

jiedi

cut（给的地址）得到模型参数后

viterbi(地址字符串)

先得到各种初始状态的以及该初始状态发射到当前可观测状态的概率

接着遍历剩下的字符 对于每个字符都去遍历当前各个隐藏状态和到达前一个各个隐藏状态的概率的最终得到一个，然后得到当前各个隐藏状态对应的最大概率并更新相关参数

最后找到最后一个时刻的各个隐藏状态对应的最大概率，从而得到该隐藏状态

revise_address_cut
(切分后的地址字符串)

判断地址字符串的各个部分是否存在于样本对于属性列中，有则筛选出存在的，并进入下一个属性继续筛选，遍历完省市属性。剩余部分之间作为详细地址部分。如果存在空属性的部分并且非空的都是存在于筛选后的样本（出现在同一行的）并且空的属性小于最大非空的（按照 $P < C < A < D$ ）那么就从筛选后样本中找第一行的对应的属性补全

cal_trans_BE_BM_MM_ME（省或市或区或详细地址该列的所有名字）计算省市详细地址内的转移概率

根据每个名字的字符串长度来判断BME的个数进而判断概率B到M B到E M到M M到E的 M到M的概率

得到PB到PM PB到PE PM到PM PM到PE CB到CE CB到CM CM到CM CM到CE AB到AM AB到AE AM到AM AM到AE DB到DE DB到DM DM到DE DM到DM的转移概率

build_porb
训练得到隐马尔可夫模型参数

cal_trans_p_c_a_d（地址样本）计算省市详细地址之间的转移概率

根据行中prov列空不空 city列空不空 distract列空不空 detail列空不空结合得到PE到CB PE到AB PE到DB CE到AB CE到DB AE到DB的个数进而求概率

cal_start_p（地址样本）计算初始概率

根据行中prov列空不空 city列空不空 distract列空不空 detail列空不空结合得到为首的是prov 或city或distract或detail的个数进而得到概率

cal_emit_p（省或市或区或详细地址该列的所有名字）计算发射概率矩阵

遍历当前属性列所有字符得到BME各个位置对应的所有字符的各类个数从而求得BME发射到某个字符的概率