

Leveraging Offensive Language for Sarcasm and Sentiment Detection in Arabic

Fatemah Husain, Kuwait University, f.husain@ku.edu.kw
Ozlem Uzuner, George Mason University, ouzuner@gmu.edu



Introduction

- **Sarcasm detection is one of the top challenging** tasks in text classification, particularly for informal Arabic with high syntactic and semantic ambiguity.
- Previous offensive language detection studies highlight **sarcastic content as a cause of confusion** in classifying offensive language.
- Previous studies demonstrate the **outperformance of dialectal pre-trained model on multiple Twitter text classification tasks** over Modern Standard Arabic (MSA) pre-trained model.

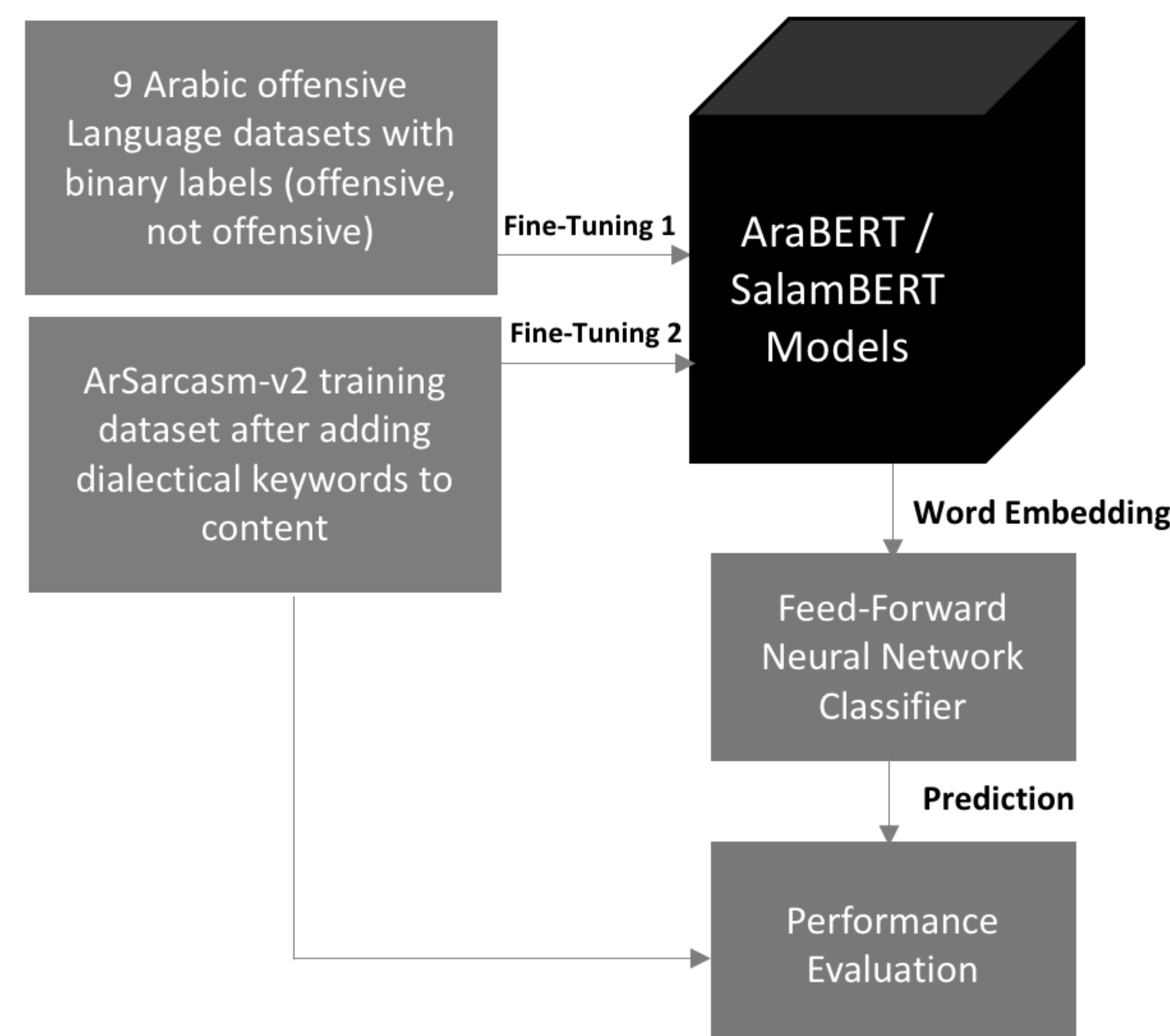


Research Questions

- What is the effect of **sequential transfer learning across offensive language detection, sarcasm detection, and sentiment analysis?**
- What is the variation in performance between a **modern standard Arabic pre-trained language model (AraBERT)** and its **continued dialectal Arabic pre-trained version (SalamBERT)** on the tasks of sarcasm detection and sentiment analysis?

Methodology

- **Method:** Sequential transfer learning from offensive language detection to sarcasm detection and sentiment analysis.
- **Datasets:** ArSarcasm-v2 and nine offensive language datasets.
- **Preprocessing:** adding an Arabic keyword token to refer to the dialect of the tweet (e.g., خليجي /Gulfian).
- **Language Model:** SalamBERT (added dialect) and AraBERT (MSA).
- **Classification Model:** Feed Forward Neural Network layer.
- **Fine-Tuning:** first fine-tuned on offensive language datasets, then on the target task training sets, and evaluated on its testing set.



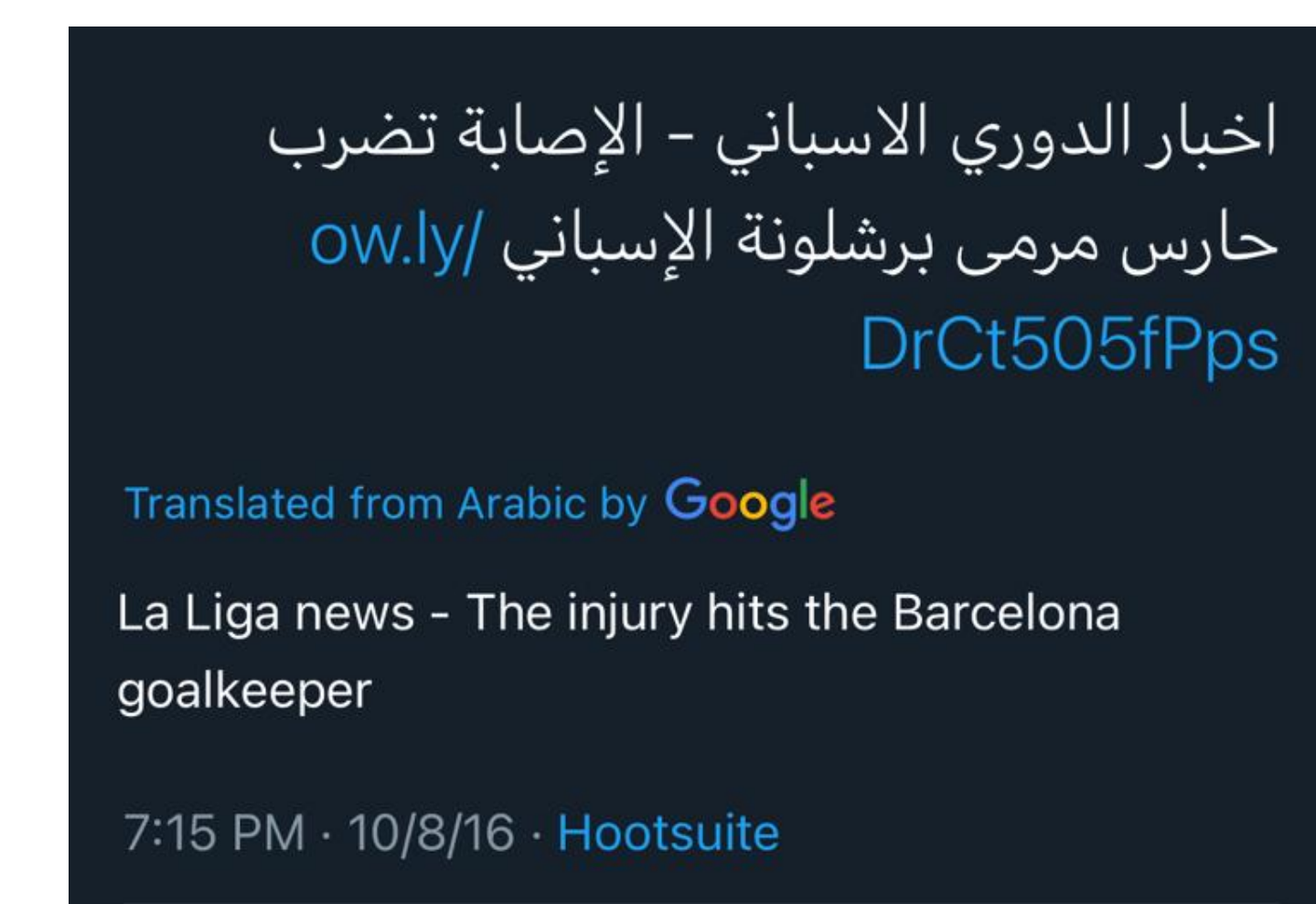
Results

Metric	AraBERT	SalamBERT
F-PN	0.6877	0.6259
Precision	0.6136	0.5580
Recall	0.6318	0.5813
Macro-F1	0.6210	0.5635
Accuracy	0.6630	0.6073

Official shared task results for the testing dataset from sentiment analysis sub-task

Metric	AraBERT	SalamBERT
F1-Sarcasm	0.5041	0.5348
Precision	0.6950	0.7128
Recall	0.6622	0.6807
Macro-F1	0.6732	0.6922
Accuracy	0.7607	0.7727

Official shared task results for the testing dataset from sarcasm detection sub-task



Example of a tweet misclassified by both SalamBERT and AraBERT

Conclusions

- Dialect is more prominent in sarcasm than sentiment:
 - AraBERT does better on sentiment
 - SalamBERT does better on sarcasm