Word2Vec practice

April 10, 2023

```
[1]: import gensim
     import pandas as pd
```

```
[5]: df=pd.read_json(r'C:\Users\Rakesh\Downloads\reviews_Sports_and_Outdoors_5.
     ↪json\Sports_and_Outdoors_5.json', lines=True)
```

```
[6]: df
```

```
[6]:            reviewerID         asin  \
     0          AIXZKN4ACSKI  1881509818
     1          A1L5P841VIO02V 1881509818
     2          AB2W04NI4OEAD  1881509818
     3          A148SVSWKTJKU6 1881509818
     4          AAAWJ6LW9WMOO  1881509818
     ...               ...        ...
     296332  A2XX2A4OJCDNLZ  B00LFPS0CY
     296333  A3LGT6UZL99IW1  B00LFPS0CY
     296334   ASKZO80Z1RKTR  B00LFPS0CY
     296335   APRNS6DB68LLV  B00LFPS0CY
     296336  A3UJRNI8UR4871  B00LFPS0CY


                                            reviewerName helpful  \
     0                                      David Briner  [0, 0]
     1                                  Jason A. Kramer  [1, 1]
     2                                        J. Fernald  [2, 2]
     3                      Jusitn A. Watts "Maverick9614"  [0, 0]
     4                                      Material Man  [0, 0]
     ...                                           ...       ...
     296332                          RatherLiveInKeyWest  [2, 3]
     296333  Richard C. Drew "Anaal Nathra/Uthe vas Bethod…  [0, 0]
     296334                                      Robin Lee  [0, 0]
     296335             Rob Slaven "slavenrm@gmail. com"  [1, 1]
     296336                         Wulfstan "wulfstan"  [2, 2]


                                              reviewText  overall  \
     0       This came in on time and I am veru happy with …        5
     1       I had a factory Glock tool that I was using fo…        5
     2       If you don't have a 3/32 punch or would like t…        4
```

```
3       This works no better than any 3/32 punch you w…        4
4       I purchased this thinking maybe I need a speci…        4
…                                                      …        …
296332  This is a water bottle done right. It is a ver…        5
296333  If you're looking for an insulated water bottl…        5
296334  This Hydracentials Sporty 25 OZ, double insula…        5
296335  As usual I received this item free in exchange…        5
296336  Hydracentials insulated 25 oz water bottle.Thi…        5

                                             summary  unixReviewTime  \
0                                      Woks very good      1390694400
1                         Works as well as the factory tool      1328140800
2                              It's a punch, that's all.      1330387200
3                            It's a punch with a Glock logo.      1328400000
4                   Ok,tool does what a regular punch does.      1366675200
…                                                    …               …
296332  Hydracentials Sporty 25 Oz Stainless Steel Wat…      1405036800
296333          Large, incredibly well made water bottle!      1405641600
296334            "Great Water Bottle For Hot Day"…        1405900800
296335  A pretty impressive water bottle.  Best I've s…      1405900800
296336                                    "Cool" Bottle!      1405382400

          reviewTime
0        01 26, 2014
1         02 2, 2012
2        02 28, 2012
3         02 5, 2012
4        04 23, 2013
…                  …
296332   07 11, 2014
296333   07 18, 2014
296334   07 21, 2014
296335   07 21, 2014
296336   07 15, 2014

[296337 rows x 9 columns]
```

[7]: `df.shape`

[7]: (296337, 9)

[8]: `df['reviewText']`

```
[8]: 0        This came in on time and I am veru happy with …
     1        I had a factory Glock tool that I was using fo…
     2        If you don't have a 3/32 punch or would like t…
     3        This works no better than any 3/32 punch you w…
```

```
4          I purchased this thinking maybe I need a speci…
                                    …
296332     This is a water bottle done right. It is a ver…
296333     If you're looking for an insulated water bottl…
296334     This Hydracentials Sporty 25 OZ, double insula…
296335     As usual I received this item free in exchange…
296336     Hydracentials insulated 25 oz water bottle.Thi…
Name: reviewText, Length: 296337, dtype: object
```

[9]: `review_text=df.reviewText.apply(gensim.utils.simple_preprocess)`

[10]: `review_text`

[10]:
```
0          [this, came, in, on, time, and, am, veru, happ…
1          [had, factory, glock, tool, that, was, using, …
2          [if, you, don, have, punch, or, would, like, t…
3          [this, works, no, better, than, any, punch, yo…
4          [purchased, this, thinking, maybe, need, speci…
                                    …
296332     [this, is, water, bottle, done, right, it, is,…
296333     [if, you, re, looking, for, an, insulated, wat…
296334     [this, hydracentials, sporty, oz, double, insu…
296335     [as, usual, received, this, item, free, in, ex…
296336     [hydracentials, insulated, oz, water, bottle, …
Name: reviewText, Length: 296337, dtype: object
```

[12]:
```python
model=gensim.models.Word2Vec(
    window=10,
    min_count=2,
    workers=4,
)
```

[13]: `model.build_vocab(review_text,progress_per=1000)`

[14]: `model.train(review_text, total_examples=model.corpus_count, epochs=model.epochs)`

[14]: `(91341859, 121496535)`

[15]: `model.save("./word2vec-sports-review")`

[16]: `model.wv.most_similar('thinking')`

[16]:
```
[('worried', 0.7020843029022217),
 ('leary', 0.6579857468605042),
 ('curious', 0.6427538394927979),
 ('concerned', 0.6259337663650513),
 ('leery', 0.6248724460601807),
 ('bitching', 0.6051909923553467),
```

```
('apprehensive', 0.5941770672798157),
('wary', 0.5940709710121155),
('complaining', 0.5918372273445129),
('skeptical', 0.5849025249481201)]
```

[17]: `model.wv.similarity(w1='thinking',w2='concerned')`

[17]: 0.6259338

[ ]:

[ ]: