parameters

Gerbrich Ferdinands

1/14/2020

ASReview takes the following parameters/arguments:

- number of training data (included/excluded)
- a model
- a query strategy
- a balance strategy (fixed)
- feature extraction

	Configurations
Models	Naive Bayes, Random Forest, Support Vector Machine, Logistic Regression
Query Strategies	Cluster Sampling, Maximum Sampling, Cluster * Maximum Sampling, Maximum * Uncertainty Sampling, Maximum * Random Sampling, Cluster * Uncertainty Sampling, Cluster * Random Sampling
Feature extraction strategies	Doc2Vec, tf-idf, sbert, embeddingIdf
Training data [included/excluded]	10/10, 5/5, 5/10

This leads to 273 combinations of configurations.

- Naive bayes only goes with tfidf feature extraction.
- For the feature extraction strategies we will focus on doc2vec and tfidf. (but will compute all 4)
- This leads to 3 * 7 * 4 * 3 + 1 * 7 * 1 * 3 = 273 combinations.

Model	Query Strategy	Feature extraction strategy	Training data [included/excluded]
nb	cluster	tfidf	10/10
nb	max	tfidf	10/10
nb	max * cluster	tfidf	10/10
nb	max * uncertainty	tfidf	10/10
nb	max * random	tfidf	10/10
nb	cluster * uncertainty	tfidf	10/10
nb	cluster * random	tfidf	10/10
nb	cluster	tfidf	5/5
nb	max	tfidf	5/5
nb	max * cluster	tfidf	5/5
nb	max * uncertainty	tfidf	5/5
nb	max * random	tfidf	5/5
nb	cluster * uncertainty	tfidf	5/5
nb	cluster * random	tfidf	5/5
nb	cluster	tfidf	5/10
nb	max	tfidf	5/10
nb	max * cluster	tfidf	5/10
nb	max * uncertainty	tfidf	5/10
nb	max * random	tfidf	5/10

(continued)

Model	Query Strategy	Feature extraction strategy	Training data [included/excluded]
nb	cluster * uncertainty	tfidf	5/10
nb	cluster * random	tfidf	5/10
rf	cluster	doc2vec	10/10
rf	max	doc2vec	10/10
rf	max * cluster	doc2vec	10/10
rf	max * uncertainty	doc2vec	10/10
rf	max * random	doc2vec	10/10
rf	cluster * uncertainty	doc2vec	10/10
rf	cluster * random	doc2vec	10/10
rf	cluster	doc2vec	5/5
rf	max	doc2vec	5/5
rf	max * cluster	doc2vec	5/5
rf	max * uncertainty	doc2vec	5/5
rf	max * random	doc2vec	5/5
rf	cluster * uncertainty	doc2vec	5/5
rf	cluster * random	doc2vec	5/5
rf	cluster	doc2vec	5/10
rf	max	doc2vec	5/10
rf	max * cluster	doc2vec	5/10
rf	max * uncertainty	doc2vec	5/10
rf	max * random	doc2vec	5/10
rf	cluster * uncertainty	doc2vec	5/10
rf	cluster * random	doc2vec	5/10
rf	cluster	tfidf	10/10
rf	max	tfidf	10/10
rf	max * cluster	tfidf	10/10
rf	max * uncertainty	tfidf	10/10
rf	max * random	tfidf	10/10
rf	cluster * uncertainty	tfidf	10/10
rf	cluster * random	tfidf	10/10
rf	cluster	tfidf	5/5
rf	max	tfidf	5/5
rf	max * cluster	tfidf	5/5
rf	max * uncertainty	tfidf	5/5
rf	max * random	tfidf	5/5
rf	cluster * uncertainty	tfidf	5/5
rf	cluster * random	tfidf	5/5
rf	cluster	tfidf	5/10
rf	max	tfidf	5/10
rf	max * cluster	tfidf	5/10
rf	max * uncertainty	tfidf	5/10
rf	max * random	tfidf	5/10
rf	cluster * uncertainty	tfidf	5/10
rf	cluster * random	tfidf	5/10
rf	cluster	sbert	10/10
rf	max	sbert	10/10

Model	Query Strategy	Feature extraction strategy	Training data [included/excluded]
rf	max * cluster	sbert	10/10
rf	max * uncertainty	sbert	10/10
rf	max * random	sbert	10/10
rf	cluster * uncertainty	sbert	10/10
rf	cluster * random	sbert	10/10
rf	cluster	sbert	5/5
rf	max	sbert	5/5
rf	max * cluster	sbert	5/5
rf	max * uncertainty	sbert	5/5
rf	max * random	sbert	5/5
rf	cluster * uncertainty	sbert	5/5
rf	cluster * random	sbert	5/5
rf	cluster	sbert	5/10
rf	max	sbert	5/10
rf	max * cluster	sbert	5/10
rf	max * uncertainty	sbert	5/10
rf	max * random	sbert	5/10
rf	cluster * uncertainty	sbert	5/10
rf	cluster * random	sbert	5/10
rf	cluster	embeddingIdf	10/10
rf	max	embeddingIdf	10/10
rf	max * cluster	embeddingIdf	10/10
rf	max * uncertainty	embeddingIdf	10/10
rf	max * random	embeddingIdf	10/10
rf	cluster * uncertainty	embeddingIdf	10/10
rf	cluster * random	embeddingIdf	10/10
rf	cluster	embeddingIdf	5/5
rf	max	embeddingIdf	5/5
rf	max * cluster	embeddingIdf	5/5
rf	max * uncertainty	embeddingIdf	5/5
rf	max * random	embeddingIdf	5/5
rf	cluster * uncertainty	embeddingIdf	5/5
rf	cluster * random	embeddingIdf	5/5
rf	cluster	embeddingIdf	5/10
rf	max	embeddingIdf	5/10
rf	max * cluster	embeddingIdf	5/10
rf	max * uncertainty	embeddingIdf	5/10
rf	max * random	embeddingIdf	5/10
rf	cluster * uncertainty	embeddingIdf	5/10
rf	cluster * random	embeddingIdf	5/10
svm	cluster	doc2vec	10/10
svm	max	doc2vec	10/10
svm	max * cluster	doc2vec	10/10
svm	max * uncertainty	doc2vec	10/10
svm	max * random	doc2vec	10/10
svm	cluster * uncertainty	doc2vec	10/10

(continued)

Model	Query Strategy	Feature extraction strategy	Training data [included/excluded]
svm	cluster * random	doc2vec	10/10
svm	cluster	doc2vec	5/5
svm	max	doc2vec	5/5
svm	max * cluster	doc2vec	5/5
svm	max * uncertainty	doc2vec	5/5
svm	max * random	doc2vec	5/5
svm	cluster * uncertainty	doc2vec	5/5
svm	cluster * random	doc2vec	5/5
svm	cluster	doc2vec	5/10
svm	max	doc2vec	5/10
svm	max * cluster	doc2vec	5/10
svm	max * uncertainty	doc2vec	5/10
svm	max * random	doc2vec	5/10
svm	cluster * uncertainty	doc2vec	5/10
svm	cluster * random	doc2vec	5/10
svm	cluster	tfidf	10/10
svm	max	tfidf	10/10
svm	max * cluster	tfidf	10/10
svm	max * uncertainty	tfidf	10/10
svm	max * random	tfidf	10/10
svm	cluster * uncertainty	tfidf	10/10
svm	cluster * random	tfidf	10/10
svm	cluster	tfidf	5/5
svm	max	tfidf	5/5
svm	max * cluster	tfidf	5/5
svm	max * uncertainty	tfidf	5/5
svm	max * random	tfidf	5/5
svm	cluster * uncertainty	tfidf	5/5
svm	cluster * random	tfidf	5/5
svm	cluster	tfidf	5/10
svm	max	tfidf	5/10
svm	max * cluster	tfidf	5/10
svm	max * uncertainty	tfidf	5/10
svm	max * random	tfidf	5/10
svm	cluster * uncertainty	tfidf	5/10
svm	cluster * random	tfidf	5/10
svm	cluster	sbert	10/10
svm	max	sbert	10/10
svm	max * cluster	sbert	10/10
svm	max * uncertainty	sbert	10/10
svm	max * random	sbert	10/10
svm	cluster * uncertainty	sbert	10/10
svm	cluster * random	sbert	10/10
svm	cluster	sbert	5/5
svm	max	sbert	5/5
svm	max * cluster	sbert	5/5
svm	max * uncertainty	sbert	5/5

(continued)

Model	Query Strategy	Feature extraction strategy	Training data [included/excluded]
svm	max * random	sbert	5/5
svm	cluster * uncertainty	sbert	5/5
svm	cluster * random	sbert	5/5
svm	cluster	sbert	5/10
svm	max	sbert	5/10
svm	max * cluster	sbert	5/10
svm	max * uncertainty	sbert	5/10
svm	max * random	sbert	5/10
svm	cluster * uncertainty	sbert	5/10
svm	cluster * random	sbert	5/10
svm	cluster	embeddingIdf	10/10
svm	max	embeddingIdf	10/10
svm	max * cluster	embeddingIdf	10/10
svm	max * uncertainty	embeddingIdf	10/10
svm	max * random	embeddingIdf	10/10
svm	cluster * uncertainty	embeddingIdf	10/10
svm	cluster * random	embeddingIdf	10/10
svm	cluster	embeddingIdf	5/5
svm	max	embeddingIdf	5/5
svm	max * cluster	embeddingIdf	5/5
svm	max * uncertainty	embeddingIdf	5/5
svm	max * random	embeddingIdf	5/5
svm	cluster * uncertainty	embeddingIdf	5/5
svm	cluster * random	embeddingIdf	5/5
svm	cluster	embeddingIdf	5/10
svm	max	embeddingIdf	5/10
svm	max * cluster	embeddingIdf	5/10
svm	max * uncertainty	embeddingIdf	5/10
svm	max * random	embeddingIdf	5/10
svm	cluster * uncertainty	embeddingIdf	5/10
svm	cluster * random	embeddingIdf	5/10
lr	cluster	doc2vec	10/10
lr	max	doc2vec	10/10
lr	max * cluster	doc2vec	10/10
lr	max * uncertainty	doc2vec	10/10
lr	max * random	doc2vec	10/10
lr	cluster * uncertainty	doc2vec	10/10
lr	cluster * random	doc2vec	10/10
lr	cluster	doc2vec	5/5
lr	max	doc2vec	5/5
lr	max * cluster	doc2vec	5/5
lr	max * uncertainty	doc2vec	5/5
lr	max * random	doc2vec	5/5
lr	cluster * uncertainty	doc2vec	5/5
lr	cluster * random	doc2vec	5/5
lr	cluster	doc2vec	5/10
lr	max	doc2vec	5/10

Model	Query Strategy	Feature extraction strategy	Training data [included/excluded]
lr	max * cluster	doc2vec	5/10
lr	max * uncertainty	doc2vec	5/10
lr	max * random	doc2vec	5/10
lr	cluster * uncertainty	doc2vec	5/10
lr	cluster * random	doc2vec	5/10
lr	cluster	tfidf	10/10
lr	max	tfidf	10/10
lr	max * cluster	tfidf	10/10
lr	max * uncertainty	tfidf	10/10
lr	max * random	tfidf	10/10
lr	cluster * uncertainty	tfidf	10/10
lr	cluster * random	tfidf	10/10
lr	cluster	tfidf	5/5
lr	max	tfidf	5/5
lr	max * cluster	tfidf	5/5
lr	max * uncertainty	tfidf	5/5
lr	max * random	tfidf	5/5
lr	cluster * uncertainty	tfidf	5/5
lr	cluster * random	tfidf	5/5
lr	cluster	tfidf	5/10
lr	max	tfidf	5/10
lr	max * cluster	tfidf	5/10
lr	max * uncertainty	tfidf	5/10
lr	max * random	tfidf	5/10
lr	cluster * uncertainty	tfidf	5/10
lr	cluster * random	tfidf	5/10
lr	cluster	sbert	10/10
lr	max	sbert	10/10
lr	max * cluster	sbert	10/10
lr	max * uncertainty	sbert	10/10
lr	max * random	sbert	10/10
lr	cluster * uncertainty	sbert	10/10
lr	cluster * random	sbert	10/10
lr	cluster	sbert	5/5
lr	max	sbert	5/5
lr	max * cluster	sbert	5/5
lr	max * uncertainty	sbert	5/5
lr	max * random	sbert	5/5
lr	cluster * uncertainty	sbert	5/5
lr	cluster * random	sbert	5/5
lr	cluster	sbert	5/10
lr	max	sbert	5/10
lr	max * cluster	sbert	5/10
lr	max * uncertainty	sbert	5/10
lr	max * random	sbert	5/10
lr	cluster * uncertainty	sbert	5/10

(continued)

Model	Query Strategy	Feature extraction strategy	Training data [included/excluded]
lr	cluster * random	sbert	5/10
lr	cluster	embeddingIdf	10/10
lr	max	embeddingIdf	10/10
lr	max * cluster	embeddingIdf	10/10
lr	max * uncertainty	embeddingIdf	10/10
lr	max * random	embeddingIdf	10/10
lr	cluster * uncertainty	embeddingIdf	10/10
lr	cluster * random	embeddingIdf	10/10
lr	cluster	embeddingIdf	5/5
lr	max	embeddingIdf	5/5
lr	max * cluster	embeddingIdf	5/5
lr	max * uncertainty	embeddingIdf	5/5
lr	max * random	embeddingIdf	5/5
lr	cluster * uncertainty	embeddingIdf	5/5
lr	cluster * random	embeddingIdf	5/5
lr	cluster	embeddingIdf	5/10
lr	max	embeddingIdf	5/10
lr	max * cluster	embeddingIdf	5/10
lr	max * uncertainty	embeddingIdf	5/10
lr	max * random	embeddingIdf	5/10
lr	cluster * uncertainty	embeddingIdf	5/10
lr	cluster * random	embeddingIdf	5/10