Project 1: Sentiment Classification with Deep Learning

Mingxiao Ye

Abstract

Sentiment analysis or opinion mining refers to the task of classifying texts based on their affective states i.e. whether the expressed opinion in text is a positive, negative or neutral one. Some more advanced opinion mining tasks classify texts on the basis of mood (e.g. happy, angry, sad, etc.), expression (e.g. sarcastic, didactic, philosophical, etc.) or intention (e.g. question, complaint, request, etc.)

For this project, we will consider a very simple binary classification problem that identifies whether the sentiment associated with a chunk of text is positive or negative.

In this report, I will use the simple dataset abstracted from IMDB Movie review that contains 40 training and 40 testing examples. The model will be built/trained on the training examples and tested against the test examples.

1. Introduction

In this project, I will try to implement three different deep learning models:

1. embedding layer + fully connected (FCN) layer
2. embedding layer + Vanilla RNN layer + FCN layer
3. embedding layer + self-attention layer + FCN layer

for performing sentiment classification of movie reviews.

2. Load the data

2.1 read file

To take in a tar file and produces two lists, one containing the movie reviews and another containing classes. For example:

[u'the happy bastard\'s quick movie review \ndamn that y2k bug . \n\'s got a head start in this movie starring jamie lee curtis and another baldwin brother (william this time) in a story regarding a crew of a tugboat that comes across a deserted russian tech ship that has a strangeness to it when they kick the power back on . \nlittle do they know the power within . . . \ngoing for the gore and bringing on a few action sequences here and there , virus still feels very empty , like a movie going for all flash and no substance . \nwe don\'t know why the crew was really out in the middle of nowhere , we don\'t know the origin of what took over the ship (just that a big pink flashy thing hit the mir) , and , of course , we don\'t know why donald sutherland is stumbling around drunkenly throughout . \nhere , it\'s just " hey , let\'s chase these people around with some robots " . \nthe acting is below average , even from the likes of curtis . \nyou\'re more likely to get a kick out of her work in halloween h20 . \nsutherland is wasted and baldwin , well , he\'s acting like a baldwin , of course . \nthe real star here are stan winston\'s robot design , some schnazzy cgi , and the occasional good gore shot , like picking into someone\'s brain . \nso , if robots and body parts really turn you on , here\'s your movie . \notherwise , it\'s pretty much a sunken ship of a movie . ', u'plot : two teen couples go to a church party , drink and then drive . \nthey get into an accident . \none of the guys dies , but his girlfriend continues to see him in her life , and has nightmares . \nwhat\'s the deal ? \nwatch the movie and " sorta " find out . . \ncritique : a mind-fuck movie for the teen generation that touches on a very cool idea , but presents it in a very bad package . \nwhich is what makes this review an even harder one to write , since i generally applaud films which attempt to break the mold , mess with your head and such (lost highway & memento) , but there are good and bad ways of making all types of films , and these folks just didn\'t snag this one correctly . \nthey seem to have taken this pretty neat concept , but executed it terribly . \nso what are the problems with the movie ? \nwell , its main problem is that it\'s simply too jumbled . \nit starts off " normal " but then downshifts into this " fantasy " world in which you , as an audience member , have no idea what\'s going on . \nthere are dreams , there are characters coming back from the dead , there are others who look like the dead , there are strange apparitions , there are disappearances , there are a loooooot of chase scenes , there are tons of weird things that happen , and most of it is simply not explained . \nnow i personally don\'t mind trying to unravel a film every now and then , but when all it does is give me the same clue over and over again , i get kind of fed up after a while , which is this film\'s biggest problem . \nit\'s obviously got this big secret to hide , but it seems to want to hide it completely until its final five minutes . \nand do they make things entertaining , thrilling or even engaging , in the meantime ? \nnot really . \nthe sad part is that the arrow and i both dig on flicks like this , so we actually figured most of it out by the half-way point , so all of the strangeness after that did start to make a little bit of sense , but it still didn\'t the make the film all that more entertaining . \ni guess the bottom line with movies like this is that you should always make sure that the audience is " into it " even before they are given the secret password to enter your world of understanding . \ni mean , showing melissa sagemiller running away from visions for about 20 minuses throughout the movie is just plain lazy ! ! \nokay , we get it . . . there \nare people chasing her and we don\'t know who they are . \ndo we really need to see it over and over again ? \nhow about giving us different scenes offering further insight into all of the strangeness going down in the movie ? \napparently , the studio took this film away from its director and chopped it up themselves , and it shows . \nthere might \'ve been a pretty decent teen mind-fuck movie in here somewhere , but i guess " the suits " decided that turning it into a music video with little edge , would make more sense . \nthe actors are pretty good for the most part , although wes bentley just seemed to be playing the exact same character that he did in american beauty , only in a new neighborhood . \nbut my biggest kudos go out to sagemiller , who holds her own throughout the entire film , and actually has you feeling her character\'s unraveling . \noverall , the film doesn\'t stick because it doesn\'t entertain , it\'s confusing , it rarely excites and it feels pretty redundant for most of its runtime , despite a pretty cool ending and explanation to all of the craziness that came before it . \noh , and by the way , this is not a horror or teen slasher flick . . . it\'s \njust packaged to look that way because someone is apparently assuming that the genre is still hot with the kids . \nit also wrapped production two years ago and has been sitting on the shelves ever since . \nwhatever . . . skip \nit ! \nwhere\'s joblo coming from ? \na nightmare of elm street 3 (7/10) - blair witch 2 (7/10) - the crow (9/10) - the crow : salvation (4/10) - lost highway (10/10) - memento (10/10) - the others (9/10) - stir of echoes (8/10)', u'films adapted from comic books have had plenty of success , whether they\'re about superheroes (batman, superman , spawn) , or geared toward kids (casper) or the arthouse crowd (ghost world) , but there\'s never really been a comic book like from hell before . \nfor starters , it was created by alan moore (and eddie campbell) , who brought the medium to a whole new level in the mid \'80s with a 12-part series called the watchmen . \nto say moore and campbell thoroughly researched the subject of jack the ripper would be like saying michael jackson is starting to look a little odd . \nthe book (or " graphic novel , " if you will) is over 500 pages long and includes nearly 30 more that consist of nothing but footnotes . \nin other words , don\'t dismiss this film because of its source . \nif you can get past the whole comic book thing , you might find another stumbling block in from hell\'s directors , albert and allen hughes . \ngetting the hughes brothers to direct this seems almost as ludicrous as casting carrot top in , well , anything , but riddle me this : who better to direct a film that\'s set in the ghetto and features really violent street crime than the mad geniuses behind menace ii society ? \nthe ghetto in question is , of course , whitechapel in 1888 london\'s east end . \nit\'s a filthy , sooty place where the whores (called " unfortunates ") are starting to get a little nervous about this mysterious psychopath who has been carving through their profession with surgical precision . \nwhen the first stiff turns up , copper peter godley (robbie coltrane , the world is not enough) calls in inspector frederick abberline (johnny depp , blow) to crack the case . \nabberline , a widower , has prophetic dreams he unsuccessfully tries to quell with copious amounts of absinthe and opium . \nupon arriving in whitechapel , he befriends an unfortunate named mary kelly (heather graham , say it isn\'t so) and proceeds to investigate the horribly gruesome crimes that even the police surgeon can\'t stomach . \ni don\'t think anyone needs to be briefed on jack the ripper , so i won\'t go into the particulars here , other than to say moore and campbell have a unique and interesting theory about both the identity of the killer and the reasons he chooses to slay . \nin the comic , they don\'t bother cloaking the identity of the ripper , but screenwriters terry hayes (vertical limit) and rafael yglesias (les mis ? rables) do a good job of keeping him hidden from viewers until the very end . \nit\'s funny to watch the locals blindly point the finger of blame at jews and indians because , after all , an englishman could never be capable of committing such ghastly acts . \nand from hell\'s ending had me whistling the stonecutters song from the simpsons for days (" who holds back the electric car/who made steve guttenberg a star ? ") . \ndon\'t worry - it\'ll all make sense when you see it . \nnow onto from hell\'s appearance : it\'s certainly dark and bleak enough , and it\'s surprising to see how much more it looks like a tim burton film than planet of the apes did (at times , it seems like sleepy hollow 2) . \nthe print i saw wasn\'t completely finished (both color and music had not been finalized , so no comments about marilyn manson) , but cinematographer peter deming (don\'t say a word) ably captures the dreariness of victorian-era london and helped make the flashy killing scenes remind me of the crazy flashbacks in twin peaks , even though the violence in the film pales in comparison to that in the black-and-white comic . \noscar winner martin childs\' (shakespeare in love) production design turns the original prague surroundings into one creepy place . \neven the acting in from hell is solid , with the dreamy depp turning in a typically strong performance and deftly handling a british accent . \nians holm (joe gould\'s secret) and richardson (102 dalmatians) log in great supporting roles , but the big surprise here is graham . \ni cringed the first time she opened her mouth , imagining her attempt at an irish accent , but it actually wasn\'t half bad . \nthe film , however , is all good . \n2 : 00 - r for strong violence/gore , sexuality , language and drug content ', u'every now and then a movie comes along from a suspect studio , with every indication that it will be a stinker , and to everybody \'s surprise (perhaps even the studio) the film becomes a critical darling . \nmtv films\' _election , a high school comedy starring matthew broderick and reese witherspoon , is a current example . \ndid anybody know this film existed a week before it opened ? \nthe plot is deceptively simple . \ngeorge washington carver high school is having student elections . \ntracy flick (reese witherspoon) is an over-achiever with her hand raised at nearly every question , way , way , high . \nmr _ m " (matthew broderick) , sick of the megalomaniac student , encourages paul , a popular-but-slow jock to run . \nand paul\'s nihilistic sister jumps in the race as well , for personal reasons . \nthe dark side of such sleeper success is that , because expectations were so low going in , the fact that this was quality stuff made the reviews even more enthusiastic than they have any right to be . \nyou can\'t help going in with the baggage of glowing reviews , which is in contrast to the negative baggage that the reviewers were likely to have . \n_election , a good film , does not live up to its hype . \nwhat makes _election_ so disappointing is that it contains significant plot details lifted directly from _rushmore_ , released a few months earlier . \nthe similarities are staggering : \ntracy flick (_election_) is the president of an extraordinary number of clubs , and is involved with the school play . \nmax fischer (_rushmore_) is the president of an extraordinary number of clubs , and is involved with the school play . \nthe most significant tension of _election_ is the potential relationship between a teacher and his student . \nthe most significant tension of _rushmore_ is the potential relationship between a teacher and his student . \ntracy flick is from a single parent home , which has contributed to her drive . \nmax fischer is from a single parent home , which has contributed to his drive . \nthe male bumbling adult in _election_ (matthew broderick) pursues an extramarital affair , gets caught , and his whole life is ruined . \nhe even gets a bee sting . \nthe male bumbling adult in _rushmore_ (bill murray) pursues an extramarital affair , gets caught , and his whole life is ruined . \nhe gets several bee stings . \nand so on . \nwhat happened ? \nhow is it that an individual screenplay (_rushmore_) and a novel (_election_) contain so many significant plot points , and yet both films were probably not even aware of each other , made from two different studios , from a genre (the high school geeks revenge movie) that hadn\'t been fully formed yet ? \neven so , the strengths of _election_ rely upon its fantastic performances from broderick , witherspoon , and newcomer jessica campbell , as paul\'s anti-social sister , tammy . \nbroderick here is playing the mr . rooney role from _ferri s bueller_ , and he seems to be having the most fun he\'s had since then . \nwitherspoon is a revelation . \nit\'s early in the year , it\'s a comedy , and teenagers have little clout , but for my money , witherspoon deserves an oscar nomination . \nand once campbell\'s character gets going , like in her fantastic speech in the gymnasium , then you\'re won over . \none thing that\'s been bothering me since i\'ve seen it . \nthere is an extraordinary amount of sexuality in this film . \ni suppose that , coming from mtv films , i should expect no less . . . \nbut the film starts off light and airy , like a sitcom . \nas the screws tighten , and the tensions mount , alexander payne decides to add elements that , frankly , distract from the story . \nit is bad enough that mr . m doesn\'t like tracy\'s determination to win at all costs , but did they have to throw in the student/teacher relationship ? \neven so , there\'s no logical reason why mr . m has an affair when he does . \nthere\'s a lot to like in _election_ , but the plot similarities to _rushmore_ , and the tonal nosedive it takes as it gets explicitly sex-driven , mark this as a disappointment . ']

['neg', 'neg', 'pos', 'pos']

2.2 preprocess

To create a vocabulary dictionary and associates the tokens with a unique integer.

For example:

{u'all': 72, u'concept': 225, u'affair': 859, u'skip': 423, u'graham': 597, u"film's": 281, u'steve': 660, u'surgical': 560, u'9/10': 436, u'run': 797, u'go': 162, u'indians': 645, u'once': 906, u'seemed': 369, u'certainly': 666, u'william': 28, u'student/teacher': 937, u'flash': 73, u'suits': 358, u'presents': 193, u'to': 44, u'watchmen': 480, u'fischer': 844, u'reviewers': 821, u'sitting': 419, u'very': 69, u'horror': 405, u'continues': 176, u'burton': 672, u'simpsons': 655, u'chooses': 620, u'handling': 723, u'every': 269, u'jack': 485, u'surprise': 737, u'surgeon': 605, u'bringing': 59, u'expectations': 806, u"might've": 354, u'cool': 191, u'identity': 617, u'entire': 384, u'school': 766, u'tonal': 941, u'broderick': 769, u'level': 474, u'did': 310, u'turns': 564, u'dig': 303, u'michael': 488, u'brother': 26, u'solid': 717, u'race': 802, u'flashy': 91, u'quick': 4, u"it'll": 663, u'flick': 407, u'unsuccessfully': 582, u'neighborhood': 379, u'sooty': 548, u'crow': 435, u'direct': 519, u'past': 509, u'likely': 120, u'street': 428, u'design': 135, u'further': 345, u'befriends': 591, u'odd': 491, u'jessica': 890, u'even': 115, u'what': 85, u'hide': 285, u'giving': 341, u'saw': 679, u'kudos': 381, u'current': 772, u'suspect': 754, u'new': 378, u'ever': 421, u'critical': 760, u'contrast': 819, u'body': 147, u'supporting': 735, u'lee': 21, u'never': 462, u'here': 63, u'understanding': 324, u'quell': 584, u'entertain': 392, u'studio': 349, u'others': 248, u'102': 731, u'along': 753, u'strong': 720, u'teacher': 849, u'great': 734, u'substance': 75, u'involved': 841, u'30': 500, u'reason': 939, u'creepy': 716, u'amount': 914, u'nervous': 553, u'action': 61, u'newcomer': 889, u'weird': 257, u"someone's": 143, u'makes': 197, u'rafael': 629, u'named': 593, u'winner': 708, u'love': 712, u'although': 366, u'win': 934, u'black-and-white': 706, u'explained': 262, u'clout': 902, u'brought': 471, u'color': 682, u'stings': 868, u'ians': 726, u'drunkenly': 101, u'existed': 775, u'tammy': 892, u'plot': 157, u'fed': 277, u'disappearances': 253, u'from': 116, u'takes': 943, u'would': 364, u'&': 212, u'revelation': 898, u'yglesias': 630, u'single': 850, u'two': 159, u'throw': 936, u'few': 60, u'live': 822, u'explicitly': 944, u'music': 361, u'films': 204, u'themselves': 352, u'until': 289, u'sex-driven': 945, u'more': 119, u'teen': 160, u'comparison': 705, u'halloween': 124, u'sister': 800, u'clue': 274, u'payne': 926, u'holm': 727, u'stick': 390, u'given': 321, u'had': 446, u'me': 272, u'arthouse': 458, u'keeping': 635, u'graphic': 492, u'word': 689, u'this': 18, u'work': 123, u'london': 694, u'reviews': 812, u'movies': 316, u'can': 508, u'rooney': 893, u'mr': 789, u'making': 215, u'male': 854, u'my': 380, u'example': 773, u'crazy': 698, u'months': 833, u'give': 271, u"london's": 544, u'geeks': 880, u"didn't": 218, u'days': 656, u'high': 765, u'coltrane': 569, u'carrot': 523, u'want': 287, u'sense': 312, u'appearance': 665, u'story': 31, u'simple': 778, u'pink': 90, u'needs': 610, u'_rushmore_': 831, u'end': 546, u'turn': 149, u'holds': 382, u'copious': 585, u'write': 199, u'how': 340, u'hot': 412, u'worry': 662, u'chase': 106, u'stir': 440, u'pales': 704, u'!': 336, u'okay': 337, u'beauty': 376, u'mess': 208, u'overall': 388, u'after': 279, u'blame': 642, u'screenwriters': 624, u'tries': 583, u'mad': 535, u'coming': 246, u'such': 209, u'guys': 171, u'types': 216, u'popular-but-slow': 795, u'a': 14, u'downshifts': 238, u'chasing': 338, u'light': 918, u'robot': 134, u'redundant': 396, u'deserted': 38, u'perhaps': 758, u'so': 145, u'enter': 323, u'cloaking': 623, u'playing': 371, u"that's": 528, u'imagining': 742, u'chopped': 351, u'executed': 226, u'help': 816, u"don't": 77, u'unfortunates': 552, u'over': 87, u'insight': 346, u'murray': 866, u'years': 417, u'course': 95, u'through': 557, u'ghastly': 650, u'hell': 464, u'still': 67, u'its': 229, u'before': 320, u'dreamy': 718, u'apes': 674, u'thrilling': 295, u"he's": 128, u'somewhere': 357, u'exact': 372, u',': 65, u'crew': 33, u'meantime': 298, u'parts': 148, u'whole': 473, u'hidden': 636, u'main': 230, u'might': 510, u'happened': 869, u'then': 166, u'bumbling': 855, u'good': 139, u'curtis': 22, u'washington': 780, u'killer': 618, u'regarding': 32, u'victorian-era': 693, u'break': 206, u'bothering': 911, u'they': 47, u'half': 744, u'not': 261, u'now': 263, u'viewers': 637, u'suppose': 915, u'comedy': 767, u'robbie': 568, u'always': 318, u'stiff': 563, u'bleak': 668, u'mark': 946, u'arrow': 301, u'each': 878, u'side': 804, u"isn't": 598, u'mean': 325, u'flashbacks': 699, u'lifted': 829, u'directly': 830, u'bentley': 368, u'tension': 845, u'generation': 189, u'series': 478, u'crimes': 603, u'idea': 192, u'police': 604, u'engaging': 297, u'society': 540, u'disappointing': 825, u'books': 445, u'expect': 916, u'eddie': 469, u'year': 900, u'happen': 259, u'really': 80, u'jackson': 489, u'opened': 740, u'kids': 413, u'sutherland': 97, u'since': 200, u'investigate': 600, u'content': 752, u'half-way': 308, u'acting': 112, u'max': 843, u'print': 678, u'got': 13, u'darling': 761, u'caught': 861, u'highway': 211, u'shows': 353, u'blair': 432, u'turning': 360, u'theory': 616, u"doesn't": 389, u'several': 867, u'little': 52, u'geared': 455, u'extraordinary': 838, u'completely': 288, u'log': 733, u"character's": 386, u'distract': 931, u'language': 750, u'shelves': 420, u'starts': 235, u'could': 647, u'success': 448, u'british': 724, u'times': 675, u'thing': 92, u'american': 375, u'glowing': 818, u'place': 549, u'guttenberg': 661, u'won': 910, u'footnotes': 503, u'onto': 664, u'think': 608, u'jumbled': 234, u'first': 562, u'origin': 84, u'copper': 565, u'features': 531, u'student': 783, u'1888': 543, u'number': 839, u'one': 170, u'fantasy': 239, u'logical': 938, u'another': 24, u'president': 837, u'decides': 927, u'comic': 444, u'visions': 331, u'george': 779, u'researched': 483, u"you're": 118, u'guess': 313, u'"': 103, u'quality': 810, u'filthy': 547, u'top': 524, u'accent': 725, u'anyone': 609, u'their': 558, u'2': 434, u'too': 233, u'wrapped': 415, u'airy': 919, u'final': 290, u'slasher': 406, u"paul's": 798, u'stomach': 607, u'lot': 409, u'godley': 567, u'that': 8, u'explanation': 400, u'jamie': 20, u'albert': 514, u'took': 86, u'alan': 467, u'megalomaniac': 792, u'part': 300, u'blindly': 640, u'wasted': 126, u'attempt': 205, u'than': 534, u"here's": 151, u'rables': 633, u'kind': 276, u"hell's": 512, u'roles': 736, u'clubs': 840, u'bother': 622, u'scenes': 255, u"childs'": 710, u'r': 747, u'were': 807, u'video': 362, u'feeling': 385, u'and': 23, u'slay': 621, u'mind': 265, u'teenagers': 901, u'sad': 299, u'say': 481, u'prophetic': 581, u'have': 222, u'need': 339, u'allen': 515, u'seem': 221, u'apparently': 348, u'any': 814, u'superman': 453, u'relationship': 847, u'baggage': 817, u'thoroughly': 482, u"i've": 912, u'-': 431, u'h20': 125, u'ably': 690, u'also': 414, u'potential': 846, u'which': 196, u'sleeper': 805, u'performance': 721, u'mis': 632, u'mir': 94, u'pursues': 857, u'reese': 770, u'play': 842, u'sure': 319, u'committing': 649, u'dalmatians': 732, u'normal': 237, u'who': 249, u'violence/gore': 748, u"where's": 424, u'most': 260, u'virus': 66, u'mouth': 741, u"gould's": 729, u'plenty': 447, u'nothing': 502, u'tracy': 785, u'contain': 872, u'why': 78, u'surprising': 669, u'released': 832, u'performances': 888, u'average': 114, u'frederick': 573, u'drive': 167, u'ways': 214, u'points': 874, u'tech': 40, u'entertaining': 294, u'typically': 719, u'away': 330, u'came': 402, u'saying': 487, u'shot': 140, u'witherspoon': 771, u'contributed': 853, u'ending': 399, u'deftly': 722, u'gruesome': 602, u'find': 186, u'touches': 190, u'deceptively': 777, u'craziness': 401, u'(': 27, u'finalized': 683, u'behind': 537, u'should': 317, u'starters': 465, u'surroundings': 715, u'only': 377, u'going': 56, u'pretty': 154, u'joblo': 425, u'folks': 217, u'finger': 641, u'do': 53, u'his': 174, u'hit': 93, u'get': 121, u'bueller_': 896, u'sitcom': 920, u'watch': 184, u'enthusiastic': 813, u'feels': 68, u'nearly': 499, u'words': 505, u'despite': 398, u'_election': 764, u'him': 178, u'becomes': 759, u'ghetto': 530, u"bastard's": 3, u'frankly': 930, u'brain': 144, u'whitechapel': 542, u"wasn't": 680, u'bad': 194, u'stuff': 811, u'briefed': 611, u'she': 739, u'abberline': 574, u'hype': 823, u"won't": 612, u'where': 550, u'looks': 670, u'set': 529, u'captures': 691, u'dreariness': 692, u'prague': 714, u'deming': 688, u'lazy': 335, u'see': 177, u'decided': 359, u'are': 131, u'contains': 826, u'bee': 863, u'acts': 651, u'locals': 639, u'subject': 484, u'00': 746, u'parent': 851, u'manson': 686, u'sorta': 185, u'movie': 5, u'review': 6, u'horribly': 601, u'3': 429, u"there's": 461, u'unravel': 267, u'melissa': 327, u'correctly': 220, u'between': 848, u'probably': 876, u'flicks': 304, u'jock': 796, u'across': 37, u'we': 76, u'vertical': 627, u'alexander': 925, u'packaged': 408, u'however': 745, u'killing': 696, u'nightmares': 180, u'job': 634, u'extramarital': 858, u'joe': 728, u"films'": 763, u'genre': 411, u'20': 333, u'picking': 141, u'marilyn': 685, u'memento': 213, u'both': 302, u'figured': 306, u'stinker': 756, u'crime': 533, u'drug': 751, u'personal': 803, u'negative': 820, u'les': 631, u'disappointment': 947, u'interesting': 615, u'witch': 433, u'geniuses': 536, u'adapted': 443, u'comes': 36, u'otherwise': 153, u'riddle': 526, u'screenplay': 871, u'point': 309, u'reasons': 619, u'character': 373, u'whatever': 422, u'hollow': 677, u"hadn't": 882, u'tons': 256, u'carver': 781, u'simply': 232, u'likes': 117, u'church': 163, u'throughout': 102, u'nomination': 905, u'staggering': 836, u'better': 527, u'nosedive': 942, u'capable': 648, u"it's": 12, u'schnazzy': 136, u'.': 11, u'secret': 284, u"winston's": 133, u'damn': 7, u'mary': 594, u'empty': 70, u'life': 179, u'encourages': 793, u"what's": 181, u'dismiss': 506, u'precision': 561, u'nihilistic': 799, u'finished': 681, u'accident': 169, u'determination': 933, u'personally': 264, u'stonecutters': 653, u'revenge': 881, u'case': 579, u'hughes': 516, u'novel': 493, u'look': 250, u'these': 107, u'plain': 334, u'mount': 924, u'bill': 865, u'harder': 198, u'elections': 784, u'hayes': 626, u'will': 494, u'apparitions': 252, u'while': 280, u'sexuality': 749, u'neat': 224, u'depp': 576, u'many': 873, u'fun': 897, u'spawn': 454, u'almost': 520, u'is': 98, u'ii': 539, u'it': 45, u'sting': 864, u'couples': 161, u'middle': 82, u'someone': 409, u'in': 17, u'inspector': 572, u'helped': 695, u'superheroes': 451, u'if': 146, u'funny': 638, u'different': 343, u'absinthe': 587, u')': 30, u'things': 258, u'make': 293, u"everybody's": 757, u'same': 273, u'proceeds': 599, u'member': 243, u'matthew': 768, u'strange': 251, u'speech': 908, u'unfortunate': 592, u'party': 164, u'gets': 860, u'week': 776, u'tighten': 922, u'peaks': 701, u'applaud': 203, u'drink': 165, u'upon': 589, u'hand': 787, u'director': 350, u'running': 329, u'cinematographer': 687, u'characters': 245, u'nightmare': 426, u'terribly': 227, u'12-part': 477, u'early': 899, u'off': 236, u'englishman': 646, u'fully': 883, u'i': 201, u'salvation': 437, u'well': 127, u'anybody': 774, u'obviously': 283, u'edge': 363, u"can't": 606, u'kelly': 595, u'echoes': 441, u'the': 1, u'over-achiever': 786, u'just': 88, u'less': 917, u'when': 46, u'ludicrous': 521, u'sleepy': 676, u'fantastic': 887, u'mid': 475, u'seen': 913, u'arriving': 590, u'violent': 532, u'casper': 457, u'campbell': 470, u'paul': 794, u'baldwin': 25, u'yet': 875, u'mysterious': 554, u'blow': 577, u'unique': 614, u'elm': 427, u'similarities': 835, u'excites': 395, u'seems': 286, u'snag': 219, u'wes': 367, u'source': 507, u'add': 928, u'other': 504, u'crack': 578, u'role': 894, u'has': 42, u'kick': 48, u'real': 129, u'casting': 522, u'around': 100, u'big': 89, u'showing': 326, u'dark': 667, u'five': 291, u'know': 54, u'world': 240, u'bit': 311, u'password': 322, u'10/10': 439, u'dreams': 244, u'strengths': 885, u'like': 71, u'lost': 210, u'screws': 921, u'twin': 700, u'remind': 697, u'audience': 242, u'indication': 755, u'_election_': 824, u'martin': 709, u"campbell's": 907, u'looooot': 254, u'anti-social': 891, u'because': 391, u'jews': 644, u'deal': 182, u'people': 108, u'mind-fuck': 188, u'sequences': 62, u'some': 110, u'back': 50, u'dead': 247, u'oh': 403, u'donald': 96, u'elements': 929, u'unraveling': 387, u'critique': 187, u'home': 852, u"tracy's": 932, u'peter': 566, u'confusing': 393, u'rarely': 394, u'for': 57, u'bottom': 314, u'though': 702, u'comments': 684, u'does': 270, u'shakespeare': 711, u'assuming': 410, u'?': 183, u'cgi': 137, u'be': 370, u'car/who': 658, u'power': 49, u'jumps': 801, u'ripper': 486, u'nowhere': 83, u'starring': 19, u'_ferris': 895, u'by': 307, u'earlier': 834, u'on': 51, u'about': 332, u'y2k': 9, u'opium': 588, u'anything': 525, u'package': 195, u'tensions': 923, u'of': 34, u'violence': 703, u'costs': 935, u'deserves': 904, u'decent': 356, u'runtime': 397, u'or': 296, u'block': 511, u'raised': 788, u'own': 383, u'significant': 827, u'strangeness': 43, u'into': 142, u'within': 55, u'profession': 559, u'studios': 879, u'getting': 517, u'down': 347, u'right': 815, u'widower': 580, u'been': 355, u'carving': 556, u'8/10': 442, u'actually': 305, u'particulars': 613, u'your': 152, u'east': 545, u'rely': 886, u'crowd': 459, u'her': 122, u'there': 64, u'question': 541, u'hey': 104, u'long': 497, u'pages': 496, u'start': 16, u"'80s": 476, u'much': 155, u'created': 466, u'low': 808, u'way': 404, u'whores': 551, u'biggest': 282, u':': 158, u'was': 79, u'happy': 2, u'ghost': 460, u'head': 15, u'medium': 472, u'toward': 456, u'offering': 344, u'enough': 570, u'batman': 452, u'but': 173, u'gymnasium': 909, u'taken': 223, u'gore': 58, u'russian': 39, u'line': 315, u'trying': 266, u'with': 109, u'bug': 10, u'production': 416, u'he': 374, u"they're": 450, u'made': 659, u'consist': 501, u'whether': 449, u'stumbling': 99, u'mtv': 762, u'up': 278, u'us': 342, u'ruined': 862, u'planet': 673, u'below': 113, u'limit': 628, u'500': 495, u'problem': 231, u'minutes': 292, u'called': 479, u'irish': 743, u'electric': 657, u'money': 903, u'moore': 468, u'johnny': 575, u'm': 790, u'4/10': 438, u'cringed': 738, u'an': 168, u'as': 241, u'at': 643, u'girlfriend': 175, u'ship': 41, u'psychopath': 555, u'mold': 207, u'sagemiller': 328, u'film': 268, u'again': 275, u'heather': 596, u'sunken': 156, u'no': 74, u'amounts': 586, u'generally': 202, u'occasional': 138, u'tim': 671, u'book': 463, u'actors': 365, u'details': 828, u'sick': 791, u'oscar': 707, u'7/10': 430, u'you': 150, u'out': 81, u'formed': 884, u'star': 130, u'brothers': 518, u'dies': 172, u'problems': 228, u'robots': 111, u'whistling': 652, u'includes': 498, u"let's": 105, u'terry': 625, u'stan': 132, u'song': 654, u'aware': 877, u'ago': 418, u'tugboat': 35, u'individual': 870, u'calls': 571, u'having': 782, u'directors': 513, u'fact': 809, u'time': 29, u'richardson': 730, u'menace': 538, u'starting': 490, u'original': 713, u'adult': 856}

2.3 Returns integer-encoded movie reviews.

[[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 5, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 18, 29, 30, 17, 14, 31, 32, 14, 33, 34, 14, 35, 8, 36, 37, 14, 38, 39, 40, 41, 8, 42, 14, 43, 44, 45, 46, 47, 48, 1, 49, 50, 51, 11, 52, 53, 47, 54, 1, 49, 55, 11, 11, 11, 56, 57, 1, 58, 23, 59, 51, 14, 60, 61, 62, 63, 23, 64, 65, 66, 67, 68, 69, 70, 65, 71, 14, 5, 56, 57, 72, 73, 23, 74, 75, 11, 76, 77, 54, 78, 1, 33, 79, 80, 81, 17, 1, 82, 34, 83, 65, 76, 77, 54, 1, 84, 34, 85, 86, 87, 1, 41, 27, 88, 8, 14, 89, 90, 91, 92, 93, 1, 94, 30, 65, 23, 65, 34, 95, 65, 76, 77, 54, 78, 96, 97, 98, 99, 100, 101, 102, 11, 63, 65, 12, 88, 103, 104, 65, 105, 106, 107, 108, 100, 109, 110, 111, 103, 11, 1, 112, 98, 113, 114, 65, 115, 116, 1, 117, 34, 22, 11, 118, 119, 120, 44, 121, 14, 48, 81, 34, 122, 123, 17, 124, 125, 11, 97, 98, 126, 23, 25, 65, 127, 65, 128, 112, 71, 14, 25, 65, 34, 95, 11, 1, 129, 130, 63, 131, 132, 133, 134, 135, 65, 110, 136, 137, 65, 23, 1, 138, 139, 58, 140, 65, 71, 141, 142, 143, 144, 11, 145, 65, 146, 111, 23, 147, 148, 80, 149, 150, 51, 65, 151, 152, 5, 11, 153, 65, 12, 154, 155, 14, 156, 41, 34, 14, 5, 11], [157, 158, 159, 160, 161, 162, 44, 14, 163, 164, 65, 165, 23, 166, 167, 11, 47, 121, 142, 168, 169, 11, 170, 34, 1, 171, 172, 65, 173, 174, 175, 176, 44, 177, 178, 17, 122, 179, 65, 23, 42, 180, 11, 181, 1, 182, 183, 184, 1, 5, 23, 103, 185, 103, 186, 81, 11, 11, 11, 187, 158, 14, 188, 5, 57, 1, 160, 189, 8, 190, 51, 14, 69, 191, 192, 65, 173, 193, 45, 17, 14, 69, 194, 195, 11, 196, 98, 85, 197, 18, 6, 168, 115, 198, 170, 44, 199, 65, 200, 201, 202, 203, 204, 196, 205, 44, 206, 1, 207, 65, 208, 109, 152, 15, 23, 209, 27, 210, 211, 212, 213, 30, 65, 173, 64, 131, 139, 23, 194, 214, 34, 215, 72, 216, 34, 204, 65, 23, 107, 217, 88, 218, 219, 18, 170, 220, 11, 47, 221, 44, 222, 223, 18, 154, 224, 225, 65, 173, 226, 45, 227, 11, 145, 85, 131, 1, 228, 109, 1, 5, 183, 127, 65, 229, 230, 231, 98, 8, 12, 232, 233, 234, 11, 45, 235, 236, 103, 237, 103, 173, 166, 238, 142, 18, 103, 239, 103, 240, 17, 196, 150, 65, 241, 168, 242, 243, 65, 222, 74, 192, 181, 56, 51, 11, 64, 131, 244, 65, 64, 131, 245, 246, 50, 116, 1, 247, 65, 64, 131, 248, 249, 250, 71, 1, 247, 65, 64, 131, 251, 252, 65, 64, 131, 253, 65, 64, 131, 14, 254, 34, 106, 255, 65, 64, 131, 256, 34, 257, 258, 8, 259, 65, 23, 260, 34, 45, 98, 232, 261, 262, 11, 263, 201, 264, 77, 265, 266, 44, 267, 14, 268, 269, 263, 23, 166, 65, 173, 46, 72, 45, 270, 98, 271, 272, 1, 273, 274, 87, 23, 87, 275, 65, 201, 121, 276, 34, 277, 278, 279, 14, 280, 65, 196, 98, 18, 281, 282, 231, 11, 12, 283, 13, 18, 89, 284, 44, 285, 65, 173, 45, 286, 44, 287, 44, 285, 45, 288, 289, 229, 290, 291, 292, 11, 23, 53, 47, 293, 258, 294, 65, 295, 296, 115, 297, 65, 17, 1, 298, 183, 261, 80, 11, 1, 299, 300, 98, 8, 1, 301, 23, 201, 302, 303, 51, 304, 71, 18, 65, 145, 76, 305, 306, 260, 34, 45, 81, 307, 1, 308, 309, 65, 145, 72, 34, 1, 43, 279, 8, 310, 16, 44, 293, 14, 52, 311, 34, 312, 65, 173, 45, 67, 218, 1, 293, 1, 268, 72, 8, 119, 294, 11, 201, 313, 1, 314, 315, 109, 316, 71, 18, 98, 8, 150, 317, 318, 293, 319, 8, 1, 242, 98, 103, 142, 45, 103, 115, 320, 47, 131, 321, 1, 284, 322, 44, 323, 152, 240, 34, 324, 11, 201, 325, 65, 326, 327, 328, 329, 330, 116, 331, 57, 332, 333, 292, 102, 1, 5, 98, 88, 334, 335, 336, 336, 337, 65, 76, 121, 45, 11, 11, 11, 64, 131, 108, 338, 122, 23, 76, 77, 54, 249, 47, 131, 11, 53, 76, 80, 339, 44, 177, 45, 87, 23, 87, 275, 183, 340, 332, 341, 342, 343, 255, 344, 345, 346, 142, 72, 34, 1, 43, 56, 347, 17, 1, 5, 183, 348, 65, 1, 349, 86, 18, 268, 330, 116, 229, 350, 23, 351, 45, 278, 352, 65, 23, 45, 353, 11, 64, 354, 355, 14, 154, 356, 160, 188, 5, 17, 63, 357, 65, 173, 201, 313, 103, 1, 358, 103, 359, 8, 360, 45, 142, 14, 361, 362, 109, 52, 363, 65, 364, 293, 119, 312, 11, 1, 365, 131, 154, 139, 57, 1, 260, 300, 65, 366, 367, 368, 88, 369, 44, 370, 371, 1, 372, 273, 373, 8, 374, 310, 17, 375, 376, 65, 377, 17, 14, 378, 379, 11, 173, 380, 282, 381, 162, 81, 44, 328, 65, 249, 382, 122, 383, 102, 1, 384, 268, 65, 23, 305, 42, 150, 385, 122, 386, 387, 11, 388, 65, 1, 268, 389, 390, 391, 45, 389, 392, 65, 12, 393, 65, 45, 394, 395, 23, 45, 68, 154, 396, 57, 260, 34, 229, 397, 65, 398, 14, 154, 191, 399, 23, 400, 44, 72, 34, 1, 401, 8, 402, 320, 45, 11, 403, 65, 23, 307, 1, 404, 65, 18, 98, 261, 14, 405, 296, 160, 406, 407, 11, 11, 11, 12, 88, 408, 44, 250, 8, 404, 391, 409, 98, 348, 410, 8, 1, 411, 98, 67, 412, 109, 1, 413, 11, 45, 414, 415, 416, 159, 417, 418, 23, 42, 355, 419, 51, 1, 420, 421, 200, 11, 422, 11, 11, 11, 423, 45, 336, 424, 425, 246, 116, 183, 14, 426, 34, 427, 428, 429, 27, 430, 30, 431, 432, 433, 434, 27, 430, 30, 431, 1, 435, 27, 436, 30, 431, 1, 435, 158, 437, 27, 438, 30, 431, 210, 211, 27, 439, 30, 431, 213, 27, 439, 30, 431, 1, 248, 27, 436, 30, 431, 440, 34, 441, 27, 442, 30], [204, 443, 116, 444, 445, 222, 446, 447, 34, 448, 65, 449, 450, 332, 451, 27, 452, 65, 453, 65, 454, 30, 65, 296, 455, 456, 413, 27, 457, 30, 296, 1, 458, 459, 27, 460, 240, 30, 65, 173, 461, 462, 80, 355, 14, 444, 463, 71, 116, 464, 320, 11, 57, 465, 65, 45, 79, 466, 307, 467, 468, 27, 23, 469, 470, 30, 65, 249, 471, 1, 472, 44, 14, 473, 378, 474, 17, 1, 475, 476, 109, 14, 477, 478, 479, 1, 480, 11, 44, 481, 468, 23, 470, 482, 483, 1, 484, 34, 485, 1, 486, 364, 370, 71, 487, 488, 489, 98, 490, 44, 250, 14, 52, 491, 11, 1, 463, 27, 296, 103, 492, 493, 65, 103, 146, 150, 494, 30, 98, 87, 495, 496, 497, 23, 498, 499, 500, 119, 8, 501, 34, 502, 173, 503, 11, 17, 504, 505, 65, 77, 506, 18, 268, 391, 34, 229, 507, 11, 146, 150, 508, 121, 509, 1, 473, 444, 463, 92, 65, 150, 510, 186, 24, 99, 511, 17, 116, 512, 513, 65, 514, 23, 515, 516, 11, 517, 1, 516, 518, 44, 519, 18, 286, 520, 241, 521, 241, 522, 523, 524, 17, 65, 127, 65, 525, 65, 173, 526, 272, 18, 158, 249, 527, 44, 519, 14, 268, 528, 529, 17, 1, 530, 23, 531, 80, 532, 428, 533, 534, 1, 535, 536, 537, 538, 539, 540, 183, 1, 530, 17, 541, 98, 65, 34, 95, 65, 542, 17, 543, 544, 545, 546, 11, 12, 14, 547, 65, 548, 549, 550, 1, 551, 27, 479, 103, 552, 103, 30, 131, 490, 44, 121, 14, 52, 553, 332, 18, 554, 555, 249, 42, 355, 556, 557, 558, 559, 109, 560, 561, 11, 46, 1, 562, 563, 564, 278, 65, 565, 566, 567, 27, 568, 569, 65, 1, 240, 98, 261, 570, 30, 571, 17, 572, 573, 574, 27, 575, 576, 65, 577, 30, 44, 578, 1, 579, 11, 574, 65, 14, 580, 65, 42, 581, 244, 374, 582, 583, 44, 584, 109, 585, 586, 34, 587, 23, 588, 11, 589, 590, 17, 542, 65, 374, 591, 168, 592, 593, 594, 595, 27, 596, 597, 65, 481, 45, 598, 145, 30, 23, 599, 44, 600, 1, 601, 602, 603, 8, 115, 1, 604, 605, 606, 607, 11, 201, 77, 608, 609, 610, 44, 370, 611, 51, 485, 1, 486, 65, 145, 201, 612, 162, 142, 1, 613, 63, 65, 504, 534, 44, 481, 468, 23, 470, 222, 14, 614, 23, 615, 616, 332, 302, 1, 617, 34, 1, 618, 23, 1, 619, 374, 620, 44, 621, 11, 17, 1, 444, 65, 47, 77, 622, 623, 1, 617, 34, 1, 486, 65, 173, 624, 625, 626, 27, 627, 628, 30, 23, 629, 630, 27, 631, 632, 183, 633, 30, 53, 14, 139, 634, 34, 635, 178, 636, 116, 637, 289, 1, 69, 546, 11, 12, 638, 44, 184, 1, 639, 640, 309, 1, 641, 34, 642, 643, 644, 23, 645, 391, 65, 279, 72, 65, 168, 646, 647, 462, 370, 648, 34, 649, 209, 650, 651, 11, 23, 116, 512, 399, 446, 272, 652, 1, 653, 654, 116, 1, 655, 57, 656, 27, 103, 249, 382, 50, 1, 657, 658, 659, 660, 661, 14, 130, 183, 103, 30, 11, 77, 662, 431, 663, 72, 293, 312, 46, 150, 177, 45, 11, 263, 664, 116, 512, 665, 158, 12, 666, 667, 23, 668, 570, 65, 23, 12, 669, 44, 177, 340, 155, 119, 45, 670, 71, 14, 671, 672, 268, 534, 673, 34, 1, 674, 310, 27, 643, 675, 65, 45, 286, 71, 676, 677, 434, 30, 11, 1, 678, 201, 679, 680, 288, 681, 27, 302, 682, 23, 361, 446, 261, 355, 683, 65, 145, 74, 684, 332, 685, 686, 30, 65, 173, 687, 566, 688, 27, 77, 481, 14, 689, 30, 690, 691, 1, 692, 34, 693, 694, 23, 695, 293, 1, 91, 696, 255, 697, 272, 34, 1, 698, 699, 17, 700, 701, 65, 115, 702, 1, 703, 17, 1, 268, 704, 17, 705, 44, 8, 17, 1, 706, 444, 11, 707, 708, 709, 710, 27, 711, 17, 712, 30, 416, 135, 564, 1, 713, 714, 715, 142, 170, 716, 549, 11, 115, 1, 112, 17, 116, 464, 98, 717, 65, 109, 1, 718, 576, 360, 17, 14, 719, 720, 721, 23, 722, 723, 14, 724, 725, 11, 726, 727, 27, 728, 729, 284, 30, 23, 730, 27, 731, 732, 30, 733, 17, 734, 735, 736, 65, 173, 1, 89, 737, 63, 98, 597, 11, 201, 738, 1, 562, 29, 739, 740, 122, 741, 65, 742, 122, 205, 643, 168, 743, 725, 65, 173, 45, 305, 680, 744, 194, 11, 1, 268, 65, 745, 65, 98, 72, 139, 11, 434, 158, 746, 431, 747, 57, 720, 748, 65, 749, 65, 750, 23, 751, 752], [269, 263, 23, 166, 14, 5, 36, 753, 116, 14, 754, 349, 65, 109, 269, 755, 8, 45, 494, 370, 14, 756, 65, 23, 44, 757, 737, 27, 758, 115, 1, 349, 30, 1, 268, 759, 14, 760, 761, 11, 762, 763, 764, 65, 14, 765, 766, 767, 19, 768, 769, 23, 770, 771, 65, 98, 14, 772, 773, 11, 310, 774, 54, 18, 268, 775, 14, 776, 320, 45, 740, 183, 1, 157, 98, 777, 778, 11, 779, 780, 781, 765, 766, 98, 782, 783, 784, 11, 785, 407, 27, 770, 771, 30, 98, 168, 786, 109, 122, 787, 788, 643, 499, 269, 541, 65, 404, 65, 404, 65, 765, 11, 789, 11, 103, 790, 103, 27, 768, 769, 30, 65, 791, 34, 1, 792, 783, 65, 793, 794, 65, 14, 795, 796, 44, 797, 11, 23, 798, 799, 800, 801, 17, 1, 802, 241, 127, 65, 57, 803, 619, 11, 1, 667, 804, 34, 209, 805, 448, 98, 8, 65, 391, 806, 807, 145, 808, 56, 17, 65, 1, 809, 8, 18, 79, 810, 811, 659, 1, 812, 115, 119, 813, 534, 47, 222, 814, 815, 44, 370, 11, 150, 606, 816, 56, 17, 109, 1, 817, 34, 818, 812, 65, 196, 98, 17, 819, 44, 1, 820, 817, 8, 1, 821, 807, 120, 44, 222, 11, 764, 65, 14, 139, 268, 65, 270, 261, 822, 278, 44, 229, 823, 11, 85, 197, 824, 145, 825, 98, 8, 45, 826, 827, 157, 828, 829, 830, 116, 831, 65, 832, 14, 60, 833, 834, 11, 1, 835, 131, 836, 158, 785, 407, 27, 824, 30, 98, 1, 837, 34, 168, 838, 839, 34, 840, 65, 23, 98, 841, 109, 1, 766, 842, 11, 843, 844, 27, 831, 30, 98, 1, 837, 34, 168, 838, 839, 34, 840, 65, 23, 98, 841, 109, 1, 766, 842, 11, 1, 260, 827, 845, 34, 824, 98, 1, 846, 847, 848, 14, 849, 23, 174, 783, 11, 1, 260, 827, 845, 34, 831, 98, 1, 846, 847, 848, 14, 849, 23, 174, 783, 11, 785, 407, 98, 116, 14, 850, 851, 852, 65, 196, 42, 853, 44, 122, 167, 11, 843, 844, 98, 116, 14, 850, 851, 852, 65, 196, 42, 853, 44, 174, 167, 11, 1, 854, 855, 856, 17, 824, 27, 768, 769, 30, 857, 168, 858, 859, 65, 860, 861, 65, 23, 174, 473, 179, 98, 862, 11, 374, 115, 860, 14, 863, 864, 11, 1, 854, 855, 856, 17, 831, 27, 865, 866, 30, 857, 168, 858, 859, 65, 860, 861, 65, 23, 174, 473, 179, 98, 862, 11, 374, 860, 867, 863, 868, 11, 23, 145, 51, 11, 85, 869, 183, 340, 98, 45, 8, 168, 870, 871, 27, 831, 30, 23, 14, 493, 27, 824, 30, 872, 145, 873, 827, 157, 874, 65, 23, 875, 302, 204, 807, 876, 261, 115, 877, 34, 878, 504, 65, 659, 116, 159, 343, 879, 65, 116, 14, 411, 27, 1, 765, 766, 880, 881, 5, 30, 8, 882, 355, 883, 884, 875, 183, 115, 145, 65, 1, 885, 34, 824, 886, 589, 229, 887, 888, 116, 769, 65, 771, 65, 23, 889, 890, 470, 65, 241, 798, 891, 800, 65, 892, 11, 769, 63, 98, 371, 1, 789, 11, 893, 894, 116, 895, 896, 65, 23, 374, 286, 44, 370, 782, 1, 260, 897, 128, 446, 200, 166, 11, 771, 98, 14, 898, 11, 12, 899, 17, 1, 900, 65, 12, 14, 767, 65, 23, 901, 222, 52, 902, 65, 173, 57, 380, 903, 65, 771, 904, 168, 707, 905, 11, 23, 906, 907, 373, 860, 56, 65, 71, 17, 122, 887, 908, 17, 1, 909, 65, 166, 118, 910, 87, 11, 170, 92, 528, 355, 911, 272, 200, 912, 913, 45, 11, 64, 98, 168, 838, 914, 34, 749, 17, 18, 268, 11, 201, 915, 8, 65, 246, 116, 762, 204, 65, 201, 317, 916, 74, 917, 11, 11, 11, 173, 1, 268, 235, 236, 918, 23, 919, 65, 71, 14, 920, 11, 241, 1, 921, 922, 65, 23, 1, 923, 924, 65, 925, 926, 927, 44, 928, 929, 8, 65, 930, 65, 931, 116, 1, 31, 11, 45, 98, 194, 570, 8, 789, 11, 790, 389, 71, 932, 933, 44, 934, 643, 72, 935, 65, 173, 310, 47, 222, 44, 936, 17, 1, 937, 847, 183, 115, 145, 65, 461, 74, 938, 939, 78, 789, 11, 790, 42, 168, 859, 46, 374, 270, 11, 461, 14, 940, 44, 71, 17, 824, 65, 173, 1, 157, 835, 44, 831, 65, 23, 1, 941, 942, 45, 943, 241, 45, 860, 944, 945, 65, 946, 18, 241, 14, 947, 11]]

2.4. Returns integer-encoded movie reviews labels.

LABELS = [0, 0, 1, 1]

2.5 Padding zeros to each review to make all reviews have equal length.

[[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 5, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 18, 29, 30, 17, 14, 31, 32, 14, 33, 34, 14, 35, 8, 36, 37, 14, 38, 39, 40, 41, 8, 42, 14, 43, 44, 45, 46, 47, 48, 1, 49, 50, 51, 11, 52, 53, 47, 54, 1, 49, 55, 11, 11, 11, 56, 57, 1, 58, 23, 59, 51, 14, 60, 61, 62, 63, 23, 64, 65, 66, 67, 68, 69, 70, 65, 71, 14, 5, 56, 57, 72, 73, 23, 74, 75, 11, 76, 77, 54, 78, 1, 33, 79, 80, 81, 17, 1, 82, 34, 83, 65, 76, 77, 54, 1, 84, 34, 85, 86, 87, 1, 41, 27, 88, 8, 14, 89, 90, 91, 92, 93, 1, 94, 30, 65, 23, 65, 34, 95, 65, 76, 77, 54, 78, 96, 97, 98, 99, 100, 101, 102, 11, 63, 65, 12, 88, 103, 104, 65, 105, 106, 107, 108, 100, 109, 110, 111, 103, 11, 1, 112, 98, 113, 114, 65, 115, 116, 1, 117, 34, 22, 11, 118, 119, 120, 44, 121, 14, 48, 81, 34, 122, 123, 17, 124, 125, 11, 97, 98, 126, 23, 25, 65, 127, 65, 128, 112, 71, 14, 25, 65, 34, 95, 11, 1, 129, 130, 63, 131, 132, 133, 134, 135, 65, 110, 136, 137, 65, 23, 1, 138, 139, 58, 140, 65, 71, 141, 142, 143, 144, 11, 145, 65, 146, 111, 23, 147, 148, 80, 149, 150, 51, 65, 151, 152, 5, 11, 153, 65, 12, 154, 155, 14, 156, 41, 34, 14, 5, 11, 0], [157, 158, 159, 160, 161, 162, 44, 14, 163, 164, 65, 165, 23, 166, 167, 11, 47, 121, 142, 168, 169, 11, 170, 34, 1, 171, 172, 65, 173, 174, 175, 176, 44, 177, 178, 17, 122, 179, 65, 23, 42, 180, 11, 181, 1, 182, 183, 184, 1, 5, 23, 103, 185, 103, 186, 81, 11, 11, 11, 187, 158, 14, 188, 5, 57, 1, 160, 189, 8, 190, 51, 14, 69, 191, 192, 65, 173, 193, 45, 17, 14, 69, 194, 195, 11, 196, 98, 85, 197, 18, 6, 168, 115, 198, 170, 44, 199, 65, 200, 201, 202, 203, 204, 196, 205, 44, 206, 1, 207, 65, 208, 109, 152, 15, 23, 209, 27, 210, 211, 212, 213, 30, 65, 173, 64, 131, 139, 23, 194, 214, 34, 215, 72, 216, 34, 204, 65, 23, 107, 217, 88, 218, 219, 18, 170, 220, 11, 47, 221, 44, 222, 223, 18, 154, 224, 225, 65, 173, 226, 45, 227, 11, 145, 85, 131, 1, 228, 109, 1, 5, 183, 127, 65, 229, 230, 231, 98, 8, 12, 232, 233, 234, 11, 45, 235, 236, 103, 237, 103, 173, 166, 238, 142, 18, 103, 239, 103, 240, 17, 196, 150, 65, 241, 168, 242, 243, 65, 222, 74, 192, 181, 56, 51, 11, 64, 131, 244, 65, 64, 131, 245, 246, 50, 116, 1, 247, 65, 64, 131, 248, 249, 250, 71, 1, 247, 65, 64, 131, 251, 252, 65, 64, 131, 253, 65, 64, 131, 14, 254, 34, 106, 255, 65, 64, 131, 256, 34, 257, 258, 8, 259, 65, 23, 260, 34, 45, 98, 232, 261, 262, 11, 263, 201, 264, 77, 265, 266, 44, 267, 14, 268, 269, 263, 23, 166, 65, 173, 46, 72, 45, 270, 98, 271, 272, 1, 273, 274, 87, 23, 87, 275, 65, 201, 121, 276, 34, 277, 278, 279, 14, 280, 65, 196, 98, 18, 281, 282, 231, 11, 12, 283, 13, 18, 89, 284, 44, 285, 65, 173, 45, 286, 44, 287, 44, 285, 45, 288, 289, 229, 290, 291, 292, 11, 23, 53, 47, 293, 258, 294, 65, 295, 296, 115, 297, 65, 17, 1, 298, 183, 261, 80, 11, 1, 299, 300, 98, 8, 1, 301, 23, 201, 302, 303, 51, 304, 71, 18, 65, 145, 76, 305, 306, 260, 34, 45, 81, 307, 1, 308, 309, 65, 145, 72, 34, 1, 43, 279, 8, 310, 16, 44, 293, 14, 52, 311, 34, 312, 65, 173, 45, 67, 218, 1, 293, 1, 268, 72, 8, 119, 294, 11, 201, 313, 1, 314, 315, 109, 316, 71, 18, 98, 8, 150, 317, 318, 293, 319, 8, 1, 242, 98, 103, 142, 45, 103, 115, 320, 47, 131, 321, 1, 284, 322, 44, 323, 152, 240, 34, 324, 11, 201, 325, 65, 326, 327, 328, 329, 330, 116, 331, 57, 332, 333, 292, 102, 1, 5, 98, 88, 334, 335, 336, 336, 337, 65, 76, 121, 45, 11, 11, 11, 64, 131, 108, 338, 122, 23, 76, 77, 54, 249, 47, 131, 11, 53, 76, 80, 339, 44, 177, 45, 87, 23, 87, 275, 183, 340, 332, 341, 342, 343, 255, 344, 345, 346, 142, 72, 34, 1, 43, 56, 347, 17, 1, 5, 183, 348, 65, 1, 349, 86, 18, 268, 330, 116, 229, 350, 23, 351, 45, 278, 352, 65, 23, 45, 353, 11, 64, 354, 355, 14, 154, 356, 160, 188, 5, 17, 63, 357, 65, 173, 201, 313, 103, 1, 358, 103, 359, 8, 360, 45, 142, 14, 361, 362, 109, 52, 363, 65, 364, 293, 119, 312, 11, 1, 365, 131, 154, 139, 57, 1, 260, 300, 65, 366, 367, 368, 88, 369, 44, 370, 371, 1, 372, 273, 373, 8, 374, 310, 17, 375, 376, 65, 377, 17, 14, 378, 379, 11, 173, 380, 282, 381, 162, 81, 44, 328, 65, 249, 382, 122, 383, 102, 1, 384, 268, 65, 23, 305, 42, 150, 385, 122, 386, 387, 11, 388, 65, 1, 268, 389, 390, 391, 45, 389, 392, 65, 12, 393, 65, 45, 394, 395, 23, 45, 68, 154, 396, 57, 260, 34, 229, 397, 65, 398, 14, 154, 191, 399, 23, 400, 44, 72, 34, 1, 401, 8, 402, 320, 45, 11, 403, 65, 23, 307, 1, 404, 65, 18, 98, 261, 14, 405, 296, 160, 406, 407, 11, 11, 11, 12, 88, 408, 44, 250, 8, 404, 391, 409, 98, 348, 410, 8, 1, 411, 98, 67, 412, 109, 1, 413, 11, 45, 414, 415, 416, 159, 417, 418, 23, 42, 355, 419, 51, 1, 420, 421, 200, 11, 422, 11, 11, 11, 423, 45, 336, 424, 425, 246, 116, 183, 14, 426, 34, 427, 428, 429, 27, 430, 30, 431, 432, 433, 434, 27, 430, 30, 431, 1, 435, 27, 436, 30, 431, 1, 435, 158, 437, 27, 438, 30, 431, 210, 211, 27, 439, 30, 431, 213, 27, 439, 30, 431, 1, 248, 27, 436, 30, 431, 440, 34, 441, 27, 442, 30], [204, 443, 116, 444, 445, 222, 446, 447, 34, 448, 65, 449, 450, 332, 451, 27, 452, 65, 453, 65, 454, 30, 65, 296, 455, 456, 413, 27, 457, 30, 296, 1, 458, 459, 27, 460, 240, 30, 65, 173, 461, 462, 80, 355, 14, 444, 463, 71, 116, 464, 320, 11, 57, 465, 65, 45, 79, 466, 307, 467, 468, 27, 23, 469, 470, 30, 65, 249, 471, 1, 472, 44, 14, 473, 378, 474, 17, 1, 475, 476, 109, 14, 477, 478, 479, 1, 480, 11, 44, 481, 468, 23, 470, 482, 483, 1, 484, 34, 485, 1, 486, 364, 370, 71, 487, 488, 489, 98, 490, 44, 250, 14, 52, 491, 11, 1, 463, 27, 296, 103, 492, 493, 65, 103, 146, 150, 494, 30, 98, 87, 495, 496, 497, 23, 498, 499, 500, 119, 8, 501, 34, 502, 173, 503, 11, 17, 504, 505, 65, 77, 506, 18, 268, 391, 34, 229, 507, 11, 146, 150, 508, 121, 509, 1, 473, 444, 463, 92, 65, 150, 510, 186, 24, 99, 511, 17, 116, 512, 513, 65, 514, 23, 515, 516, 11, 517, 1, 516, 518, 44, 519, 18, 286, 520, 241, 521, 241, 522, 523, 524, 17, 65, 127, 65, 525, 65, 173, 526, 272, 18, 158, 249, 527, 44, 519, 14, 268, 528, 529, 17, 1, 530, 23, 531, 80, 532, 428, 533, 534, 1, 535, 536, 537, 538, 539, 540, 183, 1, 530, 17, 541, 98, 65, 34, 95, 65, 542, 17, 543, 544, 545, 546, 11, 12, 14, 547, 65, 548, 549, 550, 1, 551, 27, 479, 103, 552, 103, 30, 131, 490, 44, 121, 14, 52, 553, 332, 18, 554, 555, 249, 42, 355, 556, 557, 558, 559, 109, 560, 561, 11, 46, 1, 562, 563, 564, 278, 65, 565, 566, 567, 27, 568, 569, 65, 1, 240, 98, 261, 570, 30, 571, 17, 572, 573, 574, 27, 575, 576, 65, 577, 30, 44, 578, 1, 579, 11, 574, 65, 14, 580, 65, 42, 581, 244, 374, 582, 583, 44, 584, 109, 585, 586, 34, 587, 23, 588, 11, 589, 590, 17, 542, 65, 374, 591, 168, 592, 593, 594, 595, 27, 596, 597, 65, 481, 45, 598, 145, 30, 23, 599, 44, 600, 1, 601, 602, 603, 8, 115, 1, 604, 605, 606, 607, 11, 201, 77, 608, 609, 610, 44, 370, 611, 51, 485, 1, 486, 65, 145, 201, 612, 162, 142, 1, 613, 63, 65, 504, 534, 44, 481, 468, 23, 470, 222, 14, 614, 23, 615, 616, 332, 302, 1, 617, 34, 1, 618, 23, 1, 619, 374, 620, 44, 621, 11, 17, 1, 444, 65, 47, 77, 622, 623, 1, 617, 34, 1, 486, 65, 173, 624, 625, 626, 27, 627, 628, 30, 23, 629, 630, 27, 631, 632, 183, 633, 30, 53, 14, 139, 634, 34, 635, 178, 636, 116, 637, 289, 1, 69, 546, 11, 12, 638, 44, 184, 1, 639, 640, 309, 1, 641, 34, 642, 643, 644, 23, 645, 391, 65, 279, 72, 65, 168, 646, 647, 462, 370, 648, 34, 649, 209, 650, 651, 11, 23, 116, 512, 399, 446, 272, 652, 1, 653, 654, 116, 1, 655, 57, 656, 27, 103, 249, 382, 50, 1, 657, 658, 659, 660, 661, 14, 130, 183, 103, 30, 11, 77, 662, 431, 663, 72, 293, 312, 46, 150, 177, 45, 11, 263, 664, 116, 512, 665, 158, 12, 666, 667, 23, 668, 570, 65, 23, 12, 669, 44, 177, 340, 155, 119, 45, 670, 71, 14, 671, 672, 268, 534, 673, 34, 1, 674, 310, 27, 643, 675, 65, 45, 286, 71, 676, 677, 434, 30, 11, 1, 678, 201, 679, 680, 288, 681, 27, 302, 682, 23, 361, 446, 261, 355, 683, 65, 145, 74, 684, 332, 685, 686, 30, 65, 173, 687, 566, 688, 27, 77, 481, 14, 689, 30, 690, 691, 1, 692, 34, 693, 694, 23, 695, 293, 1, 91, 696, 255, 697, 272, 34, 1, 698, 699, 17, 700, 701, 65, 115, 702, 1, 703, 17, 1, 268, 704, 17, 705, 44, 8, 17, 1, 706, 444, 11, 707, 708, 709, 710, 27, 711, 17, 712, 30, 416, 135, 564, 1, 713, 714, 715, 142, 170, 716, 549, 11, 115, 1, 112, 17, 116, 464, 98, 717, 65, 109, 1, 718, 576, 360, 17, 14, 719, 720, 721, 23, 722, 723, 14, 724, 725, 11, 726, 727, 27, 728, 729, 284, 30, 23, 730, 27, 731, 732, 30, 733, 17, 734, 735, 736, 65, 173, 1, 89, 737, 63, 98, 597, 11, 201, 738, 1, 562, 29, 739, 740, 122, 741, 65, 742, 122, 205, 643, 168, 743, 725, 65, 173, 45, 305, 680, 744, 194, 11, 1, 268, 65, 745, 65, 98, 72, 139, 11, 434, 158, 746, 431, 747, 57, 720, 748, 65, 749, 65, 750, 23, 751, 752, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0], [269, 263, 23, 166, 14, 5, 36, 753, 116, 14, 754, 349, 65, 109, 269, 755, 8, 45, 494, 370, 14, 756, 65, 23, 44, 757, 737, 27, 758, 115, 1, 349, 30, 1, 268, 759, 14, 760, 761, 11, 762, 763, 764, 65, 14, 765, 766, 767, 19, 768, 769, 23, 770, 771, 65, 98, 14, 772, 773, 11, 310, 774, 54, 18, 268, 775, 14, 776, 320, 45, 740, 183, 1, 157, 98, 777, 778, 11, 779, 780, 781, 765, 766, 98, 782, 783, 784, 11, 785, 407, 27, 770, 771, 30, 98, 168, 786, 109, 122, 787, 788, 643, 499, 269, 541, 65, 404, 65, 404, 65, 765, 11, 789, 11, 103, 790, 103, 27, 768, 769, 30, 65, 791, 34, 1, 792, 783, 65, 793, 794, 65, 14, 795, 796, 44, 797, 11, 23, 798, 799, 800, 801, 17, 1, 802, 241, 127, 65, 57, 803, 619, 11, 1, 667, 804, 34, 209, 805, 448, 98, 8, 65, 391, 806, 807, 145, 808, 56, 17, 65, 1, 809, 8, 18, 79, 810, 811, 659, 1, 812, 115, 119, 813, 534, 47, 222, 814, 815, 44, 370, 11, 150, 606, 816, 56, 17, 109, 1, 817, 34, 818, 812, 65, 196, 98, 17, 819, 44, 1, 820, 817, 8, 1, 821, 807, 120, 44, 222, 11, 764, 65, 14, 139, 268, 65, 270, 261, 822, 278, 44, 229, 823, 11, 85, 197, 824, 145, 825, 98, 8, 45, 826, 827, 157, 828, 829, 830, 116, 831, 65, 832, 14, 60, 833, 834, 11, 1, 835, 131, 836, 158, 785, 407, 27, 824, 30, 98, 1, 837, 34, 168, 838, 839, 34, 840, 65, 23, 98, 841, 109, 1, 766, 842, 11, 843, 844, 27, 831, 30, 98, 1, 837, 34, 168, 838, 839, 34, 840, 65, 23, 98, 841, 109, 1, 766, 842, 11, 1, 260, 827, 845, 34, 824, 98, 1, 846, 847, 848, 14, 849, 23, 174, 783, 11, 1, 260, 827, 845, 34, 831, 98, 1, 846, 847, 848, 14, 849, 23, 174, 783, 11, 785, 407, 98, 116, 14, 850, 851, 852, 65, 196, 42, 853, 44, 122, 167, 11, 843, 844, 98, 116, 14, 850, 851, 852, 65, 196, 42, 853, 44, 174, 167, 11, 1, 854, 855, 856, 17, 824, 27, 768, 769, 30, 857, 168, 858, 859, 65, 860, 861, 65, 23, 174, 473, 179, 98, 862, 11, 374, 115, 860, 14, 863, 864, 11, 1, 854, 855, 856, 17, 831, 27, 865, 866, 30, 857, 168, 858, 859, 65, 860, 861, 65, 23, 174, 473, 179, 98, 862, 11, 374, 860, 867, 863, 868, 11, 23, 145, 51, 11, 85, 869, 183, 340, 98, 45, 8, 168, 870, 871, 27, 831, 30, 23, 14, 493, 27, 824, 30, 872, 145, 873, 827, 157, 874, 65, 23, 875, 302, 204, 807, 876, 261, 115, 877, 34, 878, 504, 65, 659, 116, 159, 343, 879, 65, 116, 14, 411, 27, 1, 765, 766, 880, 881, 5, 30, 8, 882, 355, 883, 884, 875, 183, 115, 145, 65, 1, 885, 34, 824, 886, 589, 229, 887, 888, 116, 769, 65, 771, 65, 23, 889, 890, 470, 65, 241, 798, 891, 800, 65, 892, 11, 769, 63, 98, 371, 1, 789, 11, 893, 894, 116, 895, 896, 65, 23, 374, 286, 44, 370, 782, 1, 260, 897, 128, 446, 200, 166, 11, 771, 98, 14, 898, 11, 12, 899, 17, 1, 900, 65, 12, 14, 767, 65, 23, 901, 222, 52, 902, 65, 173, 57, 380, 903, 65, 771, 904, 168, 707, 905, 11, 23, 906, 907, 373, 860, 56, 65, 71, 17, 122, 887, 908, 17, 1, 909, 65, 166, 118, 910, 87, 11, 170, 92, 528, 355, 911, 272, 200, 912, 913, 45, 11, 64, 98, 168, 838, 914, 34, 749, 17, 18, 268, 11, 201, 915, 8, 65, 246, 116, 762, 204, 65, 201, 317, 916, 74, 917, 11, 11, 11, 173, 1, 268, 235, 236, 918, 23, 919, 65, 71, 14, 920, 11, 241, 1, 921, 922, 65, 23, 1, 923, 924, 65, 925, 926, 927, 44, 928, 929, 8, 65, 930, 65, 931, 116, 1, 31, 11, 45, 98, 194, 570, 8, 789, 11, 790, 389, 71, 932, 933, 44, 934, 643, 72, 935, 65, 173, 310, 47, 222, 44, 936, 17, 1, 937, 847, 183, 115, 145, 65, 461, 74, 938, 939, 78, 789, 11, 790, 42, 168, 859, 46, 374, 270, 11, 461, 14, 940, 44, 71, 17, 824, 65, 173, 1, 157, 835, 44, 831, 65, 23, 1, 941, 942, 45, 943, 241, 45, 860, 944, 945, 65, 946, 18, 241, 14, 947, 11, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]]

2.6 Loading the embedding word dictionary (simplified version as test)

```
{0: tensor([0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
        0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
        0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
        0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
        0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
        0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
        0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
        0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
        0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
        0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
        0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
        0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.,
        0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.]), 1: tensor([ 8.9700e-02,  1.6000e-02, -5.7100e-02,  4.0500e-02, -6.9600e-02,
        -1.2370e-01,  3.0100e-02,  2.4800e-02, -3.0300e-02,  1.7400e-02,
         6.3000e-03,  1.8400e-02,  2.1700e-02, -2.5700e-02,  3.5000e-02,
        -2.4200e-02,  2.9000e-03,  1.8800e-02, -5.7000e-02,  2.5200e-02,
        -2.1000e-02, -8.0000e-04,  3.6000e-02, -7.2900e-02, -6.6500e-02,
         9.8900e-02,  6.7600e-02,  8.5200e-02, -8.9000e-03,  3.1300e-02,
        -6.9000e-03, -3.2000e-03, -4.6200e-02,  4.9700e-02,  2.6100e-02,
         2.6800e-02, -3.1000e-02, -1.3610e-01, -6.2000e-03,  3.7500e-02,
        -3.2000e-02, -1.0600e-02,  5.3400e-02, -1.8700e-02,  6.3800e-02,
         9.4000e-03,  4.7000e-03, -5.3000e-02,  9.3000e-03, -8.7000e-03,
         4.0000e-04,  4.9300e-02, -6.2960e-01,  2.2200e-02,  1.9000e-02,
         2.6800e-02, -4.2600e-02,  5.7000e-03, -1.6830e-01,  2.4400e-02,
        -2.1300e-02, -1.8100e-02,  4.2100e-02, -3.0900e-02, -8.9000e-03,
         3.2000e-03,  1.0800e-02, -4.9000e-03,  2.5800e-02,  2.7800e-02,
        -1.6300e-02,  2.0000e-02,  1.6400e-02, -9.5400e-02, -3.2000e-03,
         4.3000e-03,  1.0400e-02, -8.8000e-03,  7.0000e-04,  3.5000e-02,
        -2.0600e-02, -8.3000e-03, -1.1400e-02, -1.8690e-01,  2.5800e-02,
         1.0000e-03,  8.5000e-03,  1.5100e-02,  2.1250e-01,  7.1000e-03,
         3.1900e-02, -4.8200e-02,  6.2100e-02,  6.2600e-02,  1.5900e-02,
        -1.3000e-03,  8.7000e-03,  6.8600e-02, -3.4000e-03,  2.3800e-02,
        -4.5200e-02, -1.9800e-02,  1.1200e-02,  1.0900e-02, -1.0220e-01,
        -2.7200e-02,  2.3370e-01, -4.6500e-02,  1.5920e-01, -4.0700e-02,
        -1.0290e-01, -4.8700e-02, -6.7600e-02,  6.7600e-02, -3.2800e-02,
         3.2300e-02,  7.7000e-03,  1.9000e-02,  1.7000e-03, -2.9740e-01,
         1.1000e-03, -3.5600e-02,  6.9300e-02, -4.8000e-02, -8.2100e-02,
        -6.4400e-02, -2.8400e-02, -1.9100e-02, -2.3300e-02,  3.5300e-02,
        -4.6300e-02,  6.5600e-02,  1.9000e-03, -2.1200e-02, -3.0900e-02,
        -3.5340e-01, -3.0900e-02,  7.6000e-03, -4.1900e-02,  4.5700e-02,
        -3.0600e-02,  3.5700e-02,  6.6700e-02,  3.6590e-01,  1.4900e-02,
        -4.4300e-02,  6.8000e-03, -3.7800e-02,  1.4600e-02,  2.1500e-02,
         1.0810e-01,  1.2400e-02, -4.3700e-02, -4.3000e-02,  2.5800e-02,
         2.1300e-02, -3.0900e-02, -1.8000e-03, -6.7000e-03,  1.7200e-02,
         8.9000e-03, -1.7100e-02,  2.7500e-02, -5.1800e-02, -1.8400e-01,
        -1.3000e-02, -2.4100e-02,  5.2600e-02, -2.8000e-02,  5.1000e-03,
         1.6300e-02, -1.6500e-02,  1.6100e-02,  1.2370e-01,  8.0400e-02,
        -7.8900e-02,  3.8600e-02, -3.8920e-01,  1.5700e-02, -2.4600e-02,
         4.7700e-02, -4.5000e-03, -2.1400e-02,  1.7300e-02, -1.9100e-02,
        -1.3820e-01, -1.1100e-02,  7.1200e-02,  1.5140e-01,  2.9100e-02,
         5.5500e-02, -3.9000e-03,  2.8000e-03, -2.7700e-02, -2.7500e-02,
        -1.7700e-02, -3.3800e-02, -3.7200e-02,  2.0710e-01,  4.6000e-02,
        -2.9400e-02,  4.3500e-02, -1.6900e-02, -1.2100e-02,  2.5300e-02,
         1.9800e-02,  9.1800e-02,  1.9300e-02,  6.6800e-02,  2.8800e-02,
         4.0000e-03, -4.3900e-02, -3.0200e-02,  6.4000e-03,  3.6400e-02,
         5.4300e-02, -3.3800e-02,  1.5900e-02,  6.1700e-02, -9.4100e-02,
        -8.6000e-03, -9.2000e-03,  3.0000e-02, -2.4100e-02, -3.5000e-02,
        -6.2100e-02,  1.7500e-02,  3.7400e-02,  3.4000e-03,  3.4400e-02,
         1.2860e-01, -2.6700e-02,  1.8610e-01,  4.8900e-02, -3.2000e-03,
         1.8000e-02, -2.2800e-02,  2.4140e-01, -9.3500e-02,  6.1200e-02,
        -2.0900e-02,  1.3600e-02,  3.9200e-02, -1.3500e-02, -2.5300e-02,
         3.3500e-02,  9.5000e-03,  4.1900e-02,  7.6000e-03,  4.5220e-01,
        -1.8800e-02,  2.3300e-02, -4.7400e-02,  1.5900e-02, -9.0000e-03,
         2.6500e-02,  3.3600e-02,  2.2100e-02,  4.7200e-02,  4.8000e-03,
         9.6200e-02,  3.4400e-02, -5.1500e-02, -8.7000e-03, -9.8000e-02,
        -2.8800e-02,  3.7700e-02,  2.0200e-02, -2.9790e-01, -3.8700e-02,
        -1.9800e-02, -1.6100e-02, -4.5000e-03,  8.7000e-03, -3.8700e-02,
         4.2100e-02,  3.8300e-02,  2.5800e-02,  6.9000e-03, -2.9800e-02,
        -1.9800e-02, -1.5200e-02,  3.3000e-03,  7.5000e-03,  3.5800e-02,
        -1.5500e-02, -1.1100e-02,  7.6000e-02, -4.5200e-02,  6.9700e-02,
         2.9900e-02, -2.9000e-03, -3.4800e-02, -2.7000e-02,  3.5100e-02,
         5.5900e-02,  5.9100e-02,  1.5590e-01, -2.5400e-02, -2.5900e-02]), ….
```

Note that some tensors are zeros because the embedding file does not have embedding vectors for these tokens.

2.7 Create a TensorDataset and DataLoader

For example:

```
(tensor([[  1,   2,   3,  ...,   0,   0,   0],
        [157, 158, 159,  ...,  27, 442,  30],
        [204, 443, 116,  ...,   0,   0,   0],
        [269, 263,  23,  ...,   0,   0,   0]]), tensor([0, 0, 1, 1]))
```

3. Define the baseline model, train and test.

The epoch is 5-10 and learning rate is 0.001-0.1

Accuracy of the deep learning on the test samples: 48-52 %

4. Add an RNN layer to the baseline model, train and test

The epoch is 5-10 and learning rate is 0.001-0.01

The number of RNN layers: 1

Accuracy of the deep learning on the test samples: 48-52 %

The number of RNN layers: 2

Accuracy of the deep learning on the test samples: 50-55 %

To make the RNN layer bi-directional.

Accuracy of the deep learning on the test samples: 50-55 %

5. Replace the RNN by self-attention, train and test

The epoch is 5-10 and learning rate is 0.001-0.01

Accuracy of the deep learning on the test samples: 50-52 %

Conclusion

The average accuracy of all three deep learning models on the test samples are around 50%