TF-IDF

Here we will run TF-IDF algorithm with different classifiers (Linear SVC, Logistic Regression, neural network MLP classifier)

SVM classifier with alpha = 1.0

Accuracy: 0.85

SVM classifier with alpha = 0.5

Accuracy: 0.8425

logistic classifier with alpha = 1.0

Accuracy: 0.8325

logistic classifier with alpha = 0.5

Accuracy: 0.815

neural network MLP classifier with hidden layers = 50 and maxIterations = 1000

Accuracy: 0.8175

neural network MLP classifier with hidden layers = 20 and maxIterations = 900

Accuracy: 0.82

Best accuracy is with SVM classifier with alpha = 1.0

Accuracy: 0.85

Word2vec

Here we will run word2vec algorithm with different classifiers (Linear SVC, Logistic Regression, neural network MLP classifier) and with different parameters

We got those results:

change window and min_Count parameters

parameters = (vecSize=200,window=10,minCount=2,epochs=10,workers=5,sg=1)

SVM classifier with alpha = 1.0 Accuracy: 0.8575

SVM classifier with alpha = 0.5 Accuracy: 0.84

logistic classifier with alpha = 1.0 Accuracy: 0.7825

logistic classifier with alpha = 0.5 Accuracy: 0.7575

neural network MLP classifier with hidden layers = 120 and maxIterations = 1000

Accuracy: 0.86

neural network MLP classifier with hidden layers = 150 and maxIterations = 800

Accuracy: 0.8625

parameters = (vecSize=150,window=8,minCount=2,epochs=10,workers=10,sg=1)

SVM classifier with alpha = 1.0 Accuracy: 0.84

SVM classifier with alpha = 0.5 Accuracy: 0.8375

logistic classifier with alpha = 1.0 Accuracy: 0.77

logistic classifier with alpha = 0.5 Accuracy: 0.74

neural network MLP classifier with hidden layers = 120 and maxIterations = 1000

Accuracy: 0.85

neural network MLP classifier with hidden layers = 150 and maxIterations = 800

Accuracy: 0.8375

parameters = (vecSize=300,window=5,minCount=2,epochs=10,workers=10,sg=1)

SVM classifier with alpha = 1.0 Accuracy: 0.8125

SVM classifier with alpha = 0.5 Accuracy: 0.8075

logistic classifier with alpha = 1.0 Accuracy: 0.765

logistic classifier with alpha = 0.5 Accuracy: 0.74

neural network MLP classifier with hidden layers = 120 and maxIterations = 1000

Accuracy: 0.82

neural network MLP classifier with hidden layers = 150 and maxIterations = 800

Accuracy: 0.815

parameters = (vecSize=500,window=10,minCount=5,epochs=10,workers=10,sg=1)

SVM classifier with alpha = 1.0 Accuracy: 0.8525

SVM classifier with alpha = 0.5 Accuracy: 0.845

logistic classifier with alpha = 1.0 Accuracy: 0.7825

logistic classifier with alpha = 0.5 Accuracy: 0.75

neural network MLP classifier with hidden layers = 120 and maxIterations = 1000

Accuracy: 0.85

neural network MLP classifier with hidden layers = 150 and maxIterations = 800

Accuracy: 0.87

change workers parameter

parameters = (vecSize=150,window=10,minCount=2,epochs=10,workers=5,sg=1)

SVM classifier with alpha = 1.0 Accuracy: 0.8475

SVM classifier with alpha = 0.5 Accuracy: 0.8375

logistic classifier with alpha = 1.0 Accuracy: 0.775

logistic classifier with alpha = 0.5 Accuracy: 0.745

neural network MLP classifier with hidden layers = 120 and maxIterations = 1000

Accuracy: 0.855

neural network MLP classifier with hidden layers = 150 and maxIterations = 800

Accuracy: 0.8475

parameters = (vecSize=300,window=5,minCount=2,epochs=10,workers=10,sg=1)

SVM classifier with alpha = 1.0 Accuracy: 0.825

SVM classifier with alpha = 0.5 Accuracy: 0.81

logistic classifier with alpha = 1.0 Accuracy: 0.7725

logistic classifier with alpha = 0.5 Accuracy: 0.7425

neural network MLP classifier with hidden layers = 120 and maxIterations = 1000

Accuracy: 0.82

neural network MLP classifier with hidden layers = 150 and maxIterations = 800

Accuracy: 0.82

parameters = (vecSize=500,window=10,minCount=5,epochs=10,workers=20,sg=1)

SVM classifier with alpha = 1.0 Accuracy: 0.86

SVM classifier with alpha = 0.5 Accuracy: 0.8525

logistic classifier with alpha = 1.0 Accuracy: 0.7875

logistic classifier with alpha = 0.5 Accuracy: 0.765

neural network MLP classifier with hidden layers = 120 and maxIterations = 1000

Accuracy: 0.86

neural network MLP classifier with hidden layers = 150 and maxIterations = 800

Accuracy: 0.83

change epochs parameter

parameters = (vecSize=150,window=10,minCount=2,epochs=20,workers=10,sg=1)

SVM classifier with alpha = 1.0 Accuracy: 0.8525

SVM classifier with alpha = 0.5 Accuracy: 0.855

logistic classifier with alpha = 1.0 Accuracy: 0.825

logistic classifier with alpha = 0.5 Accuracy: 0.7925

neural network MLP classifier with hidden layers = 120 and maxIterations = 1000

Accuracy: 0.875

neural network MLP classifier with hidden layers = 150 and maxIterations = 800

Accuracy: 0.8475

parameters = (vecSize=300,window=5,minCount=2,epochs=10,workers=10,sg=1)

SVM classifier with alpha = 1.0 Accuracy: 0.8275

SVM classifier with alpha = 0.5 Accuracy: 0.815

logistic classifier with alpha = 1.0 Accuracy: 0.785

logistic classifier with alpha = 0.5 Accuracy: 0.745

neural network MLP classifier with hidden layers = 120 and maxIterations = 1000

Accuracy: 0.8175

neural network MLP classifier with hidden layers = 150 and maxIterations = 800

Accuracy: 0.825

parameters = (vecSize=200,window=10,minCount=5,epochs=15,workers=10,sg=1)

SVM classifier with alpha = 1.0 Accuracy: 0.87

SVM classifier with alpha = 0.5 Accuracy: 0.865

logistic classifier with alpha = 1.0 Accuracy: 0.8075

logistic classifier with alpha = 0.5 Accuracy: 0.7675

neural network MLP classifier with hidden layers = 120 and maxIterations = 1000

Accuracy: 0.855

neural network MLP classifier with hidden layers = 150 and maxIterations = 800

Accuracy: 0.86

parameters = (vecSize=500,window=10,minCount=5,epochs=25,workers=10,sg=1)

SVM classifier with alpha = 1.0 Accuracy: 0.8725

SVM classifier with alpha = 0.5 Accuracy: 0.87

logistic classifier with alpha = 1.0 Accuracy: 0.8325

logistic classifier with alpha = 0.5 Accuracy: 0.8

neural network MLP classifier with hidden layers = 120 and maxIterations = 1000

Accuracy: 0.84

neural network MLP classifier with hidden layers = 150 and maxIterations = 800

Accuracy: 0.8475

change sg =0 (CBOW)

parameters = (vecSize=200,window=10,minCount=2,epochs=10,workers=5,sg=0)

SVM classifier with alpha = 1.0 Accuracy: 0.7725

SVM classifier with alpha = 0.5 Accuracy: 0.76

logistic classifier with alpha = 1.0 Accuracy: 0.7675

logistic classifier with alpha = 0.5 Accuracy: 0.7625

neural network MLP classifier with hidden layers = 120 and maxIterations = 1000

Accuracy: 0.7475

neural network MLP classifier with hidden layers = 150 and maxIterations = 800

Accuracy: 0.7675

parameters = (vecSize=500,window=10,minCount=5,epochs=25,workers=10,sg=0)

SVM classifier with alpha = 1.0 Accuracy: 0.845

SVM classifier with alpha = 0.5 Accuracy: 0.845

logistic classifier with alpha = 1.0 Accuracy: 0.845

logistic classifier with alpha = 0.5 Accuracy: 0.825

neural network MLP classifier with hidden layers = 120 and maxIterations = 1000

Accuracy: 0.82

neural network MLP classifier with hidden layers = 150 and maxIterations = 800

Accuracy: 0.815

parameters = (vecSize=150,window=10,minCount=2,epochs=20,workers=10,sg=0)

SVM classifier with alpha = 1.0 Accuracy: 0.825

SVM classifier with alpha = 0.5 Accuracy: 0.82

logistic classifier with alpha = 1.0 Accuracy: 0.81

logistic classifier with alpha = 0.5 Accuracy: 0.81

neural network MLP classifier with hidden layers = 120 and maxIterations = 1000

Accuracy: 0.7825

neural network MLP classifier with hidden layers = 150 and maxIterations = 800

Accuracy: 0.785

Best accuracy is with

parameters = (vecSize=150,window=10,minCount=2,epochs=20,workers=10,sg=1)

with neural network MLP classifier with hidden layers = 120 and max Iterations = 1000 **Accuracy: 0.875**

and with

parameters = (vecSize=500,window=10,minCount=5,epochs=25,workers=10,sg=1)

with SVM classifier with alpha = 1.0 **Accuracy: 0.8725**

Comparison Results

Results of two algorithms are similar
TF-IDF = **0.85** and word2vec Skip diagram = **0.875**

word2vec Skip diagram is better