

SENTIMENT ANALYTICS

ON ARABIC SPEECH-ACT TWEET



BACKGROUND

SENTIMENT ANALYSIS IS TO BUILD MACHINE LEARNING MODELS THAT CAN DETERMINE THE TONE (POSITIVE, NEGATIVE, NEUTRAL) OF THE TEXTS (E.G., MOVIE REVIEWS, TWEETS...). IT IS ONE OF THE MOST IMPORTANT AND STANDARD TASKS IN NLP. HOWEVER, ARABIC SENTIMENT ANALYSIS HAS NOT BEEN STUDIED AT A LEVEL AS HIGH AS OTHER LANGUAGES (E.G., ENGLISH, CHINESE, FRENCH). ONE OF THE KEY FACTORS IS THE LACK OF HIGH-QUALITY AND LARGE-SCALE TRAINING DATA.

METHOD

USING ARABIC SPEECH-ACT AND SENTIMENT CORPUS OF TWEETS DATASET WITH 21K TWEETS AND 4 CLASSES: NEGATIVE, POSITIVE, MIXED AND NEUTRAL. IN THIS MODEL WHICH IS A FUSION OF THE CNN AND BiLSTM MODELS, WE KEPT THE SAME ARCHITECTURES PREVIOUSLY USED WITH ONLY ADDING A 0.2 DROPOUT RATE TO THE LSTM CELLS.

IMPORT LIBRARY. CONNECT TO GOOGLE DRIVE TO UPLOAD THE DATASET. CHECKING THE TYPES OF THE FIELDS IN THE DATA AND THEN PRINTING THE DISTRIBUTION OF THE CLASSES. DELETING UNUSED FIELDS AND DO PREPROCESSING DATA. DATA PRE-PROCESSING : REMOVAL OF NOISE, URLs, HASHTAG AND USER-MENTIONS, EMOTICONS AND EMOJIS. REMOVING PUNCTUATIONS. LETTER NORMALIZATION AND REMOVING ELONGATION AND ARABIC DIACRITICS. REMOVING STOPWORDS, STEMMING, AND LEMMATIZATION STEPS WERE IGNORED.

PROCESS

HYPER PARAMETERS

PARAMETER	VALUE
BATCH SIZE	128
LEARNING RATE	0.01
OPTIMIZER	ADAM
NUMBER OF EPOCHS	20

EVALUATION

PRECISION	77%
RECALL	69%
F1 SCORE	72%
ACCURACY	73.02%

CONCLUSION

IN THIS STUDY GAVE QUITE GOOD RESULTS WITH AN AVERAGE EVALUATION MATRIX IN THE RANGE OF 70. IT IS FELT THAT THE MERGER OF CNN AND BiLSTM HAS NOT YIELDED SIGNIFICANT RESULTS. FOR FURTHER RESEARCH CAN CONDUCT EXPERIMENTS USING OTHER METHODS