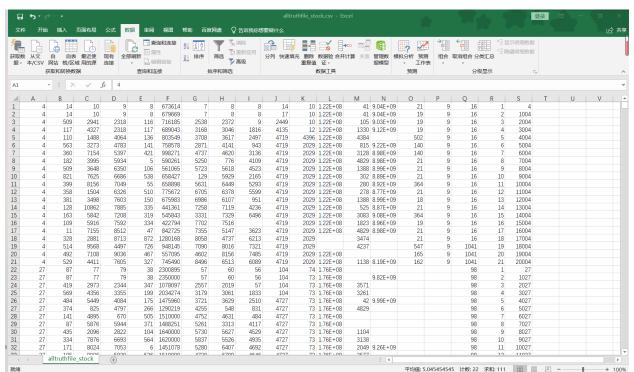
Every source: 21×1000 f sample 舒 sample 由 A: source B: sample 中一碱 G: volum (神然似)
S:天数
21天. 每天 1000支股票

现对 truth 恐A列科序: 如下图



现象可得对 A列(应是 Sample 序号)
YaeA, R列取值 1-21 (是整的), 形数
truth中由A列、F列, A列, 唯一可能一个sample
: 从data中可抽取的 Sample (那, 在truth 中

·从data中可抽取的 sample (那,在truta P) 可找到对应的, 应是上图中排序后. A中100个不同 sample (如4,27等)

一抽取策略城。

O首先从truth中(sof) Sample 序号确定M个样本(M=10)

②对原data 中S丁数据源(S=55),从bs∈S 抽取 VmeM. 得到新丽子样本的S条数据 PS;这里应保证每次排取是同一天(随机碗是~~)即同一天.同一支.不同源。对应了truth中唯一一条 设施条为七

m,; S条

③对Mn,其S条桐源数据和t,利用Jaccard 计第二条向量√

行约约,1 了,--- 了 Sin分为对应别比值与 七相同的数据源标号

@ embedding 相似度的较; 可利用 embedding 生成样本和t 6的 embedding

得到二者相似度,和一种的火气吸

医引机的 (<簸状>)