

## tf.keras.preprocessing.sequence.skipgrams

```
skipgrams(  
    sequence,  
    vocabulary_size,  
    window_size=4,  
    negative_samples=1.0,  
    shuffle=True,  
    categorical=False,  
    sampling_table=None,  
    seed=None  
)
```

Defined in [tensorflow/python/keras/\\_impl/keras/preprocessing/sequence.py](#).

Generates skipgram word pairs.

Takes a sequence (list of indexes of words), returns couples of [word\_index, other\_word index] and labels (1s or 0s), where label = 1 if 'other\_word' belongs to the context of 'word', and label=0 if 'other\_word' is randomly sampled

### Arguments:

- **sequence**: a word sequence (sentence), encoded as a list of word indices (integers). If using a **sampling\_table**, word indices are expected to match the rank of the words in a reference dataset (e.g. 10 would encode the 10-th most frequently occurring token). Note that index 0 is expected to be a non-word and will be skipped.
- **vocabulary\_size**: int. maximum possible word index + 1
- **window\_size**: int. actually half-window. The window of a word  $w_i$  will be  $[i - \text{window\_size}, i + \text{window\_size} + 1]$
- **negative\_samples**: float  $\geq 0$ . 0 for no negative (=random) samples. 1 for same number as positive samples. etc.
- **shuffle**: whether to shuffle the word couples before returning them.
- **categorical**: bool. if False, labels will be integers (eg. [0, 1, 1 ..]), if True labels will be categorical eg. [[1,0],[0,1],[0,1] ..]
- **sampling\_table**: 1D array of size **vocabulary\_size** where the entry  $i$  encodes the probability to sample a word of rank  $i$ .
- **seed**: Random seed.

### Returns:

couples, labels: where `couples` are int pairs and  
`labels` are either 0 or 1.

### Note

By convention, index 0 in the vocabulary is  
a non-word and will be skipped.

the [Apache 2.0 License](#). For details, see our [Site Policies](#). Java is a registered trademark of Oracle and/or its affiliates.

Last updated November 2, 2017.

Stay Connected

- Blog
- GitHub
- Twitter

Support

- Issue Tracker
- Release Notes
- Stack Overflow

English

Terms | Privacy