Report ClayRS

Report ClayRS

This HTML document was generated from YAML files for the purpose

of replicability of experiments done with

ClayRS

.

It contains information about the dataset, preprocessing methods, analysis algorithms,

and the results of the experimental evaluation.

ClayRS configurations of the experiment

Dataset

In this experiment, the

Dataset

was used.

The statistics of the dataset used are reported in the following

Table of iteractions

:

The embendding techniques used during the processing of the document are the following:

Gensim glove-twitter-25

Sbert

Preprocessing

The preprocessing used is NLTK, a leading platform for building Python programs to work with human language data.

It provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet,

along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing,

and semantic reasoning.

In this experiment, those operations of NLTK were used:

Strip multiple whitespace

,

an operation that removes multiple whitspaces between words.

Stopwords removal

,

used to remove the stopwords occurring in the text.

spaCy is a free, open-source library for advanced Natural Language Processing (NLP) in Python.

spaCy is designed specifically for production use and helps you build applications that process

and "understand" large volumes of text. It can be used to build information extraction or natural

language understanding systems, or to pre-process text for deep learning.

In this experiment, those spaCy operations have been used:

Strip multiple whitespace

, an operation that removes multiple whitspaces between words.

Stopwords removal

, used to remove the stopwords occurring in the text.

Partitioning

The partitioning used is the Hold-Out Partitioning. This approach splits the dataset in use into a "train" set and a "test" set. The training set is what the model is trained on, and the test set is used to see how well the model will perform on new, unseen data.

The train set size of this experiment is the
80.0%
of the original dataset, while the test set is the remaining
20.0%.

The data has been shuffled before being splitted into batches.

Metrics

In ClayRS the Precision metric is calculated as such for the single user :

Precision u =

tp

u

/

(tp

u

+ fp

u

)

Where:

tp

u

:

is the number of items which are in the recommendation list of the user and have a

rating &ge

fp

u

:

is the number of items which are in the recommendation list of the user and have a

rating <

In ClayRS, Precision metric needs the following parameters:

the

relevant_threshold

, is a parameter needed to discern

relevant items and non-relevant items for every user. If not specified,

the mean rating score of every user will be used, in this experiment it

has been set to

None

.

sys_average

, a parameter that specifies how

the system average must be computed the default value is 'macro',

in this experiment the value of the sys_average is

macro

.

Precision at k

is the proportion of recommended

items in the top-k set that are relevant. The Precision@K metric is

calculated as such for the

single user

:

Precision@K

u

=

tp@K

u

/

(tp@K

u

+ fp@K

u

)

Where:

tp@K

u

:

is the number of items which are in the

recommendation list of the user and have a rating >

fp@K

u

:

is the number of items which are in the

recommendation list of the user and have a rating <

In this experiment the value

k is 2

,

the sys_average is

macro

Results

In the following table, we present the results of the evaluation

Table of the results

:

Table of the results