Q5

The words I chosen are: jack, students, lovely, fly, California, vehicles

|  |  |  |
| --- | --- | --- |
| Word | Top-10 similar words | Explanation |
| jack | 1: sam (0.804650868501)  2: jim (0.784263389501)  3: adam (0.777474342932)  4: ed (0.775161593077)  5: chris (0.770623148763)  6: anthony (0.759454281466)  7: bruce (0.748197814965)  8: brian (0.745924331721)  9: steve (0.744650198971)  10: ray (0.744424710639) | The list for similarity words of “jack”. They may used in similar position of contexts. |
| students | 1: teachers (0.776016213554)  2: classes (0.71812726122)  3: schools (0.700161455767)  4: parents (0.674142815621)  5: teenagers (0.672505183034)  6: children (0.664030015393)  7: people (0.662068342066)  8: families (0.661573812503)  9: adults (0.651930970632)  10: kids (0.631532216187) | The list for similarity words of “students”. They may used in similar position of contexts. |
| lovely | 1: gorgeous (0.838816307984)  2: beautiful (0.831308690299)  3: fantastic (0.772205114374)  4: sexy (0.762152422243)  5: wonderful (0.751102054375)  6: bright (0.717905617737)  7: sweet (0.716062645517)  8: strange (0.712895931406)  9: handsome (0.709523083931)  10: cute (0.692549223189) | The list for similarity words of “lovely”. They may used in similar position of contexts. |
| fly | 1: throw (0.746638393515)  2: shoot (0.72376486247)  3: walk (0.712950055472)  4: ride (0.699915822465)  5: catch (0.690512352079)  6: hop (0.686922695181)  7: kick (0.686550473883)  8: go (0.669932079829)  9: pull (0.659382655827)  10: stick (0.653277848617) | The list for similarity words of “fly”. They may used in similar position of contexts. |
| California | 1: connecticut (0.775855966702)  2: florida (0.732917165034)  3: pennsylvania (0.725749564901)  4: texas (0.725168724457)  5: massachusetts (0.715188050659)  6: ohio (0.698119898535)  7: virginia (0.684975746318)  8: southern (0.587589066007)  9: westchester (0.551599496908)  10: state (0.543420680191) | The list for similarity words of “california”. They may used in similar position of contexts. |
| vehicles | 1: cars (0.852364641925)  2: units (0.765972728127)  3: equipment (0.731489939507)  4: vehicle (0.697273171078)  5: passengers (0.68085937059)  6: drivers (0.68058587509)  7: jobs (0.671838220246)  8: plants (0.639343052376)  9: stations (0.635773815114)  10: businesses (0.634113707873) | The list for similarity words of “vehicles”. They may used in similar position of contexts. |

According to the sheet I can see that the similarity between pairs are same irrespective of orders.

Screenshot as follow:

jack:   
1: sam (0.804650868501)2: jim (0.784263389501)3: adam (0.777474342932)4: ed (0.775161593077)5: chris (0.770623148763)6: anthony (0.759454281466)7: bruce (0.748197814965)8: brian (0.745924331721)9: steve (0.744650198971)10: ray (0.744424710639)  
students:   
1: teachers (0.776016213554)2: classes (0.71812726122)3: schools (0.700161455767)4: parents (0.674142815621)5: teenagers (0.672505183034)6: children (0.664030015393)7: people (0.662068342066)8: families (0.661573812503)9: adults (0.651930970632)10: kids (0.631532216187)  
lovely:   
1: gorgeous (0.838816307984)2: beautiful (0.831308690299)3: fantastic (0.772205114374)4: sexy (0.762152422243)5: wonderful (0.751102054375)6: bright (0.717905617737)7: sweet (0.716062645517)8: strange (0.712895931406)9: handsome (0.709523083931)10: cute (0.692549223189)  
fly:   
1: throw (0.746638393515)2: shoot (0.72376486247)3: walk (0.712950055472)4: ride (0.699915822465)5: catch (0.690512352079)6: hop (0.686922695181)7: kick (0.686550473883)8: go (0.669932079829)9: pull (0.659382655827)10: stick (0.653277848617)  
california:   
1: connecticut (0.775855966702)2: florida (0.732917165034)3: pennsylvania (0.725749564901)4: texas (0.725168724457)5: massachusetts (0.715188050659)6: ohio (0.698119898535)7: virginia (0.684975746318)8: southern (0.587589066007)9: westchester (0.551599496908)10: state (0.543420680191)  
vehicles:   
1: cars (0.852364641925)2: units (0.765972728127)3: equipment (0.731489939507)4: vehicle (0.697273171078)5: passengers (0.68085937059)6: drivers (0.68058587509)7: jobs (0.671838220246)8: plants (0.639343052376)9: stations (0.635773815114)10: businesses (0.634113707873)

For further analysis, I choose the first word ‘jack’to analyze its most-similar list. Do analyze to ‘sam’and ‘jim’, the results as follow:

|  |  |  |
| --- | --- | --- |
| Word | Top-10 similar words | Explanation |
| sam | 1: adam (0.82784158777)  2: max (0.809248078918)  3: jim (0.808969699176)  4: jack (0.804650868501)  5: eric (0.794762293329)  6: justin (0.791423227251)  7: jonathan (0.784367445232)  8: chris (0.778453303452)  9: ben (0.771410333419)  10: harry (0.770777400036) | The list for similarity words of “adam”. They may used in similar position of contexts. |
| jim | 1: steve (0.911116937863)  2: brian (0.898181077512)  3: jeff (0.895135638932)  4: kevin (0.894086516126)  5: chris (0.889899246767)  6: ed (0.884879643028)  7: tom (0.876603180981)  8: tim (0.870422636008)  9: dan (0.867952469503)  10: adam (0.852130097542) | The list for similarity words of “james”. They may used in similar position of contexts. |

According to this sheet it can be observed that pairs similarity (‘jack’, ‘adam’) = (‘adam’, jack’) . The list of ‘jack’, ‘james’ and ‘adam’has some different words and probabilities.

sam:   
1: adam (0.82784158777)2: max (0.809248078918)3: jim (0.808969699176)4: jack (0.804650868501)5: eric (0.794762293329)6: justin (0.791423227251)7: jonathan (0.784367445232)8: chris (0.778453303452)9: ben (0.771410333419)10: harry (0.770777400036)  
jim:   
1: steve (0.911116937863)2: brian (0.898181077512)3: jeff (0.895135638932)4: kevin (0.894086516126)5: chris (0.889899246767)6: ed (0.884879643028)7: tom (0.876603180981)8: tim (0.870422636008)9: dan (0.867952469503)10: adam (0.852130097542)

Compared to this method and the sparse method of Q3, this method is better.

First reason is this method used dense numpy vectors by applying words embedding so can run the data more quickly. What’s more this method gets lower probabilities in general.