《汉书·艺文志》目录分类再审视*

诸雨辰 / 北京师范大学文学院
李　绅 / 北京师范大学中文信息处理研究所

摘　要：为考察《汉书·艺文志》文献分类的多元可能性，文章运用基于神经词向量的K-means++、Gaussian Mixture Model和Spectral Clustering模型，对《汉书·艺文志》中的存世文献进行自动聚类。聚类结果显示四分至六分的聚类较为稳定有效。聚类结果显示出基于言说方式或文本内容的两种基本分类原则，为超越“《汉志》主义”提供了更多可能。后人在四部分类法的框架下审视《汉志》的六分法，提出诸如“史部何以不立”等命题，亦显示出四部分类法在构建知识体系中潜移默化的影响。书目在古代文化中已超越单纯的分类目录，成为富有潜在影响力的思维方式。

关键词：《汉书·艺文志》《七略》 分类　自动聚类

汉代刘向、刘歆父子所撰《七略》是中国古代目录学的奠基之作，影响深远，虽然其书已亡佚，但后人却可根据《汉书·艺文志》（以下简称《汉志》）窥其大略。清代章学诚称赞道：“校雠之义，盖自刘向父子部次条别，将以辨章学术，考镜源流；非深明于道术精微、群言得失之故者，不足与此。”（《校雠通义·叙》）[①]后人沿此，也多从学术史、思想史角度称赞其价值，比如邓骏捷归纳《七略》“学术出于王官”的学术史观，从经、子之关系角度，论述了《七略》建构的学术体系。段恺、童庆炳也阐释了《汉志》建构儒学话语权威的思想史意义。另一方面，也有学者反思刘向父子的编目工作，认为无论《七略》还是《汉志》都以书籍目录为基本性质，缺乏系统、严密的学术观照。或者从早期文本生成与

*本文系国家社科基金重大项目“基于大数据技术的古代文学经典文本分析与研究”（18ZDA238）阶段性成果。
① 章学诚著，叶瑛校注：《文史通义校注》，北京：中华书局，2014年，第1101页。

流传的角度反思“《汉志》主义”对文本开放性的封闭，以及对周秦思想、学术与文学的建构与遮蔽，可谓是对学界一向认定的《汉志》“考镜源流”的反思。①

论述《七略》或《汉志》的思想史建构，其前提是对刘向父子既有文献分类的认定；而反思其学术性质与价值，则很大程度上来自对刘向父子分类的批判性考察。无论支持与否，都涉及长期以来围绕《汉志》形成的另一焦点——分类标准问题。章学诚就有对“四部之不能返《七略》者”②原因的总结，今人亦有对史籍何以未能独成部类③、诗赋略分类是否混杂④等问题的进一步阐释。不过，今人对周秦时代文本的认识，往往难以界定文本分类的边界性与有效性，而或多或少依赖于《汉志》本身的描述。诚如徐建委所说，《汉志》“几乎成了我们描述周秦汉学术、思想与文学的‘基础结构’”，从而使观察者“无意识”地以《汉志》“分类之既定事实，去逆推其分类标准”，⑤因而往往陷入《汉志》既有的框架而不自知。徐建委由此揭示出周秦汉学术研究中的“《汉志》主义”，并试图通过限制材料的适用度，比如划分出更为细致的书、篇、章等不同层次加以考察，从而实现对“《汉志》主义”的超越。

问题在于，当我们把一部文献拆分为篇章段落后，还能否说这些篇章段落是原文献本身？而且切分到什么程度也难免受学者的主观判断的影响，因而这种路径就容易再次陷入自我封闭的阐释之中。因此，若欲追问刘向父子整理目录的诸种可能性、追问不同分类的依据，进而反思《汉志》所呈现出的文本分类是否具有必然性时，仍然需要寻找更为客观的文本分类方法。而自然语言处理技术在文本语义信息方面的应用，正可以不带任何文化语境的预设，摹拟《汉志》的分类结果，从而为重审《汉志》文本分类之多元可能性提供新视角。

本文以《汉志》著录的78种传世文献为测试对象，《六艺略》包含《周易》《尚书》等24种，《诸子略》包含《子思子》《曾子》等31种，《诗赋略》包含《屈原赋》《宋玉赋》等17种，《兵书略》包含《孙子兵法》《吴子兵法》《尉缭子》3种，《数术略》仅存《山海经》1种，《方技略》包含《黄帝内经》《难经》2种。虽然存世文献的类型与数量仅为《汉志》著录文献的一部分，但我们可以把千年以来文献的保存与否视为随机抽样的结果。因此，以传世文献作为本文分类问题的研究对

---

①参见邓骏捷：《“诸子出于王官”说与汉家学术话语》，《中国社会科学》2017年第9期；段恺、童庆炳：《论〈汉书·艺文志〉的儒学话语建构》，《山西大学学报》（哲学社会科学版）2014年第2期；杨新宾：《目录学与学术史之间——〈汉书·艺文志〉价值的再思考》，《理论月刊》2012年第6期；徐建委：《周秦汉文学研究中的〈汉志〉主义及其超越》，《文学遗产》2017年第2期。

②章学诚著，叶瑛校注：《文史通义校注》，第1114页。

③张涛：《〈七略〉中史籍未能独成部类的根本原因》，《文史哲》1992年第6期。

④陈刚：《〈汉书·艺文志·诗赋略〉赋之分类研究述略》，《文献》2011年第2期。

⑤徐建委：《周秦汉文学研究中的〈汉志〉主义及其超越》，《文学遗产》2017年第2期。

象，依然能保持统计抽样的合理性。

一、《汉志》文本分类的可能性

计算语言学通过计算文本的相似性实现对文本语义信息的理解与加工。基于Harris的词义分布假说[①]——上下文相似的词，其意义也相近，人们建立了词向量模型。在进一步改进算法基础上，Mikolov等提出了神经词向量（Neural Word Embeddings）模型，[②]在大规模语料库中，依次选取中心词及其上下文的临近词，通过神经网络构建中心词与上下文关系的相互预测模型，从而更有效地提升文本特征表示的效果。而通过词向量模型的语义表征及语义计算，也就实现了诸如文本分类、文本聚类、回归预测等语义理解任务。

考虑到古汉语和现代汉语的语义和语法有显著的区别，本研究使用了Shen Li等人基于《四库全书》训练的古汉语神经词向量模型。[③]首先对每篇文本中的字查找其在向量模型中的300维向量表示，然后对每篇文本各个字向量取平均得到文本的300维向量表示。具体来说，给定长度为n的文本$d=(w_1,w_2,\cdots,w_i,\cdots,w_{n-1},w_n)$，其中$w_i$是组成文本的字。在预训练的字向量中查找到每个$w_i$对应的向量$e_i$（维度是300），则文本d的向量表示$e_d=\frac{\sum_{i=1}^{i=n}e_i}{n}$。之后使用无监督的自动聚类模型，对相关文本进行聚类。聚类模型的基本思路是初始化K个中心样本，其余文本围绕与其最近的中心样本形成K个文本簇，经过反复迭代调整中心样本的位置，最终形成相对稳定的文本簇，完成对文本的自动聚类。由于不同聚类模型在初始化方式与原理上不尽相同，一般聚类的结果也不一样，本研究采用K-means++[④]、Gaussian Mixture Model[⑤]和Spectral Clustering[⑥]三种聚类模型，对文本自动聚类结果进行综合判断，依次取K=2至K=8，将上述78种文献自动聚为2至8类，观察基于文本相似性而形成的文本分类结果。其中，为了降低聚类的随机性，本文

①Zellig Harris, "Distributional Structure," in *Word*, vol. 10, 1954, pp. 146-162.

②Tomas Mikolov et al., "Distributed Representations of Words and Phrases and their Compositionality," Proceedings of NIPS, Lake Tahoe, USA: Neural Information Processing Systems Foundation, 2013, pp. 3111-3119.

③Shen Li et al., "Analogical Reasoning on Chinese Morphological and Semantic Relations," Proceedings of ACL, Melbourne, Australia: Association for Computational Linguistics, 2018, pp. 138-143.

④David Arthur, Sergei Vassilvitskii, "k-means++: The Advantages of Careful Seeding," in Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2007, pp. 1027-1035.

⑤Douglas Reynolds, "Gaussian Mixture Models," *Encyclopedia of Biometrics*, vol. 741, 2009, pp. 659-663.

⑥Stella X. Yu, Jianbo Shi, "Multiclass Spectral Clustering," Proceedings Ninth IEEE International Conference on Computer Vision, Nice, France, 2003, pp. 313-319.

选用的是K-means[①]的改进算法K-means++，较K-means而言，其可以根据数据计算出稳定的初始聚类中心点，同时为了充分收敛，最大迭代次数设置为500。Gaussian Mixture Model设置为每一个类使用独立的协方差矩阵，其中的EM迭代阈值设置为1e-3，最大迭代100次。Spectral Clustering采用了和聚类数量相等的本征值向量个数，并用RBF核构建相似度矩阵（affinity matrix）。此外，为了避免单次实验带来的偶然性，所有聚类实验均进行了20次，选择其中收敛程度最高的一组作为最终结果。

表1　K-means++二分法的自动聚类结果

1	周易 / 京氏易传 / 尚书 / 尚书大传 / 诗经 / 韩诗外传 / 毛诗故训传 / 周礼 / 礼记 / 大戴礼记 / 仪礼 / 司马法 / 左传 / 公羊传 / 穀梁传 / 国语 / 战国策 / 史记 / 论语 / 孝经 / 尔雅 / 小尔雅 / 急就篇 / 方言 / 晏子 / 子思子 / 曾子 / 孟子 / 荀子 / 新语 / 新书 / 春秋繁露 / 盐铁论 / 新序 / 说苑 / 列女传 / 太玄 / 法言 / 管子 / 老子 / 文子 / 庄子 / 列子 / 鹖冠子 / 商子 / 慎子 / 韩非子 / 邓析子 / 尹文子 / 公孙龙子 / 墨子 / 鬼谷子 / 越绝书 / 吕氏春秋 / 淮南子 / 孔臧赋 / 司马迁赋 / 荀卿赋 / 孙子兵法 / 吴子兵法 / 尉缭子 / 山海经 / 黄帝内经 / 难经
2	楚辞 / 宋玉赋 / 贾谊赋 / 枚乘赋 / 司马相如赋 / 淮南王赋 / 淮南小山赋 / 汉武帝赋 / 刘向赋 / 王褒赋 / 扬雄赋 / 汉高祖诗 / 出行巡狩歌 / 郊祀歌

表2　Gaussian Mixture Model二分法的自动聚类结果

1	周易 / 京氏易传 / 尚书 / 尚书大传 / 韩诗外传 / 毛诗故训传 / 周礼 / 礼记 / 大戴礼记 / 仪礼 / 司马法 / 左传 / 公羊传 / 穀梁传 / 国语 / 战国策 / 史记 / 论语 / 孝经 / 小尔雅 / 急就篇 / 方言 / 晏子 / 子思子 / 曾子 / 孟子 / 荀子 / 新语 / 新书 / 春秋繁露 / 盐铁论 / 新序 / 说苑 / 列女传 / 太玄 / 法言 / 管子 / 老子 / 文子 / 庄子 / 列子 / 鹖冠子 / 商子 / 慎子 / 韩非子 / 邓析子 / 尹文子 / 公孙龙子 / 墨子 / 鬼谷子 / 越绝书 / 吕氏春秋 / 淮南子 / 孔臧赋 / 司马迁赋 / 荀卿赋 / 孙子兵法 / 吴子兵法 / 尉缭子 / 黄帝内经 / 难经
2	诗经 / 尔雅 / 楚辞 / 宋玉赋 / 贾谊赋 / 枚乘赋 / 司马相如赋 / 淮南王赋 / 淮南小山赋 / 汉武帝赋 / 刘向赋 / 王褒赋 / 扬雄赋 / 汉高祖诗 / 出行巡狩歌 / 郊祀歌 / 山海经

表3　Spectral Clustering二分法的自动聚类结果

1	周易 / 京氏易传 / 尚书 / 尚书大传 / 诗经 / 韩诗外传 / 毛诗故训传 / 周礼 / 礼记 / 大戴礼记 / 仪礼 / 司马法 / 左传 / 公羊传 / 穀梁传 / 国语 / 战国策 / 史记 / 论语 / 孝经 / 尔雅 / 小尔雅 / 急就篇 / 方言 / 晏子 / 子思子 / 曾子 / 孟子 / 荀子 / 新语 / 新书 / 春秋繁露 / 盐铁论 / 新序 / 说苑 / 列女传 / 太玄 / 法言 / 管子 / 老子 / 文子 / 庄子 / 列子 / 鹖冠子 / 商子 / 慎子 / 韩非子 / 邓析子 / 尹文子 / 公孙龙子 / 墨子 / 鬼谷子 / 越绝书 / 吕氏春秋 / 淮南子 / 荀卿赋 / 宋玉赋 / 枚乘赋 / 司马相如赋 / 淮南王赋 / 孔臧赋 / 司马迁赋 / 扬雄赋 / 郊祀歌 / 孙子兵法 / 吴子兵法 / 尉缭子 / 山海经 / 黄帝内经 / 难经
2	楚辞 / 贾谊赋 / 淮南小山赋 / 汉武帝赋 / 刘向赋 / 王褒赋 / 汉高祖诗 / 出行巡狩歌

①James MacQueen, “Some Methods for Classification and Analysis of Multivariate Observations,” Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, 1967, pp. 281-297.

仅分两类时（表1、表2、表3），三种模型的聚类结果都显现出学术性与文学性的差异，辞赋类文献往往单独聚为一类。其中K-means++模型的结果相对更好，而Gaussian Mixture Model混入了非辞赋类的《尔雅》《山海经》，Spectral Clustering则把部分辞赋混入了学术类。

表4　K-means++三分法的自动聚类结果

1	周易 / 韩诗外传 / 礼记 / 大戴礼记 / 司马法 / 左传 / 国语 / 战国策 / 论语 / 孝经 / 晏子 / 子思子 / 曾子 / 孟子 / 荀子 / 新语 / 新书 / 春秋繁露 / 盐铁论 / 新序 / 说苑 / 列女传 / 法言 / 管子 / 老子 / 文子 / 庄子 / 列子 / 鹖冠子 / 商子 / 慎子 / 韩非子 / 邓析子 / 尹文子 / 公孙龙子 / 墨子 / 鬼谷子 / 越绝书 / 吕氏春秋 / 淮南子 / 司马迁赋 / 荀卿赋 / 孙子兵法 / 吴子兵法 / 尉缭子
2	京氏易传 / 尚书 / 尚书大传 / 诗经 / 毛诗故训传 / 周礼 / 仪礼 / 公羊传 / 穀梁传 / 史记 / 尔雅 / 小尔雅 / 急就篇 / 方言 / 太玄 / 孔臧赋 / 枚乘赋 / 司马相如赋 / 淮南王赋 / 扬雄赋 / 郊祀歌 / 山海经 / 黄帝内经 / 难经
3	楚辞 / 宋玉赋 / 贾谊赋 / 淮南小山赋 / 汉武帝赋 / 刘向赋 / 王褒赋 / 汉高祖诗 / 出行巡狩歌

表5　Gaussian Mixture Model三分法的自动聚类结果

1	韩诗外传 / 礼记 / 大戴礼记 / 司马法 / 国语 / 战国策 / 论语 / 孝经 / 晏子 / 子思子 / 曾子 / 孟子 / 荀子 / 新语 / 新书 / 春秋繁露 / 盐铁论 / 新序 / 说苑 / 列女传 / 法言 / 管子 / 老子 / 文子 / 庄子 / 列子 / 鹖冠子 / 商子 / 慎子 / 韩非子 / 邓析子 / 尹文子 / 公孙龙子 / 墨子 / 鬼谷子 / 吕氏春秋 / 淮南子 / 司马迁赋 / 荀卿赋 / 孙子兵法 / 吴子兵法 / 尉缭子
2	周易 / 京氏易传 / 尚书 / 尚书大传 / 诗经 / 毛诗故训传 / 周礼 / 仪礼 / 左传 / 公羊传 / 穀梁传 / 史记 / 越绝书 / 尔雅 / 小尔雅 / 急就篇 / 方言 / 太玄 / 孔臧赋 / 山海经 / 黄帝内经 / 难经
3	楚辞 / 宋玉赋 / 贾谊赋 / 枚乘赋 / 汉武帝赋 / 淮南王赋 / 淮南小山赋 / 王褒赋 / 司马相如赋 / 刘向赋 / 扬雄赋 / 汉高祖诗 / 郊祀歌 / 出行巡狩歌

表6　Spectral Clustering三分法的自动聚类结果

1	韩诗外传 / 礼记 / 大戴礼记 / 司马法 / 国语 / 战国策 / 论语 / 孝经 / 晏子 / 子思子 / 曾子 / 孟子 / 荀子 / 新语 / 新书 / 春秋繁露 / 盐铁论 / 新序 / 说苑 / 列女传 / 法言 / 管子 / 老子 / 文子 / 庄子 / 列子 / 鹖冠子 / 商子 / 慎子 / 韩非子 / 邓析子 / 尹文子 / 公孙龙子 / 墨子 / 鬼谷子 / 吕氏春秋 / 淮南子 / 司马迁赋 / 荀卿赋 / 孙子兵法 / 吴子兵法 / 尉缭子
2	周易 / 京氏易传 / 尚书 / 尚书大传 / 诗经 / 毛诗故训传 / 周礼 / 仪礼 / 左传 / 公羊传 / 穀梁传 / 史记 / 越绝书 / 尔雅 / 小尔雅 / 急就篇 / 方言 / 太玄 / 宋玉赋 / 孔臧赋 / 枚乘赋 / 司马相如赋 / 淮南王赋 / 扬雄赋 / 汉高祖诗 / 郊祀歌 / 山海经 / 黄帝内经 / 难经
3	楚辞 / 贾谊赋 / 淮南小山赋 / 汉武帝赋 / 刘向赋 / 王褒赋 / 出行巡狩歌

在三分法的结果中（表4、表5、表6），三种模型析出的类目2都比较混乱，出现了经学文本（如《易传》等）、史书（《史记》等）、字书（《尔雅》等）、方术书（《黄帝内经》等）与辞赋（《孔臧赋等》）混杂的现象。相对而言，

Gaussian Mixture Model混入的辞赋较少，只有《孔臧赋》一种，但其他文献混杂的现象仍然较为明显。

表 7　K-means++ 四分法的自动聚类结果

1	周易 / 韩诗外传 / 礼记 / 大戴礼记 / 司马法 / 国语 / 论语 / 孝经 / 晏子 / 子思子 / 曾子 / 孟子 / 荀子 / 新语 / 新书 / 春秋繁露 / 盐铁论 / 新序 / 说苑 / 法言 / 管子 / 老子 / 文子 / 庄子 / 列子 / 鹖冠子 / 商子 / 慎子 / 韩非子 / 邓析子 / 尹文子 / 公孙龙子 / 墨子 / 鬼谷子 / 吕氏春秋 / 淮南子 / 司马迁赋 / 荀卿赋 / 孙子兵法 / 吴子兵法 / 尉缭子
2	京氏易传 / 诗经 / 毛诗故训传 / 尔雅 / 小尔雅 / 急就篇 / 方言 / 太玄 / 孔臧赋 / 枚乘赋 / 司马相如赋 / 淮南王赋 / 扬雄赋 / 郊祀歌 / 山海经 / 黄帝内经 / 难经
3	尚书 / 尚书大传 / 周礼 / 仪礼 / 左传 / 公羊传 / 穀梁传 / 战国策 / 史记 / 列女传 / 越绝书
4	楚辞 / 宋玉赋 / 贾谊赋 / 淮南小山赋 / 汉武帝赋 / 刘向赋 / 王褒赋 / 汉高祖诗 / 出行巡狩歌

表 8　Gaussian Mixture Model 四分法的自动聚类结果

1	周易 / 司马法 / 荀子 / 新语 / 新书 / 春秋繁露 / 管子 / 老子 / 文子 / 庄子 / 列子 / 鹖冠子 / 商子 / 慎子 / 韩非子 / 邓析子 / 尹文子 / 公孙龙子 / 墨子 / 鬼谷子 / 吕氏春秋 / 淮南子 / 司马迁赋 / 荀卿赋 / 孙子兵法 / 吴子兵法 / 尉缭子
2	京氏易传 / 诗经 / 仪礼 / 尔雅 / 小尔雅 / 急就篇 / 方言 / 太玄 / 孔臧赋 / 枚乘赋 / 淮南王赋 / 司马相如赋 / 扬雄赋 / 郊祀歌 / 山海经 / 黄帝内经 / 难经
3	尚书 / 尚书大传 / 毛诗故训传 / 韩诗外传 / 周礼 / 礼记 / 大戴礼记 / 左传 / 公羊传 / 穀梁传 / 国语 / 论语 / 孝经 / 孟子 / 盐铁论 / 曾子 / 子思子 / 晏子 / 新序 / 法言 / 战国策 / 史记 / 越绝书 / 列女传 / 说苑
4	楚辞 / 宋玉赋 / 贾谊赋 / 汉武帝赋 / 淮南小山赋 / 王褒赋 / 刘向赋 / 汉高祖诗 / 出行巡狩歌

表 9　Spectral Clustering 四分法的自动聚类结果

1	周易 / 韩诗外传 / 礼记 / 大戴礼记 / 司马法 / 国语 / 战国策 / 论语 / 孝经 / 晏子 / 子思子 / 曾子 / 孟子 / 荀子 / 新语 / 新书 / 春秋繁露 / 盐铁论 / 新序 / 说苑 / 列女传 / 法言 / 管子 / 老子 / 文子 / 庄子 / 列子 / 鹖冠子 / 商子 / 慎子 / 韩非子 / 邓析子 / 尹文子 / 公孙龙子 / 墨子 / 鬼谷子 / 吕氏春秋 / 淮南子 / 司马迁赋 / 荀卿赋 / 孙子兵法 / 吴子兵法 / 尉缭子
2	诗经 / 毛诗故训传 / 尔雅 / 小尔雅 / 急就篇 / 方言 / 太玄 / 宋玉赋 / 孔臧赋 / 枚乘赋 / 司马相如赋 / 淮南王赋 / 扬雄赋 / 郊祀歌 / 山海经 / 黄帝内经 / 难经
3	京氏易传 / 尚书 / 尚书大传 / 周礼 / 仪礼 / 左传 / 公羊传 / 穀梁传 / 史记 / 越绝书
4	楚辞 / 贾谊赋 / 汉武帝赋 / 刘向赋 / 王褒赋 / 汉高祖诗 / 出行巡狩歌
5	淮南小山赋

将文本簇增加到四个时（表7、表8、表9），K-means++和Spectral Clustering

都明显把《春秋》三传、《史记》等史籍类析出为新的类目3。[①]而在史籍类析出后，原本经史混杂的类目2就呈现出以《尔雅》《小尔雅》等小学书籍与《司马相如赋》《扬雄赋》等汉大赋并列的格局。汉大赋因为喜好排比名物，尤其是排比相同偏旁的汉字，其文本特征相较于《楚辞》《宋玉赋》等文学性更强的骚赋而言，反而更接近于小学类的字书，聚类结果符合此类文献的文本特征。而类目1的学术类文献，在四分法时也更集中于诸子一类。Gaussian Mixture Model的分类效果仍不理想，史籍的类目中依然混有《礼记》《论语》《孟子》《毛诗故训传》《孝经》等。

表 10　K-means++ 五分法的自动聚类结果

1	韩诗外传 / 礼记 / 大戴礼记 / 司马法 / 国语 / 战国策 / 论语 / 孝经 / 晏子 / 子思子 / 曾子 / 孟子 / 荀子 / 新语 / 新书 / 春秋繁露 / 盐铁论 / 新序 / 说苑 / 列女传 / 法言 / 管子 / 老子 / 文子 / 庄子 / 列子 / 鹖冠子 / 商子 / 慎子 / 韩非子 / 邓析子 / 尹文子 / 公孙龙子 / 墨子 / 鬼谷子 / 吕氏春秋 / 淮南子 / 司马迁赋 / 荀卿赋 / 孙子兵法 / 吴子兵法 / 尉缭子
2	周易 / 京氏易传 / 毛诗故训传 / 尔雅 / 小尔雅 / 急就篇 / 方言 / 太玄 / 黄帝内经 / 难经
3	尚书 / 尚书大传 / 周礼 / 仪礼 / 左传 / 公羊传 / 穀梁传 / 史记 / 越绝书
4	诗经 / 宋玉赋 / 孔臧赋 / 枚乘赋 / 司马相如赋 / 淮南王赋 / 扬雄赋 / 郊祀歌 / 山海经
5	楚辞 / 贾谊赋 / 淮南小山赋 / 汉武帝赋 / 刘向赋 / 王褒赋 / 出行巡狩歌
6	汉高祖诗

表 11　Gaussian Mixture Model 五分法的自动聚类结果

1	司马法 / 荀子 / 新语 / 新书 / 春秋繁露 / 管子 / 老子 / 文子 / 庄子 / 鹖冠子 / 商子 / 慎子 / 韩非子 / 邓析子 / 尹文子 / 公孙龙子 / 墨子 / 鬼谷子 / 淮南子 / 司马迁赋 / 荀卿赋 / 孙子兵法 / 吴子兵法 / 尉缭子
2	韩诗外传 / 礼记 / 大戴礼记 / 左传 / 国语 / 论语 / 孝经 / 孟子 / 盐铁论 / 曾子 / 子思子 / 晏子 / 列子 / 吕氏春秋 / 新序 / 法言 / 战国策 / 列女传 / 说苑
3	周易 / 京氏易传 / 尚书 / 尚书大传 / 周礼 / 仪礼 / 诗经 / 毛诗故训传 / 公羊传 / 穀梁传 / 史记 / 尔雅 / 小尔雅 / 急就篇 / 方言 / 太玄 / 孔臧赋 / 越绝书 / 山海经 / 黄帝内经 / 难经
4	宋玉赋 / 枚乘赋 / 淮南王赋 / 司马相如赋 / 扬雄赋 / 汉高祖诗 / 郊祀歌
5	楚辞 / 贾谊赋 / 汉武帝赋 / 淮南小山赋 / 王褒赋 / 刘向赋 / 出行巡狩歌

① 由于K取4或5时，Spectral Clustering会把《淮南小山赋》单独聚为一类，文本量过小，很难说形成的类有何意义，因此对四分法的考察需要K值取5。

表 12　Spectral Clustering 五分法的自动聚类结果

1	周易 / 韩诗外传 / 毛诗故训传 / 礼记 / 大戴礼记 / 司马法 / 国语 / 战国策 / 论语 / 孝经 / 晏子 / 子思子 / 曾子 / 孟子 / 荀子 / 新语 / 新书 / 春秋繁露 / 盐铁论 / 新序 / 说苑 / 列女传 / 法言 / 管子 / 老子 / 文子 / 庄子 / 列子 / 鹖冠子 / 商子 / 慎子 / 韩非子 / 邓析子 / 尹文子 / 公孙龙子 / 墨子 / 鬼谷子 / 吕氏春秋 / 淮南子 / 越绝书 / 司马迁赋 / 荀卿赋 / 孙子兵法 / 吴子兵法 / 尉缭子
2	京氏易传 / 尚书大传 / 周礼 / 左传 / 公羊传 / 穀梁传 / 史记 / 太玄 / 山海经 / 黄帝内经 / 难经
3	尔雅 / 小尔雅 / 急就篇 / 方言
4	尚书 / 诗经 / 仪礼 / 宋玉赋 / 孔臧赋 / 枚乘赋 / 司马相如赋 / 淮南王赋 / 扬雄赋 / 汉高祖诗 / 郊祀歌
5	楚辞 / 贾谊赋 / 汉武帝赋 / 刘向赋 / 王褒赋 / 出行巡狩歌
6	淮南小山赋

类似Spectral Clustering出现《淮南小山赋》单独聚类的现象，K-means++的K值取5时，也只把《汉高祖诗》析出，因此五分法时K-means++和Spectral Clustering的K值需要取到6（表10、表11、表12）。两种模型都在四分法时形成相对稳定的结果，且大体一致，有理由相信对于早期文献来说，四分法确是一种较为稳定的分类方式，尤其是在处理言说特征方面具有内在的合理性。而在五分法下，K-means++将《周易》从诸子文献中调整到经书类，使得诸子类文献的立言成说性质更加明显。同时，《司马相如赋》《扬雄赋》等从经书类中分出，成为新的部类（《诗经》也被调整到这一文学性更突出的部类），经书类文献便集中于经传训诂与小学方面。Spectral Clustering的调整是把《尔雅》等小学文献析出为一类，这种聚类明显是以内容为依据的结果。类似地，Gaussian Mixture Model的五分法中出现了类目2这批较为集中地体现儒家文化观念的文献，比如《左传》虽系史籍，但其中有很多“君子曰”的体现德性的内容。可见，Gaussian Mixture Model和Spectral Clustering都开始出现基于文本内容而非言说方式的聚类结果，当然这也造成了从言说方式上看，经、史、子再度杂糅的现象。

六分法框架下，三种模型都把《尔雅》等四种小学文献独立为一类（表13、表14、表15）。而类目2的经书类在分出了小学文献后，三种模型的聚类结果都或多或少出现了与史籍、方技书籍相混的现象。值得注意的是，既Gaussian Mixture Model之后，K-means++和Spectral Clustering也出现了类目3这样集中反映儒家重德观念的文献聚类，三种聚类模型的结果再次趋同。（Spectral Clustering稍逊一筹，其中混入了《庄子》等道家典籍。）而三种模型中，汉大赋与骚赋的聚类结果在整体上也没有太大变化。综括言之，在聚类方式上形成了儒家文献、小学文献、汉大赋、骚赋的文本区分，显示出以文本内容作为聚类依据的可能性。

表 13　K-means++ 六分法的自动聚类结果

1	司马法 / 荀子 / 新语 / 新书 / 春秋繁露 / 管子 / 老子 / 文子 / 庄子 / 列子 / 鹖冠子 / 商子 / 慎子 / 韩非子 / 邓析子 / 尹文子 / 公孙龙子 / 墨子 / 鬼谷子 / 吕氏春秋 / 淮南子 / 司马迁赋 / 荀卿赋 / 孙子兵法 / 吴子兵法 / 尉缭子
2	周易 / 京氏易传 / 尚书 / 尚书大传 / 周礼 / 仪礼 / 毛诗故训传 / 公羊传 / 穀梁传 / 史记 / 越绝书 / 太玄 / 黄帝内经 / 难经
3	韩诗外传 / 礼记 / 大戴礼记 / 左传 / 国语 / 战国策 / 论语 / 孝经 / 晏子 / 子思子 / 曾子 / 孟子 / 盐铁论 / 法言 / 新序 / 说苑 / 列女传
4	诗经 / 宋玉赋 / 孔臧赋 / 枚乘赋 / 司马相如赋 / 淮南王赋 / 扬雄赋 / 郊祀歌 / 山海经
5	楚辞 / 贾谊赋 / 淮南小山赋 / 汉武帝赋 / 刘向赋 / 王褒赋 / 出行巡狩歌
6	尔雅 / 小尔雅 / 急就篇 / 方言
7	汉高祖诗

表 14　Gaussian Mixture Model 六分法的自动聚类结果

1	司马法 / 荀子 / 新语 / 新书 / 春秋繁露 / 管子 / 老子 / 文子 / 庄子 / 鹖冠子 / 商子 / 慎子 / 韩非子 / 邓析子 / 尹文子 / 公孙龙子 / 墨子 / 鬼谷子 / 列子 / 吕氏春秋 / 淮南子 / 司马迁赋 / 荀卿赋 / 孙子兵法 / 吴子兵法 / 尉缭子
2	周易 / 京氏易传 / 尚书 / 尚书大传 / 周礼 / 仪礼 / 毛诗故训传 / 公羊传 / 穀梁传 / 史记 / 太玄 / 越绝书 / 黄帝内经 / 难经
3	韩诗外传 / 礼记 / 大戴礼记 / 左传 / 国语 / 论语 / 孝经 / 孟子 / 曾子 / 子思子 / 晏子 / 战国策 / 盐铁论 / 列女传 / 说苑 / 新序 / 法言
4	诗经 / 宋玉赋 / 孔臧赋 / 枚乘赋 / 淮南王赋 / 司马相如赋 / 扬雄赋 / 汉高祖诗 / 郊祀歌 / 山海经
5	楚辞 / 贾谊赋 / 汉武帝赋 / 淮南小山赋 / 王褒赋 / 刘向赋 / 出行巡狩歌
6	尔雅 / 小尔雅 / 急就篇 / 方言

表 15　Spectral Clustering 六分法的自动聚类结果

1	周易 / 尚书大传 / 周礼 / 大戴礼记 / 司马法 / 史记 / 战国策 / 鬼谷子 / 老子 / 韩非子 / 管子 / 荀子 / 淮南子 / 公孙龙子 / 文子 / 鹖冠子 / 墨子 / 慎子 / 邓析子 / 商子 / 尹文子 / 吕氏春秋 / 孙子兵法 / 吴子兵法 / 尉缭子 / 新语 / 新书 / 盐铁论 / 春秋繁露 / 太玄 / 越绝书 / 荀卿赋
2	京氏易传 / 公羊传 / 穀梁传 / 山海经 / 黄帝内经 / 难经
3	韩诗外传 / 毛诗故训传 / 礼记 / 左传 / 国语 / 论语 / 孝经 / 庄子 / 孟子 / 曾子 / 子思子 / 晏子 / 列子 / 法言 / 列女传 / 说苑 / 新序 / 司马迁赋
4	尚书 / 诗经 / 仪礼 / 宋玉赋 / 孔臧赋 / 枚乘赋 / 司马相如赋 / 淮南王赋 / 扬雄赋 / 汉高祖诗 / 郊祀歌
5	楚辞 / 贾谊赋 / 汉武帝赋 / 刘向赋 / 王褒赋 / 出行巡狩歌
6	尔雅 / 小尔雅 / 急就篇 / 方言
7	淮南小山赋

进一步增加聚类项时，K-means++也将《淮南小山赋》独立成类，Gaussian Mixture Model则把《黄帝内经》和《难经》两种方技类文献析出为一类。整体聚类框架没有大的变化，而其他经书文献杂糅的现象则更趋明显。至此，实验可以终止。

二、文本分类的有效性与分类依据

在人文社会科学领域，分类总是依赖于因袭的概念和分类，具有根深蒂固的社会文化特性，徐建委所谓“《汉志》主义”即典型。而计算机进行的文本聚类，完全基于词汇及其与上下文之间的共现关系，这就最大限度地避免了先入之见，为跳出“《汉志》主义”提供了更多元的可能。不过，计算机是否“读懂”了文本呢？应用计算语言学方法研究古籍文本分类，还须验证其分类的有效性。

首先看分类数目的有效性。在实验中，聚类项既非越多越好，也不是越少越好，整体上以四分、五分和六分法的结果比较稳定，譬如实验中K-means++模型在K值取5以及K值取8时，新设定的聚类项并未聚集出文本簇，而是形成了《汉高祖诗》《淮南小山赋》的单一聚类，实际上仍保持了四分和六分的分类模式。目录学长期以来使用四至七分的文献分类方式，与计算语言学的结果相一致。

宏观上的聚类有效性可以从聚类数目来验证，那么微观层面自动聚类的结果又如何呢？首先看四分法的聚类结果，司马相如、扬雄等人的大赋作品与《尔雅》等小学典籍聚为一类，而并未与《楚辞》《宋玉赋》等词赋文献聚类，这恰与清代阮元对司马相如等的评价相一致。阮元称赞司马相如与扬雄，就敏锐地捕捉到其赋作与小学的关系：“古人古文小学与词赋同源共流，汉之相如、子云，无不深通古文雅训。”（《扬州隋文选楼记》）“岂有不明音韵篆文训诂，能土拟相如、子云者哉?”（《与学海堂吴学博兰修书》）[①]其后刘师培申论之：“观相如作《凡将篇》，子云作《训纂篇》，皆史篇之体，小学津梁也。足证古代文章家皆明字学。”[②]汉大赋在罗列同源词方面近似小学的特点被很好地捕捉到。此外，《司马迁赋》与《荀卿赋》也一直没有并入辞赋类，而是和诸子类文献聚合，这也符合后人对《汉志》诗赋略的探讨，刘师培推测，屈赋之属乃“写怀之赋”、陆赋之属乃“骋辞之赋”、荀赋之属乃“阐理之赋”，[③]章太炎也认为：“屈原言情，孙卿

①阮元：《揅经室集》，北京：中华书局，1993年，第388、1071页。
②刘师培：《中国中古文学史·汉魏六朝专家文研究》，北京：商务印书馆，2016年，第170页。
③刘师培：《中国中古文学史·汉魏六朝专家文研究》，第174页。

效物，陆贾赋不可见，其属有朱建、严助、朱买臣诸家，盖纵横之变也。”[①]而迁赋属于“陆赋之属”、荀赋乃“荀赋之属”，他们的表达方式本就与大部分现存的属于“屈赋之属”的辞赋不同。

辞赋文献之外，子学文献的自动聚类也有可圈之处。比如《国语》《论语》《战国策》三种文献常常与诸子文献聚在同一部类。而从目录学上说，它们却不在同一部类，《国语》《战国策》入六艺略春秋类，而《论语》成六艺略的论语类，在《四库全书》中它们也分属史部和经部。然而，《国语》与《论语》却是典型的代表春秋君子文化的“语”类文献，《战国策》在经刘向整理之后，也体现出驰骋言辞的特点，它们的核心特点都是“立言”，与诸子文献聚为一部有合理性。

当然，自动聚类的结果只在一定程度上具有相对的合理性与有效性，不能给出百分百准确的结论，它可以揭示某些现象、解释某些问题，但绝非真理。然而，计算语言学方法不带预设条件下所呈现出的相对有效的聚类结果，已足以为我们揭示出《汉志》文本分类的多种可能性。那么进一步分析这些可能的分类，它们依据的标准又如何呢？

从四分法的自动聚类结果中，可以较为明显地看出与四部分类法相对应的特点，其分类当基于言说方式的差别。从言说功能上说，《易传》《毛诗故训传》与《尔雅》等小学文献，其意在诂经而非论述；诸子文献则以立言为主；《左传》《史记》等侧重对事件的撰述，或者对历史的考证；辞赋类则更多抒怀的目的。因而从言说方式上，它们“怎么写”也各有区别，诂经则须阐释、说明，立言贵逻辑与议论，撰述和考证则以叙述性为本，抒怀便须修饰词藻、驰骋才性。因此，文本的言说方式无疑是文本分类的重要依据之一。

而六分法的自动聚类结果显然与言说方式的分类标准相龃龉，呈现出文本另一种可能的维度特征，即思想或内容上的相关性，尤其是《左传》《论语》《孟子》等集中表达儒家仁德思想，以及《尔雅》等集中于音韵训诂的内容。此外，在辞赋的区分上也有以大赋为主和以骚赋为主的不同。这种分类依据“义”，从文本的内容特征维度来分类，更侧重“写什么”。而以王官之学、私门之学、专门之学分属的《七略》，当然也是“以‘义’即书籍的内容性质作为分类标准的”。[②]

在中国古代的目录学分类中，言说方式的原则和文本内容的原则又多是交叉并用的，比如《四库全书总目》，在整体的四部分类法上依据言说方式分类，而在具体的部类中，经部、子部又是根据文本内容分类的。上述的摹拟实验中，五分法的自动聚类结果也是言说方式与文本内容两种分类依据并行的。

---

① 章太炎：《国故论衡》，北京：商务印书馆，2012年，第128页。
② 郭英德：《〈四库全书总目〉与中华学术体系的构建》，《斯文》2017年第1辑。

三、目录何以“主义”？

文本分类基于言说与内容或者并行的不同的文本特征维度，在实验中，它通过分类数量的改变而分别呈现出来。换言之，文献分类的方法往往与分类项的多寡直接相关。而一种分类既然只能依据文本的某一特征维度，那其他文本特征又如何呈现呢？

清代章学诚提出了“互著”“别裁”之说。他据《汉志》班固自注，指出兵书权谋家所录《荀卿子》《陆贾》二种、《伊尹》《太公》《管子》《鹖冠子》四种、《苏子》《蒯通》二种以及《淮南王》一种，分别互著于诸子的儒家、道家、纵横家、杂家，《墨子》并行于兵书技巧家与墨家。他又提道：“经部《易》家与子部之五行阴阳家相出入，乐家与集部之乐府、子部之艺术相出入，小学家之书法与金石之法帖相出入，史部之职官与故事相出入，谱牒与传记相出入，故事与集部之诏诰奏议相出入，集部之词曲与史部之小说相出入，子部之儒家与经部之经解相出入，史部之食货与子部之农家相出入。”（《校雠通义·互著》）[①]这是因为上古“言公”的文化语境下，文献“口耳相传”，并无私家转述，著述以“篇”为单位而无篇名，多缀辑、补充、追记之文，又常见称引他人之说，[②]所以必须使用别裁、互著之体，所谓“古人著书，或离或合，校雠编次，本无一定之规也”。正因为如此，章学诚甚至想到以篇为单位重编古书，“《月令》之于《吕氏春秋》，《三年问》《乐记》《经解》之于《荀子》，尤其显焉者也”（《校雠通义·焦竑误校汉志》）。[③]

章氏以篇为单位解构、重构经典文献的做法，在清人看来“皆极谬妄”。[④]而今人意识到早期文本具有很强的流动性、开放性，则视其说为洞见。章学诚当然是在肯定刘向、刘歆父子“部次流别，申明大道，叙列九流百氏之学，使之绳贯珠联，无少缺逸”（《校雠通义·互著》）[⑤]的意义上有所发明，但反过来也说明了《七略》的分类绝非铁板一块。因为有互著、别裁，所以文献的分类自不可能贯以一律，在这个意义上，“《汉志》主义”自然就体现出对早期文献的遮蔽。那么，当我们由《汉志》现存的文本去逆推刘向、刘歆父子在整理书目时嵌入“考

①章学诚著，叶瑛校注：《文史通义校注》，第1127页。
②参见王汎森：《对〈文史通义·言公〉的一个新认识》，《权力的毛细管作用：清代的学术、思想与心态》，北京：北京大学出版社，2015年，第447—452页。
③章学诚著，叶瑛校注：《文史通义校注》，第1179页。
④李慈铭：《越缦堂读书记》，北京：中华书局，1963年，第782页。
⑤章学诚著，叶瑛校注：《文史通义校注》，第1125页。

镜源流”的思维，最终建构“学术出于王官”的思想之前，或许应该首先追问，刘向、刘歆父子在编排文献时，到底有多少选择的余地？又有多少归类的可能？

考察《七略》的成书，时间问题非常关键。据《汉书·成帝纪》所载，刘向始校中秘书的时间是河平三年（公元前26），至绥和元年（前8）刘向去世，其间共经历了不到二十年时间。《汉书·楚元王传》又载：“向死后，歆复为中垒校尉。哀帝初即位，大司马王莽举歆宗室有材行，为侍中太中大夫，迁骑都尉、奉车光禄大夫，贵幸。复领《五经》，卒父前业。歆乃集六艺群书，种别为《七略》。”[①]可知，刘歆继承父业是在哀帝即位的绥和二年（前7）。至于刘歆上奏《七略》的时间，包括钱穆、刘汝霖、陆侃如等人都认为就在绥和二年。钟肇鹏的结论稍晚，他根据《山海经书录》标注的“建平元年四月丙戌”而推论《七略》编成当略晚于此，[②]如此则《七略》的上奏应是在建平元年（前6）稍后。而从《汉书》的描述看，刘歆“复领《五经》”之前还有一段改官历程，因而吴沂澐推测：“刘向卒后至刘歆复领《五经》期间，总领校书者暂时虚悬，或可进一步推测刘向专责的六艺、诸子与诗赋略之校书活动也可能因此短暂中断。”[③]这种结论是有可能的。且《别录》当只是刘向叙录的汇合，而没有总理群书之分类目录。[④]这样算来，留给刘歆完成《七略》的时间可能只有一年，相比于其父刘向18年的校书历程来说，实在非常急迫。

短时间内迅速成书上奏，刘歆只好延续刘向的工作方法。据《汉书·艺文志》：“诏光禄大夫刘向校经传诸子诗赋，步兵校尉任宏校兵书，太史令尹咸校数术，侍医李柱国校方技。”[⑤]各略主纂皆为官守其书的专职担任。这种工作方式有制度上的便利，也有学术上的优势。刘歆在短时间内整理《别录》而成《七略》，自然也无暇变更文献整理的操作方式。而以官守为标准的校书活动，其本质上正是基于文本内容的分类方式，因而以文本内容而不是以言说方式来进行文献分类，本就是刘歆不得不然的优势选项。

问题在于，何以本来自然而然的六分法在后世会在分类方法上形成一连串的“问题”？尤其是在史部之确立上，文献学的探讨进一步延伸到对汉代思想史、或者刘歆之学术思维的解释。[⑥]而若要考虑这个问题，也许首先应该辨析史部之

---

①班固：《汉书·楚元王传》，北京：中华书局，1962年，第1967页。
②钟肇鹏：《七略别录考》，《文献》1985年第3期。
③吴沂澐：《刘向、刘歆校书问题考述》，《斯文》2018年第3辑。
④姚名达：《中国目录学史》，北京：商务印书馆，2017年，第41页。
⑤班固：《汉书》，第1701页。
⑥如张涛提出汉代史学成为经学附庸才是《七略》中史籍未能独立成部类的根本原因。参见张涛：《〈七略〉中史籍未能独成部类的根本原因》，《文史哲》1992年第6期。

独立与否，之于刘向父子究竟是一个“真命题”，还是一个“伪命题”？

关于史部的问题，传统的结论是从文献多寡的角度考虑的。南朝梁代阮孝绪在《七录序》中总结说：“刘氏之世，史书甚寡，附见春秋，诚得其例。今众家记传，倍于经典，犹从此志，实为繁芜。”[①]因而史籍须独立一部。后人的论述基本不出阮孝绪之说，如章学诚谓：“史部日繁，不能悉隶以《春秋》家学，四部之不能返《七略》者一。名墨诸家，后世不复有其支别，四部之不能返《七略》者二。文集炽盛，不能定百家九流之名目，四部之不能返《七略》者三。”（《校雠通义·宗刘》）[②]余嘉锡也总结称：“书之有部类，犹兵之有师旅也。虽其多寡不能如卒伍之整齐划一，而要不能大相悬绝，故于可分者分之，可合者合之。”“七略之变而为四部，不过因史传之加多而分之于《春秋》，因诸子、兵书、数术、方技之渐少而合之为一部。”[③]

然而这一解释也不断引起人们的怀疑，比如姚名达指出，《七略》“往往同一种中，又复杂附绝不同类之书，如附《国语》《世本》《战国策》《楚汉春秋》《太史公》《汉大年纪》十二家之书于《春秋》，附《帝王诸侯世谱》《古来帝王年谱》于《历谱》。若谓史书甚少，不必独立；则其他各种，每有六七家百余卷即成一种者；而谓以十二家五百余篇之史书反不能另立一种乎？”[④]此说有一定道理。据《汉书·艺文志》著录，六艺略3,123篇、诸子略4,324篇、诗赋略1,318篇、兵书略790篇、数术略2,528卷、方技略868卷，可见并不是按照文献数量而取大致相同的规模。而春秋类计948篇，[⑤]其数量还超过兵书略，几乎与诗赋略规模相当。所以《七略》的分类似乎也不完全是根据多寡来分的。

在刘向、刘歆那里，文献分类基于文献整理，而文献整理时为了工作上的方便，由各类专家分领其事，最终形成以文本内容为依据的分类。但后人审视目录学的历史，却拟出一条从六分法到四分法的演化路径，并在这一线性发展的线索中，框定各目录的性质。而我们首先在实验结果上明确《七略》之四分、五分与六分各有理据，再从历史语境上论证六分法只是一种自然选择之后，似乎也有必要重审《七略》演为四部之说的理路。

文献分类由六分演化而为四分，其中关键一环是阮孝绪的《七录》。阮氏提及刘向父子的目录，主要是为了论证《七录》内篇五分法的合理性。与《七略》相比，《七录》多《记传录》而少《兵书略》，所以阮氏一方面强调“刘氏之世，

---

①严可均辑：《全上古三代秦汉三国六朝文》，北京：中华书局，1958年，第6691页。
②章学诚著，叶瑛校注：《文史通义校注》，第1114页。
③余嘉锡：《目录学发微》，北京：中华书局，2007年，第144—145、168页。
④姚名达：《中国目录学史》，第59—60页。
⑤以上数字，出于班固统计。参见陈国庆：《汉书艺文志注释汇编》，北京：中华书局，1983年。

史书甚寡”而“今众家记传，倍于经典”，一方面指出“兵书既少，不足别录，今附于子末”。[①]以此为他变更部类提供解释。或许是因为看到阮孝绪独立《记传录》而删裁《兵书略》的做法，加之《七录》的子目划分又为《隋书·经籍志》所继承，今人在考虑《七录》内篇五部时，多认为它是大略依照四分法的结果。[②]如此一来，阮孝绪本来可能只是为了自证其调整部类合理性的一番序言，便可以反推回去，成为了阐释《七略》不能设立史部的理由了。但如果加上外篇的《佛法录》和《仙道录》，统观七录而不是内篇五录的话，则说阮孝绪大略依照四部的分类可能就不完全合理，因为《七录》的整体分类同样有基于文本内容为分类依据的特征。类似地，荀勖《晋中经簿》也被认为是始创四部之祖，其实《晋中经簿》于四部外另有“佛经”一部，与《七录》都属于交叉运用言说方式与文本内容为依据的分类法，亦近似本实验中《汉志》的五分法策略。

《七录》和《晋中经簿》并未完全采用基于言说方式的分类法，后人却相信其中已有四分法的用意，这是否显示了四分法对二书的“收编”呢？在这个意义上，不仅《汉志》成为后人理解周秦汉文献的预设与前理解，四部分类法也会成为后人潜移默化地理解古代文献的基本预设。换言之，后人对古籍目录分类方法的认知，很可能潜移默化地以四分法为参照标准。于是，基于文本内容的六分法与基于言说方式的四分法，在目录学史的角度看，也就不再是一个“选择”问题，而成为“演化”问题，甚至明察如章学诚都认为“四部之与《七略》，亦势之不容不两立者也”（《校雠通义·宗刘》）。[③]

其实在文本分类时，本来存在多种可能，正如文本自动聚类实验所呈现出来的，分类的数目往往与分类方法直接相关。当类目较少，言说方式的维度更鲜明时，四部的界限会浮现出来，而当类目渐多，文本内容的维度更有优势时，四部的界限就比较模糊。四部分类法果然不可能返回《七略》式的文本内容分类吗？从内容分类到言说方式的分类，果然就是线性不可逆的吗？答案当然是否定的。在《四库全书总目》将四部分类法运用得登峰造极之后，民国时期所采用的美国学者Dewey的“十进分类法”以及当代所使用《中国图书馆图书分类法》就再次回到了基于文本内容的文献分类方式，则“四部之不能返《七略》”当然也就未必那么言之凿凿了。

或许我们应该更清晰地意识到，在中国学术文化传统中，学术体系的构建、学者知识结构的形成，无不与目录分类的“格式化”密切关联着，“图书分类在

① 严可均辑：《全上古三代秦汉三国六朝文》，第6691—6692页。
② 郭英德、于雪棠：《中国古典文献学的理论与方法》，北京：北京师范大学出版社，2008年，第306页。
③ 章学诚著，叶瑛校注：《文史通义校注》，第1117页。

现象层面是一种文化知识体系的构建方式，在本质层面则是一种学术体系的构建方式”。[①]而一旦人们处于由目录分类所构建的知识世界与学术体系之中时，往往就会在思考分类问题，或者在考虑文献之相互关系、学术之源流演进时，倾向于对目录分类的逆推。如此，则不仅《汉志》已成为“主义”，而作为“终极分类法”[②]的四部分类法，亦有可能成为某种“主义”了。意识到“目录主义”潜移默化的影响，进而思考超越之途径，亦或是今人研究中需要有所考量的。

Re–examination of the Catalogue Classification of *Hanshu Yiwenzhi*

Zhu Yuchen, Li Shen

Abstract: In order to investigate the multiple possibilities of literature classification in *Hanshu Yiwenzhi*, this paper uses the K-means++, Gaussian Mixture Model and Spectral Clustering model to automatically cluster the surviving documents in *Hanshu Yiwenzhi*. The clustering results show that the clustering from four to six classifications is stable and effective. There are two kinds of classification methods based on the principle of expression and content, which provide more possibilities to beyond *Hanzhi*. Under the framework of four-part classification, later generations examined the six division of Hanzhi and put forward such propositions as “why the historical department is not established”, which also shows the subtle influence of the four-part classification in the construction of knowledge system. In ancient culture, bibliography has gone beyond simple classification catalogue and become a thinking mode with potential influence.

Keywords: *Hanshu Yiwenzhi*; *QiLue*; Classification; Automatically Cluster

（编辑：王波）

①郭英德：《〈四库全书总目〉与中华学术体系的构建》，《斯文》2017年第1辑。
②张升：《历史文献学》，北京：北京师范大学出版社，2016年，第199页。