
Keras2Vec

Release 0.0.2

Oct 04, 2019

Contents:

1	Keras2Vec Module	1
2	Keras2Vec Data Generator	3
3	Keras2Vec Documents	5
4	Keras2Vec Encoder	7
5	Indices and tables	9
	Python Module Index	11
	Index	13

CHAPTER 1

Keras2Vec Module

```
class keras2vec.keras2vec.Keras2Vec (documents, embedding_size=16, seq_size=3,  
                                     neg_sampling=5, workers=1)
```

The Keras2Vec class is where the Doc2Vec model will be trained. By taking in a set of Documents it can begin to train against them to learn the embedding space that best represents the provided documents.

Args: documents (list of Document): List of documents to vectorize

```
build_model (infer=False)
```

Build both the training and inference models for Doc2Vec

```
fit (epochs, lr=0.1, verbose=0)
```

This function trains Keras2Vec with the provided documents

Args: epochs(int): How many times to iterate over the training dataset

```
get_doc_embedding (doc)
```

Get the vector/embedding for the provided doc Args:

doc (object): Object used in the initial generation of the model

Returns: np.array: embedding for the provided doc

```
get_doc_embeddings ()
```

Get the document vectors/embeddings from the trained model Returns:

np.array: Array of document embeddings indexed by encoded doc

```
get_label_embedding (label)
```

Get the vector/embedding for the provided label Args:

label (object): Object used in the initial generation of the model

Returns: np.array: embedding for the provided label

```
get_label_embeddings ()
```

Get the label vectors/embeddings from the trained model Returns:

np.array: Array of the label embeddings

get_word_embedding (*word*)

Get the vector/embedding for the provided word Args:

word (object): Object used in the initial generation of the model

Returns: np.array: embedding for the provided doc

get_word_embeddings ()

Get the vectors/embeddings from the trained model Returns:

np.array: Array of embeddings indexed by encoded doc

infer_vector (*infer_doc, epochs=5, lr=0.1, init_infer=True, verbose=0*)

Infer a documents vector by training the model against unseen labels and text. Currently inferred vector is passed to an attribute and not returned from this function.

Args: infer_doc (Document): Document for which we will infer a vector epochs (int): number of training cycles lr (float): the learning rate during inference init_infer (bool): determines whether or not we want to reinitialize weights for inference layer

Keras2Vec Data Generator

```
class keras2vec.data_generator.DataGenerator (documents, seq_size, neg_samples,  
                                             batch_size=100, shuffle=True,  
                                             val_gen=False)
```

The DataGenerator class is used to encode documents and generate training/testing data for a Keras2Vec instance. Currently this object is only used internally within the Keras2Vec class and not intended for direct use.

Args: documents (list of Document): List of documents to vectorize

build_vocabs ()

Build the vocabularies for the document ids, labels, and text of the provided documents

create_encodings ()

Build the encodings for each of the provided data types

encode_doc (*doc, neg_sampling=False, num_neg_samps=3*)

Encodes a document for the keras model

Args: doc(Document): The document to encode neg_sampling(Boolean): Whether or not to generate negative samples for the document **NOTE:** Currently not implemented

on_epoch_end ()

Updates indexes after each epoch

Keras2Vec Documents

class keras2vec.document.Document (*doc_id, text, labels=[]*)

The Document class is used to contain a documents content - document id, labels, text These objects are passed into the Keras2Vec class, which will process them for training

Args:

doc_id (int): The identification number for the document or collection of documents. While these should range from (1, num_docs), in theory this is not a hard constraint.

labels (list of str/int): a list of labels that contextualize the document. For example: a sports article might be labeled - ['news', 'sports'] **NOTE:** This is not fully implemented in the current version of Keras2Vec

text (str): the content of the document

gen_windows (*window_size, pad_word=""*)

Generate a sliding window, of size window_size, for the given document

Args: window_size (int): the size of the window, must be an odd number! pad_word (string): the word to pad indexes beyond the document, defaults to ""


```
class keras2vec.encoder.Encoder (items)
    Simple encoder class to fit/transform/reverse_transform data.

    Args: items (list of objects): items to encode.

    encode (items)
        Take in items to encode

        Args: items (list of objects)

    inverse_transform (index)
        Reverses the encoding for a given index

        Args: index (int): index to reverse encoding

        Returns: object: decoded object

    transform (item)
        Encodes a given object

        Args: item (object): Object to encode

        Returns: int: integer encoding of the item
```


CHAPTER 5

Indices and tables

- `genindex`
- `modindex`
- `search`

k

`keras2vec.data_generator`, 3
`keras2vec.document`, 5
`keras2vec.encoder`, 7
`keras2vec.keras2vec`, 1

B

`build_model()` (*keras2vec.keras2vec.Keras2Vec* *method*), 1
`build_vocabs()` (*keras2vec.data_generator.DataGenerator* *method*), 3

C

`create_encodings()` (*keras2vec.data_generator.DataGenerator* *method*), 3

D

`DataGenerator` (*class in keras2vec.data_generator*), 3
`Document` (*class in keras2vec.document*), 5

E

`encode()` (*keras2vec.encoder.Encoder* *method*), 7
`encode_doc()` (*keras2vec.data_generator.DataGenerator* *method*), 3
`Encoder` (*class in keras2vec.encoder*), 7

F

`fit()` (*keras2vec.keras2vec.Keras2Vec* *method*), 1

G

`gen_windows()` (*keras2vec.document.Document* *method*), 5
`get_doc_embedding()` (*keras2vec.keras2vec.Keras2Vec* *method*), 1
`get_doc_embeddings()` (*keras2vec.keras2vec.Keras2Vec* *method*), 1
`get_label_embedding()` (*keras2vec.keras2vec.Keras2Vec* *method*), 1

`get_label_embeddings()` (*keras2vec.keras2vec.Keras2Vec* *method*), 1
`get_word_embedding()` (*keras2vec.keras2vec.Keras2Vec* *method*), 2
`get_word_embeddings()` (*keras2vec.keras2vec.Keras2Vec* *method*), 2

I

`infer_vector()` (*keras2vec.keras2vec.Keras2Vec* *method*), 2
`inverse_transform()` (*keras2vec.encoder.Encoder* *method*), 7

K

`Keras2Vec` (*class in keras2vec.keras2vec*), 1
`keras2vec.data_generator` (*module*), 3
`keras2vec.document` (*module*), 5
`keras2vec.encoder` (*module*), 7
`keras2vec.keras2vec` (*module*), 1

O

`on_epoch_end()` (*keras2vec.data_generator.DataGenerator* *method*), 3

T

`transform()` (*keras2vec.encoder.Encoder* *method*), 7