Multimedia Computing & Application | CSE 563

Assignment 3 | Divyanshu Kumar Singh (2017048)

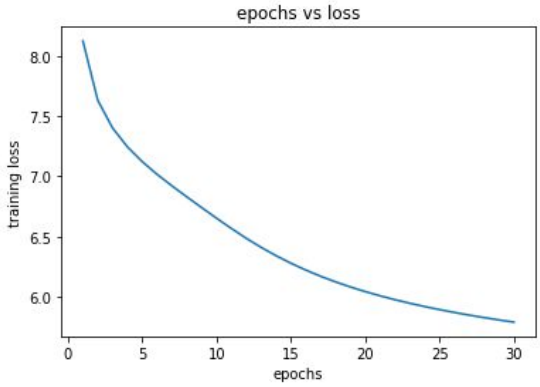
Ques 1.)

Technically, Word2Vec is a simple and shallow 2 layer Neural Network model. It takes input a corpus of text/sentence/words etc and outputs a vectors space i.e called a word vector. One thing to note here is, that word vector when made from a corpus, we have to make sure the text is clean, therefore we perform preprocessing on the corpus. We have two remove all the stopwords, numbers, special character and any symbols and etc. Finally, convert all the words to lower case.

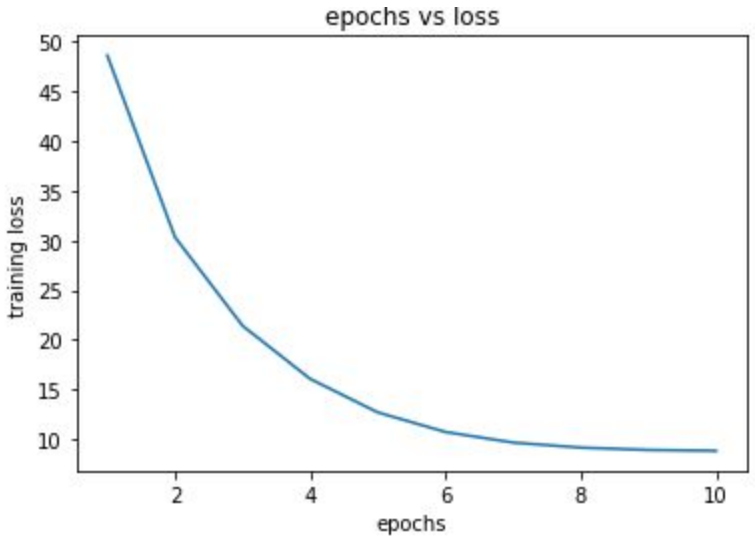
Before putting this dataset, into input space we make a dictionary/vocabulary for the words which keep a record of every unique word in dataset. We cannot input a string in the model, hence we use a one-hot encoded vector. During the training what it does is it updates the weights of vectors to minimize the loss. Now, coming to the layer, the hidden layer is the fully connected layer and its weights are none other than the word embeddings. The output layer is the probabilities of the desired word in correspondence to the vocabulary or the dictionary.

During the training we check the loss of every epoch, and like every other ML model, we expect the loss to decrease the epoch increase here are some screenshot for algorithms, for different sizes of data and epochs.

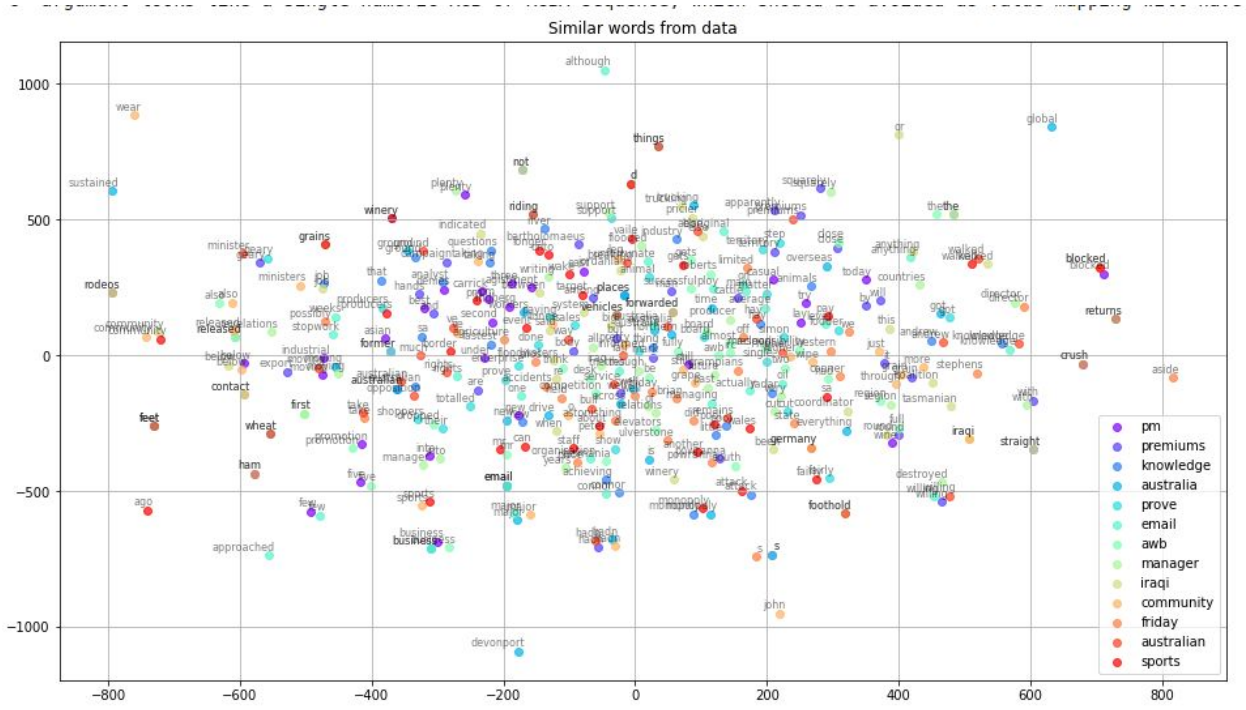
a.) This image corresponds to 100 samples from entire data i.e 100 sentences, and 30 epochs



b.) This image corresponds to 50 samples and 10 epochs



Visualizing Word Embedding using tSNE:



For different word_embedding during i.e formation of those vectors please refer to the code file: Quesition1.ipynb (gtihub: https://github.com/sdiv0/Multimedia-Computing-Application/tree/master/Assignment%203)

Ques 2.)

Approach	Baseline	Relevance Feedback	Relevance Feedback Query Expansion
MAP	**0.518385**	**0.89649**	**0.91355**

Based on the explanation of the retrieval system it was expected that both the relevance feedback approach would run much better than the baseline, it was mainly due to the reason that in each and every relevance feedback approach we consider the user feedback, which actually helps the algorithms filter out the useless result.

In this case, since, we already have the ground truth so we use it instead of actual feedback from the user. We can easily classify the relevant and non-relevant documents from the retrieved results.

Relevance feedback actually increases the query weights for relevant doc by a factor of alpha and decrease the query weight of non-relevant doc by a factor of beta.

Relevance feedback using query expansion goes one step further, for every doc in tfidf values, we sort them and find the top ten values (i.e in descending order). Along with their indices, we do this every doc in for each query. Now we update the specific location of those indices in the each query basically increasing their weight. This produce finer results but this is hardly comparable to relevance feedback approach as can seen in the above table.

```
Baseline Retrieval
Query: 1 Top relevant 10 documents: [ 72 500 965 181 360 171  15 166 513 511]
Query: 2 Top relevant 10 documents: [258 712 289 162 237 187  96 299 713 974]
Query: 3 Top relevant 10 documents: [ 70 160  62 286 230  71 234 276 277  59]
Query: 4 Top relevant 10 documents: [ 67 532 175 787 209 234 177  93 400 405]
Query: 5 Top relevant 10 documents: [308   8 327 333 330 326 329   5  10   7]
Query: 6 Top relevant 10 documents: [116 321 112 238 311 115 323 118 260 310]
Query: 7 Top relevant 10 documents: [121 261 189  92 391 247  82 387 415 392]
Query: 8 Top relevant 10 documents: [ 52 433  60 434 265 435 431 427 262 264]
Query: 9 Top relevant 10 documents: [409 415 273 194 413 114 268 421 954 840]
Query: 10 Top relevant 10 documents: [532 534 543  77 465 981 556 254 367 539]
Query: 11 Top relevant 10 documents: [ 445  990  446  228  227 1006  532  453  463  229]
Query: 12 Top relevant 10 documents: [ 19 367 368 373 193 371 220 365 936  20]
Query: 13 Top relevant 10 documents: [197 196  21 194 199 198 195 483 481 147]
Query: 14 Top relevant 10 documents: [  25   26   23  457   29  861  454 1006  445  371]
Query: 15 Top relevant 10 documents: [218 355 361 103 219 106 350 107 357 353]
Query: 16 Top relevant 10 documents: [ 36 202  99 494 484 497 492 205 812 496]
Query: 17 Top relevant 10 documents: [ 38 134 336  41 304 128 528 342 127 341]
Query: 18 Top relevant 10 documents: [ 45 525  49 134 528  81 514 517  48 516]
Query: 19 Top relevant 10 documents: [847 844 551 861 863 864 865 564 673 555]
Query: 20 Top relevant 10 documents: [177 860 435 434 431 582  60  52 262 878]
Query: 21 Top relevant 10 documents: [821 613 815 766 896 810 253 887 889 888]
Query: 22 Top relevant 10 documents: [648 108 375 647 645 905  48 758 371 636]
Query: 23 Top relevant 10 documents: [916 917 804 798 849 819 620 817 813 808]
Query: 24 Top relevant 10 documents: [674 667 939 940 936 852 938 850 675 686]
Query: 25 Top relevant 10 documents: [696 687 699 698 948 692 690 695 587 949]
Query: 26 Top relevant 10 documents: [716 723 967 966 971 970 703 973 974 710]
Query: 27 Top relevant 10 documents: [734 984 732 978 727 977 976 974 980 731]
Query: 28 Top relevant 10 documents: [777 774 772 995 989 998 776 840 994 770]
Query: 29 Top relevant 10 documents: [1016 1008 1017 1015  853 1012  740 1007 1009  750]
Query: 30 Top relevant 10 documents: [1027 1026 1023 1024 1020 1019 1033  920  867  841]
Query: 1 AveP: 0.7202185962386813
Query: 2 AveP: 0.421921563277281
Query: 3 AveP: 0.5558513746821222
Query: 4 AveP: 0.3401614027216174
Query: 5 AveP: 0.7258666056836868
Query: 6 AveP: 0.7705628746904544
Query: 7 AveP: 0.6089939128444843
Query: 8 AveP: 0.4834395577799805
Query: 9 AveP: 0.4547386952031073
Query: 10 AveP: 0.24465405738440002
Query: 11 AveP: 0.49684946239628464
Query: 12 AveP: 0.507463162086818
Query: 13 AveP: 0.8769455523884369
Query: 14 AveP: 0.6288375786617474
Query: 15 AveP: 0.48493997193350696
Query: 16 AveP: 0.5991679632592016
Query: 17 AveP: 0.25604069622416237
Query: 18 AveP: 0.42262203950986427
Query: 19 AveP: 0.44834522737921056
Query: 20 AveP: 0.20492089261229138
Query: 21 AveP: 0.25590108380895416
Query: 22 AveP: 0.27399296436255677
Query: 23 AveP: 0.46491898167016327
Query: 24 AveP: 0.7978831868602536
Query: 25 AveP: 0.7236796830654314
Query: 26 AveP: 0.4365812284246108
Query: 27 AveP: 0.5671081951580693
Query: 28 AveP: 0.5615817175885622
Query: 29 AveP: 0.7961049828931106
Query: 30 AveP: 0.4212839117806344
MAP: 0.5183859040856561
```

```
Retrieval with Relevance Feedback
Query: 1 Top relevant 10 documents: [ 13 171 511 181  15 166 509 184 500 504]
Query: 2 Top relevant 10 documents: [289 187 292 237 258 162 296 301  80  96]
Query: 3 Top relevant 10 documents: [232 277 160 230 282  73  71 283  70  78]
Query: 4 Top relevant 10 documents: [207 396  94 209  93 178 400 397 210 403]
Query: 5 Top relevant 10 documents: [327   5 330 333   8 159  10 308 329   6]
Query: 6 Top relevant 10 documents: [116 321 238 323 112 260 242 115 122 118]
Query: 7 Top relevant 10 documents: [261 189 121  92 247 391 392 385 393 388]
Query: 8 Top relevant 10 documents: [265  52 433  60 123 264 266  61 262 435]
Query: 9 Top relevant 10 documents: [126 409 422 273  64 421 124 271  56 268]
Query: 10 Top relevant 10 documents: [540 256 254  58 535 532 155 543 529 539]
Query: 11 Top relevant 10 documents: [229 441 446 448 447 148  32  63 151 444]
Query: 12 Top relevant 10 documents: [193 365  19  17  20  16 367 373 364 368]
Query: 13 Top relevant 10 documents: [144 481 199 474 479 477 471 145 478 198]
Query: 14 Top relevant 10 documents: [ 23  26 457 455 459 461 454 458  25 467]
Query: 15 Top relevant 10 documents: [219 355 353 359  33 358 357 140 102 361]
Query: 16 Top relevant 10 documents: [ 36 202 484  99 205 492 490 494 487  98]
Query: 17 Top relevant 10 documents: [ 38  41 342 336 334 128 341 337 339 348]
Query: 18 Top relevant 10 documents: [525 517 514 528  45 526 523  49 522  43]
Query: 19 Top relevant 10 documents: [847 844 865 861 564 562 858 555 545 551]
Query: 20 Top relevant 10 documents: [874 577 581 573 878 590 873 580 589 570]
Query: 21 Top relevant 10 documents: [821 815 613 894 892 888 887 898 619 896]
Query: 22 Top relevant 10 documents: [652 654 640 648 659 641 908 660 913 900]
Query: 23 Top relevant 10 documents: [811 922 822 815 918 924 808 810 927 914]
Query: 24 Top relevant 10 documents: [668 942 667 686 666 670 851 850 852 929]
Query: 25 Top relevant 10 documents: [690 699 692 696 688 687 955 951 953 952]
Query: 26 Top relevant 10 documents: [723 716 967 960 959 717 712 973 724 971]
Query: 27 Top relevant 10 documents: [738 736 739 728 734 735 977 984 737 732]
Query: 28 Top relevant 10 documents: [ 995  989  776 1000  777  774  992  994 1001  998]
Query: 29 Top relevant 10 documents: [ 853 1012 1017  740  741 1008  744 1015  768  742]
Query: 30 Top relevant 10 documents: [1027  843 1021 1033  825 1020  827 1024  823  831]
Query: 1 AveP: 0.9506896286600243
Query: 2 AveP: 0.9172576657411218
Query: 3 AveP: 0.8959869441152865
Query: 4 AveP: 0.7750238253950321
Query: 5 AveP: 0.8990161326932629
Query: 6 AveP: 0.9184012066365006
Query: 7 AveP: 0.9616785175608705
Query: 8 AveP: 0.7644144957060557
Query: 9 AveP: 0.9415327104736922
Query: 10 AveP: 0.9304930748697764
Query: 11 AveP: 0.8842958611436873
Query: 12 AveP: 0.9534662867996201
Query: 13 AveP: 0.9520263953534629
Query: 14 AveP: 0.9289779848419555
Query: 15 AveP: 0.9367975320750203
Query: 16 AveP: 0.8516453825277354
Query: 17 AveP: 0.8644689432317071
Query: 18 AveP: 0.8158444823150705
Query: 19 AveP: 0.8549181772826574
Query: 20 AveP: 0.9077840748137772
Query: 21 AveP: 0.7075908653452404
Query: 22 AveP: 0.8255244459514591
Query: 23 AveP: 0.9146206533389989
Query: 24 AveP: 0.9762461701368962
Query: 25 AveP: 0.8930608719963969
Query: 26 AveP: 0.8908489479046141
Query: 27 AveP: 0.9452531582433544
Query: 28 AveP: 0.9452322288601832
Query: 29 AveP: 0.9657010300370151
Query: 30 AveP: 0.9260412860796493
MAP: 0.8964946326710042
```

```
Query: 1 Top relevant 10 documents: [ 13 511 171 181 499 509 166  15 180 184]
Query: 2 Top relevant 10 documents: [289 187 237 292 258 162 296 301 236  80]
Query: 3 Top relevant 10 documents: [277 232 230 160 283  71  73 276  70 282]
Query: 4 Top relevant 10 documents: [207  94 178 400 396  93 209 397 403 259]
Query: 5 Top relevant 10 documents: [327   5   6 330 329 333  10 326 159   8]
Query: 6 Top relevant 10 documents: [116 321 238 112 323 260 115 122 242 243]
Query: 7 Top relevant 10 documents: [261 189  92 121 391 247 392 393 385 386]
Query: 8 Top relevant 10 documents: [265 433 264 266  52 123 435  61 429 262]
Query: 9 Top relevant 10 documents: [126 409 422 273  64  56 124 268 421 412]
Query: 10 Top relevant 10 documents: [256 254 540 533  58 535 155 543 531 532]
Query: 11 Top relevant 10 documents: [229 441 448 444 446 228 452 148 447  32]
Query: 12 Top relevant 10 documents: [ 19 193 367 365  20 364  17 366 373 371]
Query: 13 Top relevant 10 documents: [199 481 144 198 477 194 479 474 471 145]
Query: 14 Top relevant 10 documents: [ 26  23 457 455 459 461  25 456  28 458]
Query: 15 Top relevant 10 documents: [359 219 353 355 102  33 140 357 350 358]
Query: 16 Top relevant 10 documents: [ 36 202 205  99 484  98 492 490 494 487]
Query: 17 Top relevant 10 documents: [ 38  41 342 336 341 128  37 334 340 133]
Query: 18 Top relevant 10 documents: [525 514 517  45 528 516 515 526  49 524]
Query: 19 Top relevant 10 documents: [844 847 865 564 861 562 555 858 545 860]
Query: 20 Top relevant 10 documents: [581 874 577 878 573 590 580 873 585 570]
Query: 21 Top relevant 10 documents: [821 888 613 815 896 619 898 894 887 892]
Query: 22 Top relevant 10 documents: [652 648 654 640 635 659 913 908 636 644]
Query: 23 Top relevant 10 documents: [811 922 915 815 812 808 822 918 818 924]
Query: 24 Top relevant 10 documents: [668 666 942 851 674 686 667 670 935 850]
Query: 25 Top relevant 10 documents: [699 690 947 688 692 687 696 698 695 955]
Query: 26 Top relevant 10 documents: [973 967 966 960 723 716 959 717 724 971]
Query: 27 Top relevant 10 documents: [738 739 736 728 977 730 734 735 986 984]
Query: 28 Top relevant 10 documents: [ 995  989  776  994  777  992 1000  774  771  998]
Query: 29 Top relevant 10 documents: [ 853 1017 1012 1008 1015  740  742  741 1007 1013]
Query: 30 Top relevant 10 documents: [1027 1033 1026  843 1021 1020  825  827 1024  823]
Query: 1 AveP: 0.9476220318424274
Query: 2 AveP: 0.9485405219780221
Query: 3 AveP: 0.893459050565721
Query: 4 AveP: 0.8646553049283697
Query: 5 AveP: 0.9340573997944351
Query: 6 AveP: 0.9226747109100049
Query: 7 AveP: 0.9723622782446313
Query: 8 AveP: 0.6955663587749149
Query: 9 AveP: 0.9532914055027997
Query: 10 AveP: 0.9396290835345901
Query: 11 AveP: 0.9126935936793509
Query: 12 AveP: 0.9797979797979798
Query: 13 AveP: 0.9783650323499946
Query: 14 AveP: 0.9538589372229078
Query: 15 AveP: 0.966261400616328
Query: 16 AveP: 0.962730858884705
Query: 17 AveP: 0.868162713229825
Query: 18 AveP: 0.8231254420960301
Query: 19 AveP: 0.8869308664502955
Query: 20 AveP: 0.9010898497611524
Query: 21 AveP: 0.7437637376182714
Query: 22 AveP: 0.8133366062487805
Query: 23 AveP: 0.9277902354115005
Query: 24 AveP: 0.9654222904592618
Query: 25 AveP: 0.9180354716631118
Query: 26 AveP: 0.9236525255024935
Query: 27 AveP: 0.9371603251505213
Query: 28 AveP: 0.9306343614690497
Query: 29 AveP: 0.9561222496541314
Query: 30 AveP: 0.985969387755102
MAP: 0.9135587337032237
```