1

试一下beta，迭代次数不同时的效果；

Betarange0.01-0.09 0.1-0.9

数据集1

Beta 0.2 最好

[[0.26590909090909093, 0.44177777777777777, 0.49725, 0.5472857142857143, 0.5905, 0.6151, 0.64075, 0.67, 0.6925, 0.73], [0.258989898989899, 0.4538333333333333, 0.4926875, 0.5605, 0.6030833333333333, 0.6265, 0.65125, 0.6803333333333333, 0.704, 0.7405], [0.26661616161616164, 0.46416666666666667, 0.522125, 0.5717857142857142, 0.6134166666666667, 0.6351, 0.658625, 0.6873333333333334, 0.71025, 0.745], [0.23515151515151514, 0.4737777777777778, 0.527, 0.5777142857142857, 0.6178333333333333, 0.6415, 0.665, 0.693, 0.71275, 0.7485], [0.2293939393939394, 0.4920555555555556, 0.538375, 0.5846428571428571, 0.6211666666666666, 0.6458, 0.6665, 0.6956666666666667, 0.716, 0.751], [0.22116161616161617, 0.48677777777777775, 0.53575, 0.5856428571428571, 0.62375, 0.6503, 0.67, 0.696, 0.71825, 0.7525], [0.20015151515151516, 0.4955555555555556, 0.5460625, 0.5892857142857143, 0.6295, 0.6519, 0.671375, 0.6955, 0.7195, 0.758], [0.17525252525252524, 0.49372222222222223, 0.544625, 0.5891428571428572, 0.6321666666666667, 0.6543, 0.672875, 0.697, 0.72, 0.7625], [0.14651515151515151, 0.48794444444444446, 0.546875, 0.5984285714285714, 0.6320833333333333, 0.6533, 0.674125, 0.6991666666666667, 0.71975, 0.764], [0.15348484848484847, 0.4827222222222222, 0.5481875, 0.5953571428571428, 0.6335, 0.6558, 0.675625, 0.7008333333333333, 0.7215, 0.764], [0.12636363636363637, 0.4102222222222222, 0.534125, 0.5856428571428571, 0.6421666666666667, 0.6647, 0.685625, 0.7108333333333333, 0.7365, 0.7715], [0.05909090909090909, 0.282, 0.4961875, 0.5716428571428571, 0.6383333333333333, 0.6638, 0.685875, 0.7116666666666667, 0.74025, 0.777], [0.055656565656565654, 0.23244444444444445, 0.4438125, 0.5567857142857143, 0.63575, 0.6607, 0.6815, 0.7123333333333334, 0.743, 0.778], [0.05186868686868687, 0.20044444444444445, 0.384625, 0.5347142857142857, 0.6249166666666667, 0.6564, 0.681375, 0.7115, 0.7405, 0.7815], [0.0503030303030303, 0.1811111111111111, 0.3261875, 0.5076428571428572, 0.6183333333333333, 0.6517, 0.679375, 0.7095, 0.73625, 0.783], [0.05035353535353535, 0.1565, 0.284, 0.48085714285714287, 0.6105, 0.6452, 0.6775, 0.7081666666666667, 0.737, 0.786], [0.05085858585858586, 0.14794444444444443, 0.2419375, 0.44785714285714284, 0.6006666666666667, 0.6409, 0.675875, 0.7068333333333333, 0.737, 0.786], [0.05075757575757576, 0.14005555555555554, 0.2131875, 0.40785714285714286, 0.591, 0.6368, 0.673375, 0.7055, 0.73575, 0.7855]]

[[0.0504040404040404, 0.12805555555555556, 0.197, 0.3891428571428571, 0.5826666666666667, 0.6334, 0.6705, 0.705, 0.73625, 0.785], [0.04984848484848485, 0.0525, 0.1401875, 0.18921428571428572, 0.3954166666666667, 0.5971, 0.65275, 0.6928333333333333, 0.73275, 0.781], [0.04994949494949495, 0.04922222222222222, 0.1163125, 0.15478571428571428, 0.3175, 0.5694, 0.642, 0.6858333333333333, 0.7275, 0.779], [0.0498989898989899, 0.04822222222222222, 0.103625, 0.1457857142857143, 0.2325, 0.5126, 0.6365, 0.6818333333333333, 0.7215, 0.7775], [0.04984848484848485, 0.04783333333333333, 0.1004375, 0.1372142857142857, 0.19291666666666665, 0.4538, 0.630375, 0.671, 0.719, 0.7745], [0.04984848484848485, 0.0475, 0.0976875, 0.1315, 0.17291666666666666, 0.4014, 0.62575, 0.6698333333333333, 0.7165, 0.772], [0.04984848484848485, 0.04755555555555555, 0.09425, 0.1245, 0.1675, 0.3739, 0.621875, 0.6668333333333333, 0.71125, 0.7695], [0.04984848484848485, 0.04777777777777778, 0.0939375, 0.11414285714285714, 0.1585, 0.3487, 0.61725, 0.6655, 0.71, 0.7695], [0.04984848484848485, 0.04777777777777778, 0.09225, 0.10528571428571429, 0.15366666666666667, 0.3304, 0.61175, 0.6615, 0.71, 0.7665]]

数据集2

[[0.6240379891927297, 0.736448766432559, 0.7512155591572123, 0.7913868951146099, 0.8080767152890329, 0.8213938411669368, 0.8344813614262561, 0.8522420313344138, 0.8322528363047001, 0.8379254457050244], [0.4954150974291796, 0.7351881865658203, 0.7611426256077796, 0.7962491317434591, 0.8107779578606159, 0.8256077795786062, 0.8401539708265803, 0.8587250135062129, 0.8435980551053485, 0.8379254457050244], [0.6125757327656788, 0.7328471096704484, 0.7585089141004863, 0.8027321139152581, 0.8165856293895192, 0.8283630470016208, 0.8440032414910859, 0.8587250135062129, 0.8448136142625607, 0.8435980551053485], [0.5898968396921566, 0.7385197190707725, 0.7546596434359806, 0.803542486686733, 0.8183414370610481, 0.8317666126418153, 0.8446110210696921, 0.8600756347920043, 0.8415721231766613, 0.8517017828200972], [0.5311937121336172, 0.7552674230145867, 0.7562803889789304, 0.8013429034498727, 0.821042679632631, 0.8327390599675851, 0.8442058346839546, 0.8622366288492707, 0.8460291734197731, 0.8549432739059968], [0.5291468806287867, 0.7548172159193229, 0.7653970826580226, 0.8042370919194258, 0.8218530524041059, 0.8353322528363047, 0.8478525121555915, 0.8630470016207455, 0.8488654781199352, 0.8565640194489466], [0.5641886359914852, 0.75895912119575, 0.7624594813614263, 0.8077101180828895, 0.8246893571042679, 0.8380875202593193, 0.8508914100486223, 0.8633171258779039, 0.8573743922204214, 0.8589951377633711], [0.5683641722613394, 0.7647217720151269, 0.7607374392220422, 0.8107200740912248, 0.8264451647757969, 0.8418152350081037, 0.8517017828200972, 0.8654781199351702, 0.8594003241491086, 0.8614262560777958], [0.611920746684133, 0.7649018548532325, 0.7669165316045381, 0.8118777494790461, 0.8314424635332253, 0.847001620745543, 0.851904376012966, 0.8665586169638033, 0.8626418152350082, 0.8646677471636953], [0.61159325364336, 0.7738159553394561, 0.7802876823338736, 0.8129196573280852, 0.832117774176121, 0.8486223662884927, 0.8555510534846029, 0.8673689897352782, 0.8630470016207455, 0.8662884927066451], [0.5925167840183396, 0.7813794345398883, 0.8149311183144247, 0.8295901829127112, 0.8473797947055646, 0.8628849270664506, 0.8673014586709886, 0.8816855753646677, 0.8703403565640194, 0.86709886547812], [0.46487637137710824, 0.786871961102107, 0.8166531604538088, 0.8266959944431581, 0.853322528363047, 0.8696920583468395, 0.8723662884927067, 0.8862776877363587, 0.8784440842787682, 0.8792544570502431], [0.36556410676273127, 0.7762470736538808, 0.8263776337115073, 0.8306320907617504, 0.8549432739059968, 0.8669367909238249, 0.8689222042139384, 0.8876283090221502, 0.8808752025931929, 0.8865478119935171], [0.32937612575732766, 0.7366288492706645, 0.8174635332252836, 0.8345681870803426, 0.8531874662344678, 0.8674230145867099, 0.8762155591572123, 0.893030794165316, 0.8905996758508914, 0.8946515397082658], [0.2864745374160799, 0.695569962182604, 0.7853525121555915, 0.8240333410511692, 0.8534575904916262, 0.8698541329011346, 0.8741896272285251, 0.8908698001080497, 0.890194489465154, 0.8938411669367909], [0.2560176846242017, 0.6642355483522421, 0.7698541329011345, 0.8146561704098171, 0.8525121555915721, 0.8692058346839546, 0.8741896272285251, 0.8895191788222583, 0.8918152350081038, 0.8865478119935171], [0.23866055346323892, 0.6376733297316766, 0.7505064829821718, 0.8140773327159064, 0.8500810372771475, 0.8690437601296597, 0.8745948136142626, 0.8881685575364667, 0.8905996758508914, 0.8857374392220422], [0.2316194530866219, 0.5798667386998019, 0.733387358184765, 0.8107200740912248, 0.8480551053484603, 0.86709886547812, 0.873581847649919, 0.8876283090221502, 0.8910048622366289, 0.8849270664505673]]

[[0.23088259374488293, 0.5406086799927967, 0.7211304700162074, 0.8022690437601296, 0.8456239870340356, 0.8656401944894652, 0.8713533225283631, 0.8849270664505673, 0.8893841166936791, 0.8841166936790924], [0.22048468970034388, 0.3535026112011525, 0.4468192868719611, 0.7530678397777263, 0.8075364667747164, 0.8504051863857375, 0.8608184764991896, 0.873581847649919, 0.8768233387358185, 0.8768233387358185], [0.22171278860324217, 0.24356203853772734, 0.31006888168557534, 0.7079184996526974, 0.7755267423014587, 0.8385737439222042, 0.8549432739059968, 0.8657482441923284, 0.8630470016207455, 0.86709886547812], [0.2225315212051744, 0.20106248874482263, 0.22497974068071314, 0.6685575364667747, 0.7429767693138843, 0.8304700162074554, 0.8486628849270664, 0.8627768773635872, 0.859805510534846, 0.8654781199351702], [0.2225315212051744, 0.18683594453448588, 0.35848865478119935, 0.6235239638805279, 0.7285251215559158, 0.8205834683954619, 0.8421799027552674, 0.8565640194489466, 0.8585899513776337, 0.8622366288492707], [0.22122154904208285, 0.18071312803889789, 0.3247568881685575, 0.5772169483676777, 0.7150189086980011, 0.8111831442463533, 0.8403565640194489, 0.8503511615343058, 0.8589951377633711, 0.8630470016207455], [0.22122154904208285, 0.17585089141004862, 0.31432333873581847, 0.5377402176429729, 0.7028633171258779, 0.80032414910859, 0.836709886547812, 0.8490005402485143, 0.8581847649918962, 0.8614262560777958], [0.22122154904208285, 0.1655861696380335, 0.28292139384116693, 0.5123871266496874, 0.6935440302539168, 0.7949756888168558, 0.8338735818476499, 0.8476499189627229, 0.8537277147487844, 0.8614262560777958], [0.2209759292615032, 0.16612641815235007, 0.27380470016207453, 0.47302616346376475, 0.6843598055105349, 0.7896272285251216, 0.8294165316045381, 0.8452188006482982, 0.8525121555915721, 0.859805510534846]]

[[0.46494949494949495, 0.6412222222222222, 0.693125, 0.7276428571428571, 0.75325, 0.7642, 0.778875, 0.7991666666666667, 0.8165, 0.831], [0.5091414141414141, 0.6695, 0.722625, 0.7583571428571428, 0.7784166666666666, 0.7909, 0.80125, 0.8178333333333333, 0.83075, 0.842], [0.5195959595959596, 0.6887222222222222, 0.741375, 0.7709285714285714, 0.79175, 0.8049, 0.813625, 0.8258333333333333, 0.837, 0.8465], [0.4953030303030303, 0.7081666666666667, 0.75725, 0.7849285714285714, 0.804, 0.8118, 0.822125, 0.8336666666666667, 0.8425, 0.85], [0.4902525252525253, 0.7241666666666666, 0.7700625, 0.7946428571428571, 0.81025, 0.8192, 0.8285, 0.8385, 0.84425, 0.8515], [0.452979797979798, 0.7376666666666667, 0.781375, 0.7999285714285714, 0.81675, 0.8252, 0.832625, 0.842, 0.84725, 0.854], [0.40616161616161617, 0.7407777777777778, 0.7863125, 0.8048571428571428, 0.82025, 0.8285, 0.835625, 0.8443333333333334, 0.84975, 0.856], [0.3842929292929293, 0.7462777777777778, 0.793875, 0.807, 0.8234166666666667, 0.8323, 0.837125, 0.8466666666666667, 0.85225, 0.856], [0.3543939393939394, 0.7476666666666667, 0.798, 0.8122142857142857, 0.8251666666666667, 0.8356, 0.839625, 0.8483333333333334, 0.854, 0.857], [0.3301010101010101, 0.7529444444444444, 0.796875, 0.8167142857142857, 0.8303333333333334, 0.8371, 0.8415, 0.8505, 0.856, 0.8605], [0.1146969696969697, 0.7575555555555555, 0.8130625, 0.8317142857142857, 0.8456666666666667, 0.8506, 0.85225, 0.8576666666666667, 0.86275, 0.8665], [0.09722222222222222, 0.7189444444444445, 0.813625, 0.8366428571428571, 0.8476666666666667, 0.8576, 0.85725, 0.8616666666666667, 0.8635, 0.871], [0.09121212121212122, 0.6603333333333333, 0.807375, 0.8380714285714286, 0.8494166666666667, 0.8584, 0.859375, 0.8638333333333333, 0.8675, 0.8735], [0.08949494949494949, 0.5937222222222223, 0.7985625, 0.8372857142857143, 0.8494166666666667, 0.8587, 0.860625, 0.8645, 0.86875, 0.876], [0.08873737373737374, 0.5396111111111112, 0.78775, 0.8365714285714285, 0.85, 0.8574, 0.859875, 0.8651666666666666, 0.8685, 0.8745], [0.08787878787878788, 0.48488888888888887, 0.77925, 0.8318571428571429, 0.8489166666666667, 0.8571, 0.86125, 0.8653333333333333, 0.8685, 0.8745], [0.08712121212121213, 0.4478888888888889, 0.7705625, 0.8285, 0.8470833333333333, 0.8573, 0.86225, 0.8645, 0.8695, 0.873], [0.08681818181818182, 0.4007222222222222, 0.7575, 0.8262142857142857, 0.8475, 0.8573, 0.861375, 0.865, 0.87, 0.874]]

[[0.08646464646464647, 0.3616111111111111, 0.73825, 0.8221428571428572, 0.8461666666666666, 0.8556, 0.8605, 0.8656666666666667, 0.87025, 0.873], [0.0842929292929293, 0.2475, 0.5305625, 0.7892857142857143, 0.835, 0.8514, 0.857, 0.865, 0.86475, 0.871], [0.08267676767676768, 0.21266666666666667, 0.42925, 0.7370714285714286, 0.8271666666666667, 0.8452, 0.853, 0.8588333333333333, 0.86175, 0.8675], [0.08186868686868687, 0.18916666666666668, 0.374375, 0.6937142857142857, 0.81725, 0.8391, 0.851, 0.8576666666666667, 0.86, 0.865], [0.08222222222222222, 0.17444444444444446, 0.3360625, 0.6126428571428572, 0.8120833333333334, 0.8364, 0.848625, 0.8568333333333333, 0.8595, 0.8635], [0.08212121212121212, 0.15955555555555556, 0.30425, 0.5670714285714286, 0.8060833333333334, 0.8332, 0.846125, 0.8538333333333333, 0.858, 0.86], [0.08202020202020202, 0.1408888888888889, 0.2893125, 0.5177857142857143, 0.7989166666666667, 0.8277, 0.8435, 0.8523333333333334, 0.856, 0.859], [0.08202020202020202, 0.13516666666666666, 0.274875, 0.48392857142857143, 0.7913333333333333, 0.8245, 0.840125, 0.851, 0.85525, 0.859], [0.08176767676767677, 0.13166666666666665, 0.26775, 0.4635, 0.77975, 0.8228, 0.837875, 0.849, 0.8545, 0.8575]]

2

词典的大小V是动态的；

发现刚开始准确率低了不少，但后边准确率有所提升

[0.25116161616161614, 0.489, 0.545125, 0.5893571428571428, 0.6390833333333333, 0.6614, 0.68575, 0.7106666666666667, 0.73675, 0.7715]

[0.39323726870803993, 0.7451827840806771, 0.7757293354943274, 0.8197499421162306, 0.8371150729335495, 0.8578606158833063, 0.8620340356564019, 0.8752025931928687, 0.8695299837925445, 0.86709886547812]

[0.5391414141414141, 0.7866111111111111, 0.8175625, 0.8350714285714286, 0.8459166666666667, 0.8513, 0.85125, 0.8573333333333333, 0.8625, 0.8665]

0.1训练数据；不同beta；

3

self.n\_z[cluster] += wordFre \* prob

[0.39494949494949494, 0.5162222222222222, 0.5499375, 0.5920714285714286, 0.6361666666666667, 0.6513, 0.674625, 0.6983333333333334, 0.73, 0.76]

[0.4091206811855248, 0.5846389339095984, 0.684663695299838, 0.7534151423940727, 0.779443544030254, 0.8118314424635332, 0.8213128038897893, 0.839546191247974, 0.8354943273905997, 0.8500810372771475]

[0.7249494949494949, 0.7961111111111111, 0.8236875, 0.8361428571428572, 0.8465, 0.8491, 0.851125, 0.8575, 0.86125, 0.8695]

0.1训练数据；不同beta；

4

高概率document用于更新model

prob = 1 >= 0.8 or 0 < 0.8

[0.39585858585858585, 0.5161666666666667, 0.54975, 0.5919285714285715, 0.63625, 0.6514, 0.674625, 0.6983333333333334, 0.73, 0.76]

[0.4091206811855248, 0.5846389339095984, 0.684663695299838, 0.7534151423940727, 0.779443544030254, 0.8118314424635332, 0.8213128038897893, 0.839546191247974, 0.8354943273905997, 0.8500810372771475]

[0.7267171717171718, 0.7959444444444445, 0.8233125, 0.8355714285714285, 0.8464166666666667, 0.8492, 0.8505, 0.8573333333333333, 0.86075, 0.8695]

[[0.39585858585858585, 0.5161666666666667, 0.5498125, 0.592, 0.63625, 0.6514, 0.674625, 0.6983333333333334, 0.73, 0.76], [0.39585858585858585, 0.5161666666666667, 0.5498125, 0.5919285714285715, 0.63625, 0.6514, 0.674625, 0.6983333333333334, 0.73, 0.76], [0.39585858585858585, 0.5161666666666667, 0.5498125, 0.5919285714285715, 0.63625, 0.6514, 0.674625, 0.6983333333333334, 0.73, 0.76], [0.39585858585858585, 0.5161666666666667, 0.5498125, 0.5919285714285715, 0.63625, 0.6514, 0.674625, 0.6983333333333334, 0.73, 0.76], [0.39585858585858585, 0.5161666666666667, 0.54975, 0.5919285714285715, 0.63625, 0.6514, 0.674625, 0.6983333333333334, 0.73, 0.76], [0.39585858585858585, 0.5161666666666667, 0.54975, 0.5919285714285715, 0.63625, 0.6514, 0.674625, 0.6983333333333334, 0.73, 0.76], [0.39585858585858585, 0.5161666666666667, 0.54975, 0.5919285714285715, 0.63625, 0.6514, 0.674625, 0.6983333333333334, 0.73, 0.76], [0.39585858585858585, 0.5161666666666667, 0.5498125, 0.5919285714285715, 0.63625, 0.6514, 0.674625, 0.6983333333333334, 0.73, 0.76]]

[[0.4091206811855248, 0.5846389339095984, 0.684663695299838, 0.7534151423940727, 0.779443544030254, 0.8118314424635332, 0.8213128038897893, 0.839546191247974, 0.8354943273905997, 0.8500810372771475], [0.4091206811855248, 0.5846389339095984, 0.684663695299838, 0.7534151423940727, 0.779443544030254, 0.8118314424635332, 0.8213128038897893, 0.839546191247974, 0.8354943273905997, 0.8500810372771475], [0.4091206811855248, 0.5846389339095984, 0.684663695299838, 0.7534151423940727, 0.779443544030254, 0.8118314424635332, 0.8213128038897893, 0.839546191247974, 0.8354943273905997, 0.8500810372771475], [0.4091206811855248, 0.5846389339095984, 0.684663695299838, 0.7534151423940727, 0.779443544030254, 0.8118314424635332, 0.8213128038897893, 0.839546191247974, 0.8354943273905997, 0.8500810372771475], [0.4091206811855248, 0.5846389339095984, 0.684663695299838, 0.7534151423940727, 0.779443544030254, 0.8118314424635332, 0.8213128038897893, 0.839546191247974, 0.8354943273905997, 0.8500810372771475], [0.4091206811855248, 0.5846389339095984, 0.684663695299838, 0.7534151423940727, 0.779443544030254, 0.8118314424635332, 0.8213128038897893, 0.839546191247974, 0.8354943273905997, 0.8500810372771475], [0.4091206811855248, 0.5846389339095984, 0.684663695299838, 0.7534151423940727, 0.779443544030254, 0.8118314424635332, 0.8213128038897893, 0.839546191247974, 0.8354943273905997, 0.8500810372771475], [0.4091206811855248, 0.5846389339095984, 0.684663695299838, 0.7534151423940727, 0.779443544030254, 0.8118314424635332, 0.8213128038897893, 0.839546191247974, 0.8354943273905997, 0.8500810372771475]]

[[0.7245454545454545, 0.7956111111111112, 0.8229375, 0.8354285714285714, 0.8465833333333334, 0.8492, 0.8505, 0.8573333333333333, 0.86075, 0.8695], [0.7267171717171718, 0.7956666666666666, 0.8229375, 0.8355, 0.8464166666666667, 0.8491, 0.8505, 0.8573333333333333, 0.86075, 0.8695], [0.7267171717171718, 0.7957222222222222, 0.823125, 0.8355, 0.8464166666666667, 0.8491, 0.8505, 0.8573333333333333, 0.86075, 0.8695], [0.7267171717171718, 0.7958333333333333, 0.823125, 0.8355714285714285, 0.8464166666666667, 0.8491, 0.8505, 0.8573333333333333, 0.86075, 0.8695], [0.7267171717171718, 0.7958333333333333, 0.82325, 0.8355714285714285, 0.8464166666666667, 0.8492, 0.8505, 0.8573333333333333, 0.86075, 0.8695], [0.7267171717171718, 0.7958333333333333, 0.8233125, 0.8355714285714285, 0.8464166666666667, 0.8492, 0.8505, 0.8573333333333333, 0.86075, 0.8695], [0.7267171717171718, 0.7959444444444445, 0.8233125, 0.8355714285714285, 0.8464166666666667, 0.8492, 0.8505, 0.8573333333333333, 0.86075, 0.8695], [0.7267171717171718, 0.7959444444444445, 0.8233125, 0.8355714285714285, 0.8464166666666667, 0.8492, 0.8505, 0.8573333333333333, 0.86075, 0.8695]]