Automatically generate headlines to short articles

This project attempts to reproduce the results in the paper:
[Generating News Headlines with Recurrent Neural Networks]
(http://arxiv.org/abs/1512.01712)

How to run
Software
* The code is running with [jupyter notebook](http://jupyter.org/)
* Install [Keras](http://keras.io/)
* `pip install python-Levenshtein`

Data
The dataset is the signal 1 million news article dataset,
each example is made from the text
from the start of the article, which we call description (or `desc`),
and the text of the original headline (or `head`).
The texts should be already tokenized and the tokens separated by
spaces.

Once you have the data ready save it in a python pickle file as a
tuple:
`(heads, descs, keywords)` were `heads` is a list of all the head
strings,
`descs` is a list of all the article strings in the same order and
length as `heads`.
I ignore the `keywords` information so you can place `None`.

Build a vocabulary of words
The vocabulatory notebook describes how a dictionary is built for the
tokens and how
an initial embedding matrix is built from [GloVe]
(http://nlp.stanford.edu/projects/glove/)

Train a model
Train notebook describes how a model is trained on the data using
[Keras](http://keras.io/)

Use model to generate new headlines
Predict generate headlines by the trained model and
showes the attention weights used to pick words from the description.
The text generation includes a feature which was
not described in the original paper, it allows for words that are
outside
the training vocabulary to be copied from the description to the
generated headline.