1、最近一段时间的主题词：

[[('美国', 0.015757361049780514),

('中国', 0.015538518953444954),

('美元', 0.0072920358929809744),

('中美', 0.0063468132648950484),

('市场', 0.0061285953739308755),

('经济', 0.005397729759952433),

('贸易', 0.0049612571433317754),

('不', 0.0043043434118460967),

('一个', 0.0041576622899701099),

('大豆', 0.0037216060797195705),

('指数', 0.0037178810962686815),

('中', 0.0036473362638356132),

('支撑', 0.003646128831097955),

('可能', 0.0035021321911048197),

('上', 0.0035020812554577918),

('特朗普', 0.0034286878930845396),

('还', 0.0034270617711108639),

('双方', 0.0033547225056038792),

('继续', 0.0032092167147697291),

('发展', 0.0032088691226110481)]]

2、输出同一个主题在所有时间段的主题词

[[('美国', 0.016870796394135416),

('中国', 0.014700239508391864),

('市场', 0.0065123649635400902),

('关税', 0.0055980998371738359),

('特朗普', 0.0050270279707832375),

('中', 0.0046840806101060214),

('上', 0.0043415215854782684),

('影响', 0.0043033902121516771),

('贸易', 0.0042656000296625749),

('企业', 0.0039602240651817489),

('大', 0.0037700184008655911),

('公司', 0.00369371505877991),

('经济', 0.003618105227966701),

('产品', 0.0035794931081734758),

('不', 0.0035417668548707454)],

[('中国', 0.01795567379235076),

('美国', 0.016781737692609541),

('市场', 0.0079110168589284612),

('贸易', 0.0062059850711466864),

('特朗普', 0.0054021792557372682),

('贸易战', 0.0048419908470340991),

('上', 0.0047251757278120657),

('经济', 0.0046520671672431492),

('关税', 0.0046520557215123997),

('影响', 0.0042672924214122989),

('可能', 0.0041259832446484106),

('美元', 0.0040869242990826558),

('中', 0.003833763539701053),

('不', 0.003707076367611833),

('中美', 0.003502404308804127)],

[('美国', 0.014586840901834508),

('中国', 0.0108005761902612),

('市场', 0.0073334118931425284),

('影响', 0.0044821073347458325),

('中兴', 0.0044346793062350429),

('上', 0.0043566926014799862),

('中', 0.0041741007803032908),

('企业', 0.004162619698951881),

('可能', 0.0037066767487296572),

('不', 0.0036268352765421925),

('贸易', 0.0035130013025501633),

('公司', 0.0034783145077380753),

('经济', 0.0034671504515396741),

('芯片', 0.0032034506302660243),

('价格', 0.0030907483108721279)],

[('美国', 0.012976161891509921),

('中国', 0.01281887986829644),

('市场', 0.0092079752040742802),

('大豆', 0.0071139956801850113),

('贸易', 0.0069054099188113251),

('中美', 0.0061730371280687797),

('万吨', 0.0059089956132681084),

('价格', 0.0049683940545118353),

('不', 0.0048131662634483836),

('可能', 0.0047083764223292274),

('上', 0.0043947689552097735),

('影响', 0.0042895524557165663),

('问题', 0.0040797432268800289),

('进口', 0.0040272632689752818),

('国内', 0.0038179103999028421)],

[('美国', 0.015757361049780514),

('中国', 0.015538518953444954),

('美元', 0.0072920358929809744),

('中美', 0.0063468132648950484),

('市场', 0.0061285953739308755),

('经济', 0.005397729759952433),

('贸易', 0.0049612571433317754),

('不', 0.0043043434118460967),

('一个', 0.0041576622899701099),

('大豆', 0.0037216060797195705),

('指数', 0.0037178810962686815),

('中', 0.0036473362638356132),

('支撑', 0.003646128831097955),

('可能', 0.0035021321911048197),

('上', 0.0035020812554577918)]]

3、找出在不同时间段波动较大的词（计算词的概率的标准差的字典）

这些词最能体现不同时间段的特点

[('中国', 0.0024280877417560145),

('美元', 0.0022340264199100447),

('中美', 0.0018073550769372702),

('万吨', 0.0017705105599599847),

('大豆', 0.0016580269480580411),

('关税', 0.0015634522357383389),

('美国', 0.0014653477399400403),

('特朗普', 0.0013785773481711089),

('价格', 0.0013169926801330299),

('贸易战', 0.0013007050460909998),

('贸易', 0.0012403281750907043),

('芯片', 0.0012004291770208837),

('指数', 0.0010930065885543223),

('市场', 0.0010896759928995473),

('库存', 0.0010568646080702914),

('影响', 0.0010071163164050933),

('双方', 0.00099031020330752825),

('黄金', 0.00098302402331542889),

('月份', 0.00097971253230122565),

('支撑', 0.00095401730338000573),

('产品', 0.0009256826355501064),

('吨', 0.00090935358107700271),

('公司', 0.00089610976153970279),

('国内', 0.00086766631778395845),

('问题', 0.0008552550596549719),

('基金', 0.00083626365128849803),

('一个', 0.00083609559554060926),

('还', 0.00082999190352402413),

('经济', 0.00081860051359189813),

('亿美元', 0.00080658488750959257),

('…', 0.00079373385919506451),

('我国', 0.00078815429924823102),

('谈判', 0.00077181452279265314),

('进口', 0.00076312456231926624),

('元', 0.00075414912920101238),

('企业', 0.00074122165346035725),

('未来', 0.00072878142867002858),

('豆粕', 0.00069174110426627894),

('磋商', 0.00068274539802602209),

('来说', 0.00066335607357337443),

('大', 0.00065970300324673262),

('现货', 0.00065604251012367433),

('加征', 0.0006503584463514485),

('人', 0.0006358358665194087),

('增加', 0.00062792666763864077),

('机构', 0.00061099380194812596),

('继续', 0.00060219821784653093),

('种植', 0.00059154654810339468),

('重要', 0.00058321138379444179),

('今日', 0.00057924858049345006)]

4、把波动比较大的前100个词在5个时间段的概率拎出来

见excel

5、以下是我总结的每个时间段重要的主题词

* 可以画出它们的概率随时间变化的折线图，每个时期选几个主题词画一张，反映事态发展（用excel画就行）
* 词和每个时间段的概率在[word\_pro.csv](http://localhost:8888/notebooks/%E5%BA%94%E7%94%A8%E7%BB%9F%E8%AE%A1%E6%A1%88%E4%BE%8B/case2018Lab9/word_pro.csv)

#第一时期

word\_pro["美国商务部"]

word\_pro["制裁"]

word\_pro["禁令"]

word\_pro["限制"]

word\_pro["禁止"]

word\_pro["商务部"]

word\_pro["出口"]

word\_pro["技术"]

word\_pro["设备"]

word\_pro['产品']

word\_pro['影响']

word\_pro["协议"]

word\_pro['基站']

word\_pro["高通"]

#第二时期

word\_pro['基金']

word\_pro['影响']

word\_pro['芯片']

word\_pro['估值']

word\_pro["制裁"]

word\_pro["行业"]

#第三时期

word\_pro['基金']

word\_pro['芯片']

word\_pro['技术']

word\_pro['产品']

word\_pro['产业']

word\_pro['华为']

word\_pro['厂商']

word\_pro['国产']

word\_pro['核心技术']

word\_pro["行业"]

#第四

word\_pro['经济']

word\_pro['发展']

word\_pro['产业']

word\_pro['制造']

word\_pro['全球']

word\_pro['出口']

word\_pro['投资']

#第五时期

word\_pro["特朗普"]

word\_pro['发展']

word\_pro['经济']

word\_pro['中美']

word\_pro['问题']

word\_pro['技术']

word\_pro["恢复"]