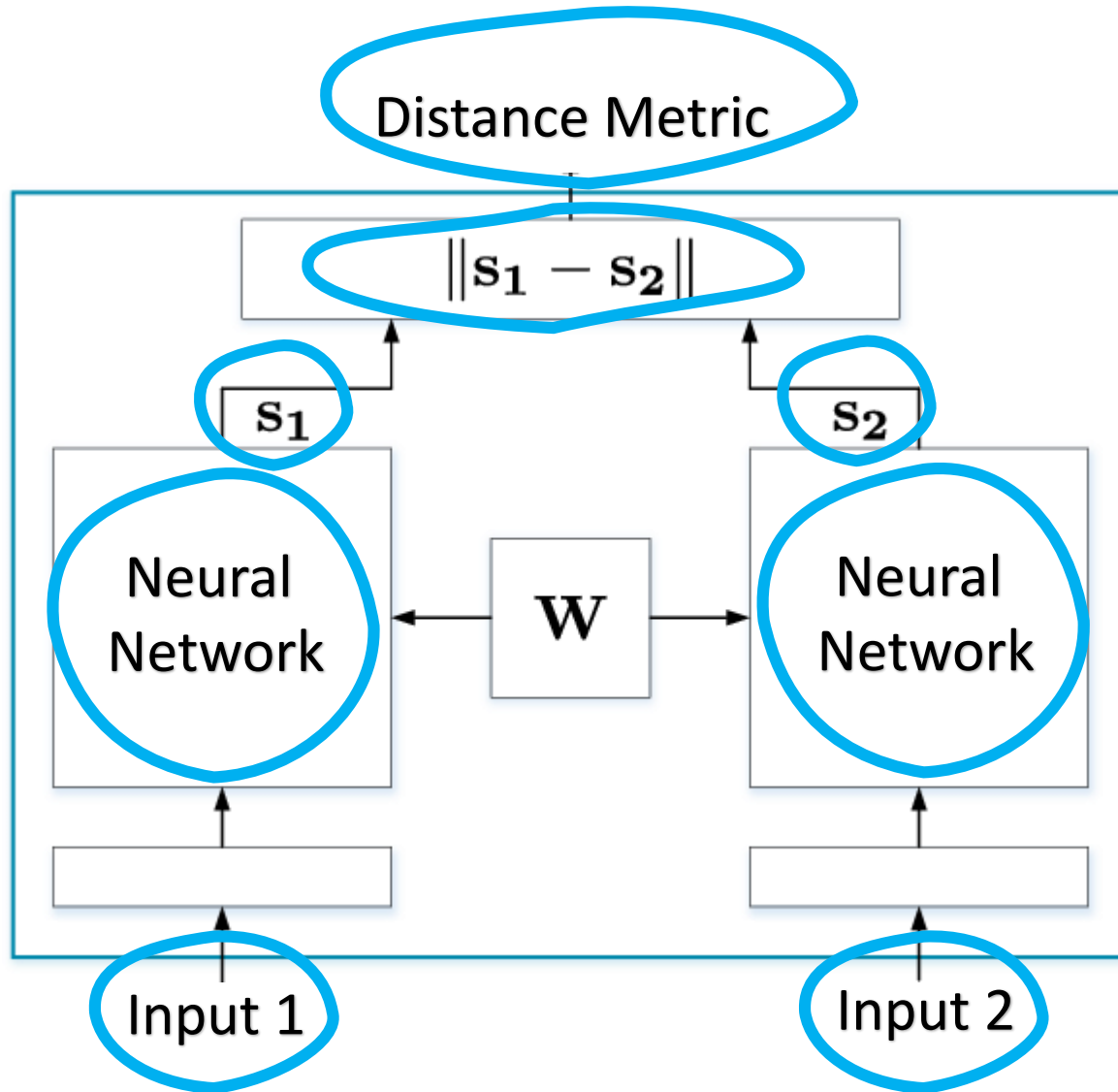
Additional Slides

Repetition Count vs Headings



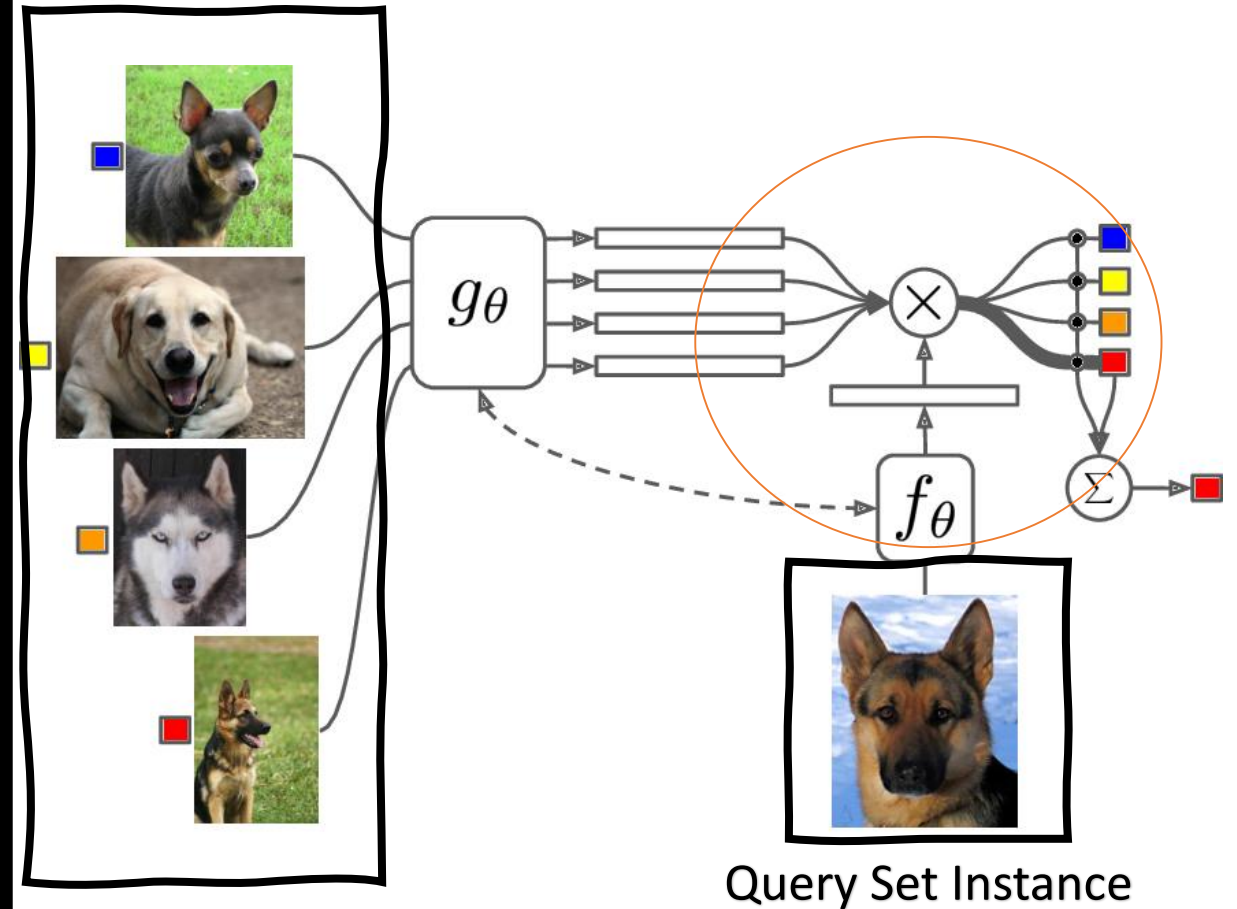
Related Work: Metric Learning

- Siamese



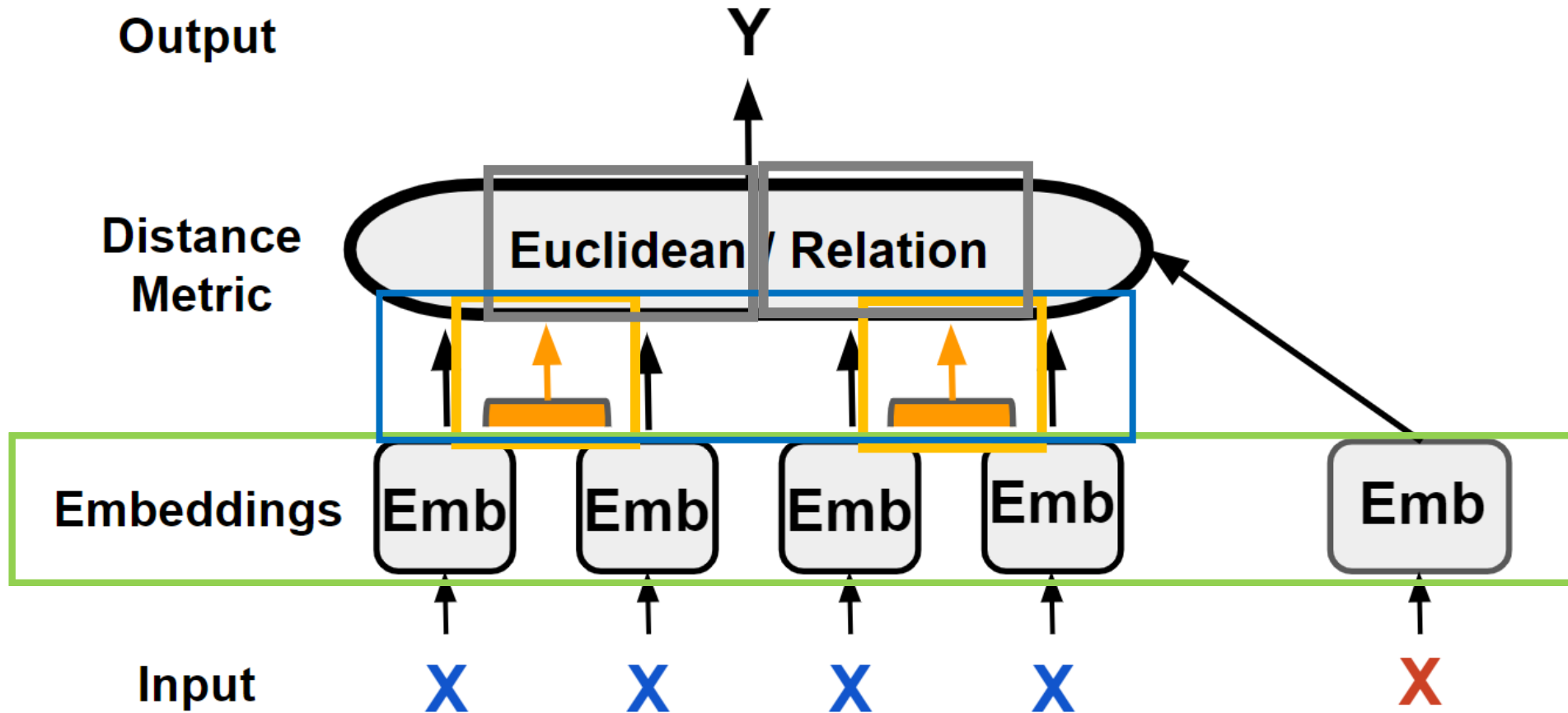
- Matching Networks

Support Set Instances



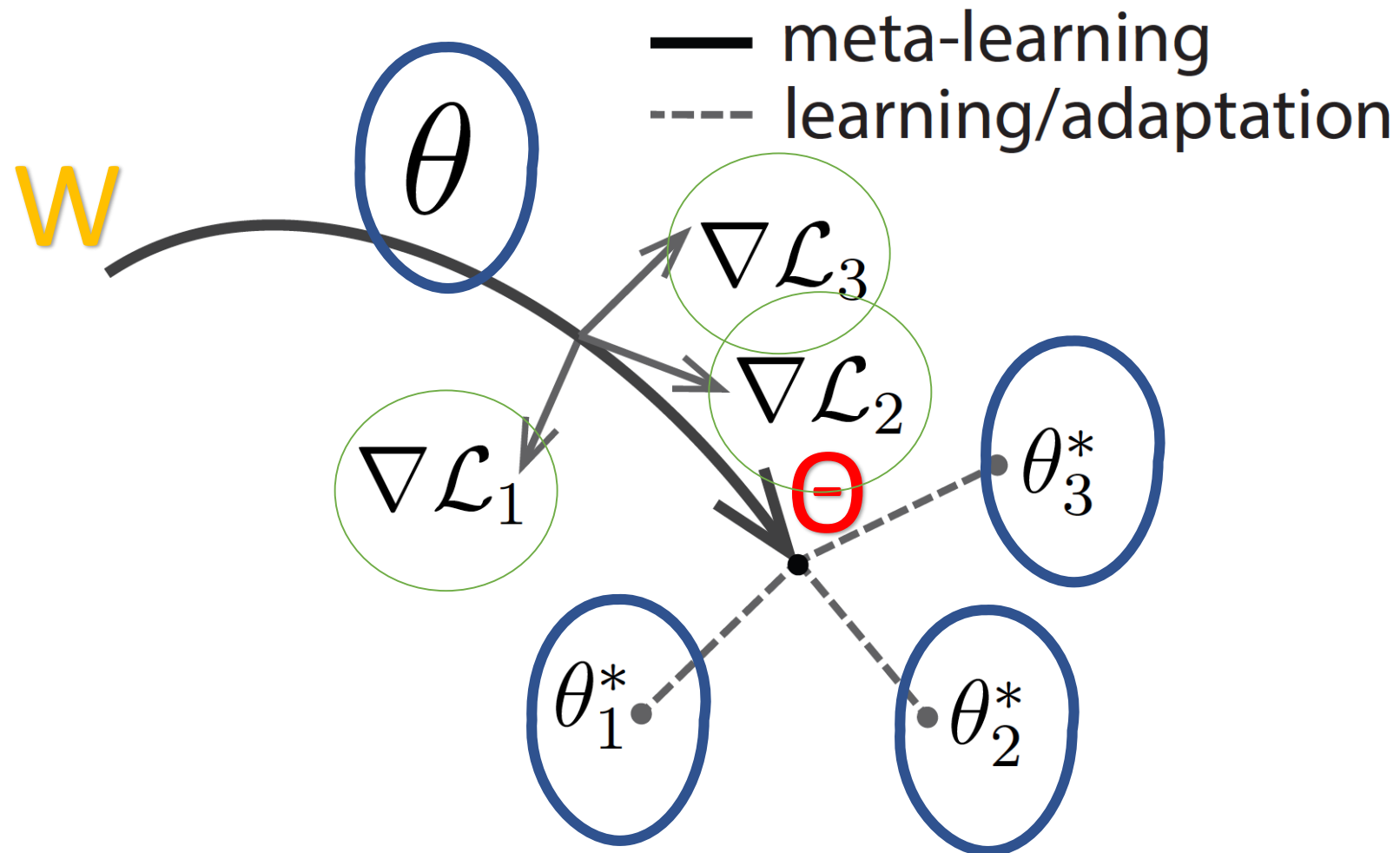
Related Work: Metric Learning

- Prototypical Networks & Relation Networks



● Mean Embeddings / Prototypes ● Query Example ● Support Examples

Related Work: Meta Learning



Related Work: Transfer Learning

- Baseline

