Pan Lingua Task-1

Survey

Machine Learning Models

Single Method Machine Learning Models:

Kernel method (SVM) Logistic regression Decision tree (J-48graft, RF)

Pros	Cons
Robustness	Low Accuracy
Flexibility	Fragmentation Issue
Scalability	Poor performance on Class Imbalanced Datasets

Hybrid Method Machine Learning Models:

K-means and cuckoo search; Naïve Bayes and logistic regression; Naïve Bayes and particle swarm optimization; Particle swarm optimization and genetic algorithm and decision tree; N-gram and support vector machine; Support vector machine, logistic regression and decision tree

Pros	Cons
Robustness	Higher Time Complexity due to more models being combined
Flexibility	
Scalability	
Adaptability to Different data sizes	
Higher Accuracy	

SVM with SVD Features:

Pros	Cons
Less Computation	Dependent on emotion speech dataset
Faster Classification	Dependent on the features extracted

Deep Learning Models

Deep4Snet to Classify Original from Fake speech

Pros	Cons
High Accuracy	Histogram features needed instead of spectrogram.
Imperfect symmetry of histogram features on the left and right help in faster classification	Poor performance on normal features and entropy-based features

2D CNN for Dari One Word Speech Classification

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 39, 173, 16)	80
max_pooling2d_1 (MaxPooling2	(None, 19, 86, 16)	0
dropout_1 (Dropout)	(None, 19, 86, 16)	0
conv2d_2 (Conv2D)	(None, 18, 85, 32)	2080
max_pooling2d_2 (MaxPooling2	(None, 9, 42, 32)	0
dropout_2 (Dropout)	(None, 9, 42, 32)	0
conv2d_3 (Conv2D)	(None, 8, 41, 64)	8256
max_pooling2d_3 (MaxPooling2	(None, 4, 20, 64)	0
dropout_3 (Dropout)	(None, 4, 20, 64)	0
conv2d_4 (Conv2D)	(None, 3, 19, 128)	32896
max_pooling2d_4 (MaxPooling2	(None, 1, 9, 128)	0
dropout_4 (Dropout)	(None, 1, 9, 128)	0
global_average_pooling2d_1 (	(None, 128)	0
dense_1 (Dense)	(None, 20)	2580

Total params: 45,892
Trainable params: 45,892
Non-trainable params: 0

Pros	Cons
Fast computation	The Model layers can always be optimized to obtain a better accuracy
Fast classification	May or may not need to resample the audio files in the dataset to increase accuracy

RNN

Pros	Cons

Any input length	No use of additional memory as it is speech data and not time series data.
Model size doesn't increase with increase of input layers.	Slow computation due to this recurrent nature