

清代奏摺数据库

数据说明

2026 年 9 月 1 日

目录

第一节 数据库简介	2
第二节 奏折样本	3
第三节 变量定义	5
第四节 网站检索	15
第五节 声明	17

第一节 数据库简介

奏折始于清代康熙朝，康熙允许其亲信官员绕过常规的官僚系统，秘密向其报告所需信息。雍正即位后将奏折制度化，雍正要求中央与地方的高级官员将重要事务和信息及时通过奏折直接向皇帝报告。阅读奏折后，皇帝会在奏折末尾用朱红色笔写下回复，称为朱批，奏折也被称为朱批奏折。随后奏折发还于奏报官员，让他们知悉皇帝的回复或指示。官员处理完相关事务后，需要定期将朱批奏折原件交到中央，存于官中档，军机处也会抄录一份作为副本保存，称作录副奏折。得益于这一执行良好的档案保存制度，大部分奏折完整地留存到了今天。

皇帝和官员围绕奏折产生的互动（信息不对称下的委托代理问题）是我在博士期间研究的核心话题。我从 24 年开始收集了北京第一历史档案馆和台北国立故宫博物院¹公开的奏折数据，构建了清代奏折数据库，并利用这些数据完成了博士论文。在原始奏折数据的基础上，清代奏折数据库中包含许多我在研究过程中新收集和整理的变量，它们共同描述了一份信息从官员到皇帝，再从皇帝到官员的报告和处理过程。

清代奏折数据库可以免费下载。

¹下称一史馆和台北故宫

第二节 奏折样本

现存奏折主要藏于北京第一历史档案馆和台北国立故宫博物院两处。

在奏折的收录与保存过程中，官员书写的奏折原件被称为“朱批奏折”，待官员根据朱批处理完对应事务后，缴回存于官中档；从乾隆朝开始，军机处会抄录一份奏折留作副本，称为“录副奏折”，存于军机处。

对第一历史档案馆所藏，条件检索档案目录筛选中的“朱批奏折”，得到 525909 份朱批奏折数据，条件检索档案目录筛选中的“录副奏折”与“满文录副奏折”，得到 737812 份录副奏折数据，一史馆数据共 1263721 份。对台北故宫所藏，筛选“官中档奏折”，得到 158614 份朱批奏折数据，筛选“军机处档折件”，得到 190889 份录副奏折数据，台北故宫数据共 349503 份。两个数据源合并，共得到 1613224 份奏折数据。覆盖康熙、雍正、乾隆、嘉庆、道光、咸丰、同治、光绪、宣统九位皇帝的在位年份。

一史馆和台北故宫关于奏折的统计口径和披露方式存在显著差异。一史馆中，一份奏折附属的单、夹片等附件均会被单独视作一份数据，且有相当一部分非奏折资料被一史馆误认为是奏折（比如印券、切结等）。因此，档案来源为一史馆的数据并非真的都是朱批奏折和录副奏折，变量 `type` 提供了详细的档案类型，使用时请关注。而台北故宫则将这些附件与主奏折视作一份数据。为了与一史馆的数据结构一致，我将台北故宫的奏折数据进行了拆分，如果一份台北故宫的奏折同

时包含了主奏折和附件，那附件就会被拆分出来作为额外的数据观测¹。变量 `type` 记录了这一拆分详情，取值“台北故宫奏折的原始奏折”为未拆分的原始记录，而取值“台北故宫奏折中的括号内容”为拆分出的附件观测。因此，清代奏折数据库中的总数据量为 1693573 份奏折。

台北故宫公布了宫中档朱批奏折和军机处录副奏折的档案图片，从而可以看到奏折正文。由于精力、时间和职业规划的限制，我只对部分台北故宫的朱批奏折正文利用 DeepSeek 进行了整理，共 113238 份台北故宫朱批奏折有正文数据。未来有机会的话可能再行补充完善。因此奏折正文检索仅供参考，样本有限，不建议直接用作研究。

请注意：由于一史馆和台北故宫可能不定期更新其数据，因此清代奏折数据库的数据样本仅供参考。使用该数据库，筛选研究样本时，应对奏折制度和奏折数据有良好了解，以档案来源 `source` 和奏折类型 `type` 等变量为参考，全面考虑后使用。

¹举例，一份台北故宫的奏折题为“奏报 xx 府被水事 (xx 府 xx 年雨水单)”，那么在数据库中，这份奏折会有两条记录，一条为题为“奏报 xx 府被水事 (xx 府 xx 年雨水单)”的原始数据（这份数据和台北故宫网站公开数据一致），另一条题为“(xx 府 xx 年雨水单)”，记录附件的信息。

第三节 变量定义

本节介绍清代奏折数据库中所有变量的详细定义和特殊取值含义。

id: 就是 id, 每条观测 unique.

以下是奏折基本信息变量:

title: 奏折题目。是对奏折全文内容的概括。实际上, 奏折题目并非是由当时的奏报官员亲自书写, 而是由今天的古籍整理者根据相关文件规定所拟, 原则上遵循“责任者(作者) + 事由”的逻辑拟题¹。

count_title: 题目字数。

title_revised: 修正后的题目, 即去掉题目中括号内容后的题目。

titlelabel_DeepSeek: DeepSeek 根据 title_revised 对奏折报告事务进行的分类。部分分类错误, 不建议直接使用, 需谨慎。

titlelabel_DeepSeek_level: DeepSeek 在分类奏折时是否认为有把握, 1 有把握, 2 无把握。

count_titlerevised: 修正后题目的字数。

prefecture: 奏折题目 title_revised 中出现过的府, 包括直接出现的府, 以及出现的县对应的府。通过匹配完成, 不完全准确, 使用时自行核查。

以下是奏报官员相关变量:

¹规定详见《明清档案著录细则》

report_position: 写奏折官员的官职。官职信息是由官员亲自书写的。

report_position_revised: 对奏报官员的官职数据进行了人工检查, 对部分明显错误(如错字)的官职记录进行了修正。

position_province: 奏报官员管辖的省份。根据 report_position_revised 进行确定。

report_official: 奏报官员。

report_official_revised: 对奏报官员数据进行了人工检查, 对部分明显错误(如错字)的官员记录进行了修正。

report_official_many: 取值为 1 表示奏报官员多于 1 人。

manhan_sanpin: 奏报官员的满汉身份, 根据人物传记资料库中部分三品官员数据确定。使用时需谨慎, 可能存在错误。

hometown_province_sanpin: 奏报官员的籍贯省, 根据人物传记资料库中部分三品官员数据确定。使用时需谨慎, 可能存在错误。

hometown_prefecture_sanpin: 奏报官员的籍贯府, 根据人物传记资料库中部分三品官员数据确定。使用时需谨慎, 可能存在错误。

manhan_rwzj: 奏报官员的满汉身份, 根据人物传记资料库数据确定。使用时需谨慎, 可能存在错误。

hometown_province_rwzj: 奏报官员的籍贯省, 根据人物传记资料库数据确定。使用时需谨慎, 可能存在错误。

hometown_prefecture_rwzj: 奏报官员的籍贯府, 根据人物传记资料库数据确定。使用时需谨慎, 可能存在错误。

manhan_xty: 奏报官员的满汉身份, 根据 Xi. T (2019) 确定。

hometown_province_xty: 奏报官员的籍贯省, 根据 Xi. T (2019) 确定。

hometown_prefecture_xty: 奏报官员的籍贯府, 根据 Xi. T (2019) 确定。

birthdeath_CN_sanpin: 奏报官员的中文生卒年份, 根据人物传记资料库中部分三品官员数据确定。使用时需谨慎, 可能存在错误。

birthdeath_sanpin: 奏报官员的生卒年份, 根据人物传记资料库中部分三品官员数据确定。使用时需谨慎, 可能存在错误。

birth_sanpin: birthdeath_CN_sanpin 中的公历生年。

birth_unsure_sanpin: birthdeath_CN_sanpin 中的公历生年不确定。

death_sanpin: birthdeath_CN_sanpin 中的公历卒年。

death_unsure_sanpin: birthdeath_CN_sanpin 中的公历卒年不确定。

birthdeath_CN_rwzj: 奏报官员的中文生卒年份, 根据人物传记资料库数据确定。使用时需谨慎, 可能存在错误。

birthdeath_rwzj: 奏报官员的生卒年份, 根据人物传记资料库数据确定。使用时需谨慎, 可能存在错误。

birth_rwzj: birthdeath_CN_rwzj 中的公历生年。

birth_unsure_rwzj: birthdeath_CN_sanpin 中的公历生年不确定。

death_rwzj: birthdeath_CN_rwzj 中的公历卒年。

death_unsure_rwzj: birthdeath_CN_rwzj 中的公历卒年不确定。

以下是奏折日期变量:

report_ymd: 具奏日期, 官员写奏折的农历日期。

report_year: 具奏年。

report_month: 具奏月。

report_monthrun: 具奏闰月。

report_day: 具奏日。

report_ymd_solar: 具奏日期对应的阳历日期, 部分数据有。

report_year_solar: 具奏阳历年。

report_month_solar: 具奏阳历月。

report_day_solar: 具奏阳历日。

以下是朱批基本信息相关变量:

comment: 朱批。皇帝在这份奏折后书写的朱批。仅台北故宫样本有朱批数据。

count_comment: 朱批 comment 的字数。

comment_revised: 修正朱批。去掉 comment 中的括号内容、去掉标点符号数字字母。

count_commentrevised: 朱批 comment_revised 的字数。

comment_ymd: 朱批日期, 皇帝写朱批的农历日期, 仅台北故宫的军机处档折件样本有。

comment_year: 朱批年。

comment_month: 朱批月。

comment_monthrune: 朱批闰月。

comment_day: 朱批日。

以下是朱批标签变量:

以下变量均为描述朱批特征的虚拟变量。为保证对朱批特征的准确刻画, 我阅读了所有朱批, 确定了以下标签来描述朱批包括决策模式在内的各类特征。阅读每份朱批后, 我为其分配对应的标签: 如果一份朱批表现出某些标签刻画的行为或特征, 这份朱批就被赋予这些对应的标签。一份朱批可能同时具有多个标签。

comment_iSee: 知道了。皇帝只回复了“览”或“知道了”等类似内容, 没有其他任何内容。

comment_ComfortingToSee: 欣慰知道。皇帝表示欣慰地看到奏报信息。

comment_AbsorbingToSee: 加额知道。皇帝表示认真地阅读了奏报信息。(与欣慰览之相似)

comment_MoodSharing: 分享心情。皇帝对奏报事务表达了他的心情。

comment_MessageNotification: 告知信息。皇帝向奏报官员告知了一些信息。

comment_IssueReported: 已有题奏。皇帝告诉奏报官员同样的信息已有其他官员报告。

comment_IamWell: 朕安。皇帝回应官员的请安奏折。

comment_Health: 关心健康。皇帝关心官员或其亲属的健康。

comment_Greeting: 问候官员。皇帝向官员或其亲属表达问候。

comment_Reward: 赏赐。皇帝奖赏奏报官员。

comment_Encouragement: 勉励。皇帝勉励、激励奏报官员。

comment_GoodJob: 甚好/做得好/奏得好。皇帝称赞奏报官员及时报告信息并且表现出色。

comment_WellDone: 甚是/做得是/奏得是。皇帝称赞奏报官员及时报告信息并且表现出色。(与甚好/做得好/奏得好相似)

comment_TakeItSeriously: 认真勉励秉公速办。皇帝告诉奏报官员认真、速速、秉公、严肃地处理奏报事务。

comment_UseDirection: 相机酌量。皇帝告诉奏报官员相机、灵活地处理奏报事务。

comment_Strictly: 严办。皇帝告诉奏报官员严格处理奏报事务，多为案件相关事务。

comment_DiscussReport: 转发议奏。皇帝将奏折转交给有关部门(或官员)

讨论处理。

`comment_YouKnowIt`: 转发知道。皇帝将奏折转交给有关部门（或官员），以便他们获取相关信息并处理。

`comment_ForwardProcess`: 转发办。皇帝将奏折转交给有关部门（或官员）直接办理。

`comment_GrandCouncilDispatch`: 廷寄。皇帝将奏折转交给军机处，由他们处理并起草发给奏报官员的指示。

`comment_ReferAudience`: 送部引见。皇帝程序化地将人事事务转交给相关部门处理。

`comment_PresentAudience`: 带来引见。皇帝程序化地将人事事务转交给相关部门处理。（该子标签与送部引见相似）

`comment_SendSomeone`: 派人去。皇帝派官员去处理。

`comment_Agree`: 同意认可。皇帝同意或认可了请求/意见/做法/奏报内容。

`comment_ApprovedAsProposed`: 依议。皇帝同意或认可了奏报内容。（固定格式）

`comment_ApprovedAsRequested`: 着照所议/请。皇帝同意或认可了奏报内容。（固定格式）

`comment_Formally`: 着。皇帝如橡皮图章般正式化了奏报提议，多为人事相关事务。（固定格式）

`comment_Granted`: （着）准。皇帝如橡皮图章般批准了奏报提议，多为人事相关事务。（固定格式）

`comment_FormallyGranted`: 着准。皇帝如橡皮图章般正式批准了奏报提议，多为人事相关事务。（固定格式）

comment_FormallyXGranted: (着) 不准。皇帝如橡皮图章般正式批准了奏报提议, 多为人事相关事务。(固定格式, 着不准意味着正式化了奏折中关于不准某类人事事务的请求)

comment_AsUsual: 照例。皇帝同意按照惯例行事。(固定格式)

comment_SummonedAudience: 来京陛见。皇帝告诉或同意官员来北京觐见。

comment_AssumedOffice: 速赴任。皇帝告诉官员尽快前去岗位赴任。

comment_ExcusedAudience: 不必来京陛见。皇帝告诉官员不用来北京觐见。

comment_ScheduledAudience: 不必即来京陛见。皇帝告诉官员在方便时到北京觐见。

comment_NotHaveTo: 不必。皇帝表示这是不必要的。(相对复杂, 存在多种可能性。)

comment_DiscussWithSomeone: 与人商量。皇帝告诉奏报官员在处理时要与其他官员共同商议。(其他官员多为上级地方官员)

comment_FollowSomeone: 听人指挥。皇帝告诉奏报官员听他其他官员指挥。(与人商量和听人指挥两个子标签暗示皇帝对奏报官员缺乏信心)

comment_Warn: 敲打。皇帝言语敲打了奏报官员, 对人。

comment_Complain: 抱怨。皇帝对繁多的事务表示不满, 并抱怨官员不够勤勉、倾向于制造麻烦。

comment_Criticize: 批评。皇帝批评了官员报告的奏报事务(的出现、处理方式或结果), 对事。

comment_ReportRequested: 要求奏来。皇帝要求官员报告某些信息。

comment_Inquiry: 询问质问。皇帝在询问他想知道的信息。

comment_Disagree: 不同意不认可。皇帝不同意请求/意见/做法/奏折内容。

comment_PersonalRegard: 识人。皇帝表达了他对相关官员能力行为处事
看法。

comment_Opinion: 看法。皇帝发表了自己对奏报事务/信息的看法。

comment_Command: 指令。皇帝在朱批中直接给出了对奏报事务的指令。

comment_NoNeedToDiscuss: 无庸议。皇帝表示没有必要再讨论了。

comment_EdictDone: 已有旨。皇帝告诉奏报官员已发布相关旨意。(固定格
式)

comment_EdictWill: 余/另/即有旨。皇帝告诉奏报官员即将会发布相关旨
意。(固定格式)

comment_Tiben: 具题。皇帝告诉奏报官员用题本报告。(固定格式)

comment_SendOut: 单并发。皇帝向相关部门转交了奏折中额外报告的清单
文件。(固定格式)

comment_KeepIn: 单留中。皇帝告诉奏报官员, 他将暂时将奏折或部分文件
留在中央而不发回。(固定格式)

comment_MarginalComment: 另有夹批。皇帝有在奏折的额外附件中写有朱
批。

comment_ViaGrandCouncil: 军机大臣奉旨。朱批的公式化开头。(主要出现
在清末朱批由摄政者代笔时期)

comment_ActingOnComment: 奉朱批。朱批的公式化开头。(主要出现在清
末朱批由摄政者代笔时期)

comment_ManchuLanguage: 有满文朱批。皇帝写有满文朱批。

comment_Meaningless: 无法分类。由于奏折损坏或朱批不完整, 导致无法分
类。

以下是奏折档案来源类别相关变量：

emperor_source：皇帝。根据档案来源确定的皇帝，不一定准。

emperor_ymd：皇帝。根据具奏日期中的年号确定的皇帝，不一定准。

type：奏折类型。台北故宫奏折中的原始奏折表示从台北故宫处收集的原始数据；台北故宫奏折中的括号内容表示从原始数据中拆分出的奏折附件。见第二节奏折样本。其他取值代表样本来自一史馆，为一史馆对这份史料备注的具体种类。

language：一史馆样本中备注的文字种类。

remark：一史馆档案的官方备注。

id_tbgg：台北故宫样本的 id。共 349503 个。

id_ysg：一史馆样本的 id。共 1263721 个。

source：档案来源。奏折来源于哪个存档单位。

source_id：档案编号。奏折在一史馆和台北故宫的原始档案编号，可用于在官网检索。

id_tbggsource：无需在意，为爬虫时 id。

id_ysgsource：无需在意，为爬虫时 id。

id_tbgginer：无需在意，为皇帝朝代内部 id。

id_ysginer：无需在意，为皇帝朝代内部 id。

id_kuohao：给台北故宫拆分出的奏折附件一个 id。

county_direct：奏折题目 title_revised 中直接出现的县。用于构造变量 prefecture。

prefecture_direct：奏折题目 title_revised 中直接出现的府。用于构造变量 prefecture。

province_direct：奏折题目 title_revised 中直接出现的省。

以下是奏折正文变量：

maintext：奏折正文。仅部分台北故宫朱批奏折有，利用 DeepSeek 对奏折直接进行识图所得，较为粗糙，仅供参考和阅读，不建议直接使用和分析。由于奏折正文占用存储空间较大，因此未直接提供下载，如希望获得，可直接邮件联系我，或者自己爬虫。由于我的时间和精力已不在学术上，因此仅简单地对部分奏折的正文进行了处理，未来有机会的话再尝试补充。

第四节 网站检索

网站的检索框和奏折框中展示了第三节中的部分变量。

检索框中，它们是：

奏折题目(title)、具奏官员(report_official_revised)、官职(report_position_revised)、奏折正文(maintext)、题目提及府(prefecture)、皇帝(emperor_source 和 emperor_ymd)、起始年结束年(report_year)、朱批(comment)、朱批标签(各类朱批标签变量)、事务分类(titlelabel_DeepSeek)、档案来源(source)、奏折类型(type)

奏折框中，除了上面提到的，还有：

具奏日期(report_ymd)、朱批日期(comment_ymd)、档案编号(source_id)

网站的检索功能支持

1. 多关键词检索，关键词之间以空格连接。
2. 多关键词间关系支持“且”、“或”两种关系，不同变量间的逻辑关系仍为“且”关系。
3. 支持检索变量是否有数据，输入 YESDATA 即可查询某变量有数据的样本，NODATA 查询某变量没有数据的样本。
4. 奏折正文检索支持简体和繁体。

5. 部分变量支持下拉框选择关键词。

第五节 声明

免责声明：原始数据来自公开的北京第一历史档案馆和台北国立故宫博物院，本网站不具有任何商业用途。大规模数据的处理难免出现遗漏、错误、误判等各种问题。使用这一数据库时，应当自行对相关数据进行检查和筛选，清代奏折数据库对史料和变量的正确性不做任何承诺。对使用该数据库产生的任何后果，我不承担任何责任。使用该数据库时无需提及或引用我。如果你使用这一数据库，则视为同意以上所有内容。

如对清代奏折数据库有任何问题或建议，请联系 peihengohuo@163.com。