

面向警务业务的数据可视化处理平台设计与实现

侯学慧^{1*}, 冯玉花¹, 肖淑蓉²

¹新疆警察学院信息安全工程系, 新疆 乌鲁木齐

²新疆维吾尔自治区人民检察院喀什分院/第一检察部, 新疆 喀什

收稿日期: 2026年2月22日; 录用日期: 2026年3月18日; 发布日期: 2026年3月25日

摘要

针对警务业务中多源异构数据难以快速理解、检索分析与直观呈现的问题, 本文面向警务数据分析与教学科研场景, 设计并实现了一套桌面端数据分析可视化处理平台。平台以Python为主要开发语言, 采用PyQt5构建图形交互界面, 并通过QtWebEngine将pyecharts生成的交互式HTML图表嵌入客户端, 实现“数据处理-分析建模-交互可视化展示”的一体化流程。系统围绕三类典型任务构建模块: 1) 网络舆情文本分析与多视图可视化, 基于中文分词与停用词过滤完成词频统计, 采用TF-IDF评估关键词重要性, 生成词云、词频/权重表、评论长度分布图, 并结合IP/地域信息进行二维地图展示; 2) 警情文本信息检索与统计展示, 对案件类别进行聚合统计并生成饼图/玫瑰图, 支持按类别条件过滤与表格化结果回显; 3) 基于机器学习的人脸识别流程, 采用PCA进行特征降维并训练SVM分类器, 通过准确率、F1值与混淆矩阵对模型性能进行评估, 同时以可视化方式展示推理结果。实验结果表明, 该平台具有较好的交互性、模块化与可扩展性, 可为警务数据可视化处理系统原型构建及教学实训提供工程化参考。

关键词

警务数据, 数据可视化, 文本挖掘, 舆情分析, PyQt5, PCA, SVM

Design and Implementation of a Data Visualization and Processing Platform for Policing Operations

Xuehui Hou^{1*}, Yuhua Feng¹, Shurong Xiao²

¹Department of Information Security Engineering, Xinjiang Police College, Urumqi Xinjiang

²First Procuratorial Department, Kashgar Branch of the Xinjiang Uygur Autonomous Region People's Procuratorate, Kashgar Xinjiang

*通讯作者。

文章引用: 侯学慧, 冯玉花, 肖淑蓉. 面向警务业务的数据可视化处理平台设计与实现[J]. 计算机科学与应用, 2026, 16(3): 782-788. DOI: 10.12677/csa.2026.163103

Abstract

To address the difficulties in rapidly understanding, retrieving, analyzing, and intuitively presenting multi-source heterogeneous data in police operations, this paper designs and implements a desktop-based data analysis and visualization processing platform oriented toward police data analysis as well as teaching and research scenarios. The platform is primarily developed using Python, with PyQt5 employed to construct the graphical interactive interface. Interactive HTML charts generated by pyecharts are embedded into the client through QtWebEngine, enabling an integrated workflow of “data processing - analytical modeling - interactive visualization.” The system consists of three task-oriented modules: 1) Online public opinion text analysis and multi-view visualization, where Chinese word segmentation and stop-word filtering are applied to perform word frequency statistics, TF-IDF is used to evaluate keyword importance, and visualizations such as word clouds, word frequency/weight tables, and comment length distribution charts are generated. Additionally, IP and geographic information are combined to display results on a two-dimensional map; 2) Police incident text information retrieval and statistical visualization, which aggregates statistics by case category and generates pie charts or rose charts, while supporting conditional filtering by category and tabular result display; 3) A machine learning-based face recognition workflow, where PCA is applied for feature dimensionality reduction and an SVM classifier is trained. Model performance is evaluated using accuracy, F1-score, and confusion matrix, while inference results are presented through visualization. Experimental results demonstrate that the platform exhibits good interactivity, modularity, and scalability, providing an engineering reference for the development of police data visualization system prototypes as well as teaching and training applications.

Keywords

Policing Data, Data Visualization, Text Mining, Public Opinion Analysis, PyQt5, PCA, SVM

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着公安大数据应用推进，警务工作对数据处理、建模分析、预测预警与可视化呈现的需求持续增强，尤其在社会治理、风险研判、网络舆情监测、案件信息查询与辅助决策等环节，需要将海量数据转化为可理解、可交互的可视化信息[1][2]，以提升研判效率与协同处置能力。已有研究指出，应强化警务大数据综合应用能力建设，突出警务数据可视化等能力，以适应复杂治安形势下的实战需求[3][4]。

本文搭建警务数据分析可视化平台系统，面向典型警务数据处理任务，构建集成式桌面应用平台系统。平台采用图形界面承载业务流程，以交互式可视化呈现关键结果，模块覆盖文本挖掘、统计可视化、空间可视化与机器学习识别等处理链路，目标是形成可复用的警务数据可视化处理范式与工程实现路径。

本文主要贡献如下：1) 提出面向警务数据处理的桌面端可视化平台总体架构，实现图表网页与 GUI 的融合式展示。2) 实现网络舆情文本分析与多视图联动展示，形成词频表、词云、字数分布与地图等可视化组合。3) 实现警情文本信息检索与类别统计展示。4) 实现基于主成分分析 PCA 与支持向量机 SVM 模型的小型人脸库识别训练与预测流程。

2. 需求分析与总体设计

2.1. 业务需求与功能划分

系统主要面向三类应用：网络舆情评论信息分析[5]、警务文本信息搜索及可视化统计[6]、基于机器学习的人脸识别[7]。据此平台划分为三大模块：模块一“网络舆情信息可视化”：评论字数统计、关键词重要性统计表、关键词词云、关联地址信息地图展示等。模块二“警情文本信息搜索”：案件类别统计饼图、按类别检索与表格展示。模块三“人脸识别”：构建人像库，图像预处理，PCA 降维与 SVM 模型训练，对新图像进行预测，并输出识别结果与评估指标。

2.2. 运行环境与关键技术

平台运行于本地计算机 CPU 环境，开发工具为 PyCharm，Python 版本为 3.8，依赖库包含 Numpy、Sklearn、Pyecharts、PyQt5、Matplotlib、PIL、sqlite3 等。关键技术路线如下：

- 1) 客户端交互：PyQt5 构建 GUI，Qt Designer 辅助界面设计，QtWebEngineView 嵌入网页图表。
- 2) 可视化生成：pyecharts 生成 Bar、Pie、Table、Map、Map3D、WordCloud 等，并渲染为 HTML 文件供客户端加载。
- 3) 文本挖掘：中文分词与停用词处理，TF-IDF 评估关键词重要性并可视化展示。
- 4) 人脸识别：PCA 降维后使用 SVM 进行分类训练，采用 Accuracy、Precision、F1、混淆矩阵等评估。

2.3. 系统架构

平台采用分层与模块化设计：1) 表示层：PyQt5 主窗口与多视图控件，负责菜单切换、输入触发、结果呈现。代码中通过多个按钮连接槽函数实现模块切换与交互，例如网络舆情模块、检索模块、人脸识别模块。2) 业务层：封装文本分析、统计汇总、检索过滤与识别推理逻辑。3) 可视化渲染层：集中在 charts 逻辑中生成 HTML 图表文件，并由 QtWebEngineView 加载显示，形成图表生产与展示解耦。4) 采用 SQLite3 作为本地持久化存储与查询引擎。系统首次运行时将原始 CSV 数据集自动导入 SQLite 数据库并完成表结构初始化；后续文本分析与检索统计均通过 SQL 查询获取数据，结合对情感字段与类别字段的索引设计提升查询效率与响应速度。

3. 警务数据可视化处理方法

3.1. 网络舆情文本处理与多视图展示

3.1.1. 数据预处理与关键词重要性计算

本文舆情实验所使用的 weibo_sentiment_labeled_dataset 数据集来源于公开微博评论数据。数据采集时间为 2025 年 3 月 28 日，覆盖 57 个社会话题，共采集文章 1986 篇，累计获得评论 15,142 条。为提升数据管理与查询效率，平台将原始 CSV 数据在首次运行时自动导入 SQLite3 数据库：系统初始化阶段完成数据库文件创建、表结构建立与索引配置(如按情感类别字段建立索引)，随后舆情分析模块不再直接读取 CSV，而是通过 SQL 查询按情感类别(如正向/负向)提取评论文本，作为后续文本挖掘输入。

在文本处理方法上，平台以 TF-IDF 为核心关键词重要性度量，并结合词频统计实现可解释的结果输出。实现流程为：从 SQLite 读取指定情感类别评论→中文分词→停用词过滤→词频统计→TF-IDF 向量化并对权重汇总排序→生成包含“切词、词频、TF-IDF 值”的关键词统计表。最终结果以多视图形式输出并可视化展示(关键词表格、词云、长度分布、地图)，形成“数据库查询 - 统计计算 - 可视化渲染”的

闭环流程。

3.1.2. 词云可视化

平台使用 pyecharts 的 WordCloud 组件, 将 TF-IDF 权重排名前 100 的关键词渲染为词云 HTML 文件, 并通过 QtWebEngine 在界面中加载展示。词云能够直观呈现高权重词项与主题特征, 适用于舆情热点词快速识别与主题聚焦分析。

3.1.3. 评论字数分布统计

平台从 SQLite 中按情感类别读取评论文本, 统计每条评论的字符长度并进行区间分桶计数, 生成条形图并配置滑动条以支持区间浏览, 分别输出正、负面评论字数分布图。该可视化用于观察文本表达强度与分布形态, 为识别异常长文本、模板化文本或高重复文本提供参考。

3.1.4. 地图可视化

平台在舆情分析中引入地域展示视图, 采用二维地图叠加正、负面数据以呈现地域分布差异, 并预留 Map3D 与 Bar3D 的扩展展示方式。地图视图用于辅助观察不同地域的舆情热度与倾向分布, 为后续结合 IP/位置字段开展更精细的空间分析预留接口。

3.2. 警情文本信息检索与统计可视化

文本信息检索实验数据采用自建的虚拟“警情文本(网络诈骗)”数据集, 以结构化表格构建, 包含内容“content”与类型“class”字段, 用于实现“按类别聚合统计(饼图/玫瑰图)以及按类别过滤回显(表格)”的检索闭环。共 1000 条文本、9 类; 类别分布: 信贷理财 239、购物消费 210、冒充类 153、平台诈骗 110、招聘兼职 133、婚恋交友 63、网络盗窃 53、冒充公检法 32、中奖诈骗 7; 描述案情文本内容长度均值约 121 个字符。

平台以虚拟数据网络诈骗类案件为例, 构建类别统计与文本检索功能, 通过饼图或玫瑰图展示类别占比, 并提供按类别查询后表格化展示的能力。饼图模块读取数据并按类别字段聚合计数生成 Pie 图, 表格模块根据输入类别进行过滤并输出 Table 组件 HTML, 界面端通过输入框与确认按钮触发更新, 并重新加载页面, 实现检索到可视化呈现的闭环。该模块可用于支持典型警情数据的快速定位与结构化浏览, 适合教学实训场景下的数据理解训练, 也可为实际业务系统提供原型参考。

3.3. 基于 PCA 与 SVM 的人脸识别可视化推理

3.3.1. 数据库构建与预处理

人脸图片来源于 LFW (Labeled Faces in the Wild) people 公开数据集, 该数据集可通过 scikit-learn 提供的接口, 在线下载与加载。本文人脸识别实验使用数据集是基于 sklearn 可获取的人脸数据(LFW people)二次整理得到的实验子集, 选择 7 位人物的人脸图像, 每人 50 张, 统一尺寸为 125×125 RGB, 文件名保持原始命名方式。在此基础上, 为验证模型对新增类别的泛化能力, 本文补充采集了同尺寸与数量(50 张)的名为 Houxue_Hui 人脸图像作为新增数据, 共 8 类。

3.3.2. PCA 降维与 SVM 分类

为验证本文方法的有效性, 构建多模型对比实验。首先对高维像素特征进行主成分分析(PCA)降维, 主成分数设置为 120 维, 并采用 `whiten = True` 进行标准化处理, 以减少特征冗余并提升计算效率。对比模型包括: Linear SVM、PCA + Linear SVM、PCA + KNN ($k=3$)、PCA + Logistic Regression、PCA + RBF-SVM, 其中本文方法采用 RBF 核函数的支持向量机, 关键参数设置为: `kernel = 'rbf'`, `C = 3`, `gamma = 0.002`。实验结果用于比较不同特征处理与分类器组合对识别性能的影响, 从而验证 PCA + RBF-SVM 在

非受控环境下的人脸识别优势。

3.3.3. 评估指标与结果可视化

模型评估使用 Accuracy、Macro-F1、Kappa、MCC、ROC-AUC 等评估指标。其中：Macro-F1 衡量各类别平均性能，适合多分类均衡性分析；Kappa 与 MCC 用于衡量分类一致性与整体相关性；ROC-AUC 反映模型对各类别的整体区分能力。在可视化推理阶段，系统对新输入图像进行预测，并将预测结果叠加到图像上生成可视化输出文件，用于 GUI 中展示模型推理效果，实现“训练评估 - 结果分析 - 前端可视化展示”的完整流程。

4. 平台实现与界面集成

4.1. GUI 组织与模块切换

平台采用主窗口承载多页面控件，通过按钮触发隐藏与显示不同模块区域，实现模块切换。界面包括：舆情文本信息分析(如图 1 所示)、警情文本信息搜索及人脸识别系统页面(如图 2 所示)，形成统一入口与一致交互。

4.2. 图表嵌入机制

平台的核心工程实现特点是将 pyecharts 生成的 HTML 图表作为中间产物，通过 QtWebEngineView 加载本地 HTML 文件，实现动态图表在桌面端的嵌入式展示。对多个 browser 控件设置透明背景并隐藏滚动条，然后加载对应 HTML 文件，如词云、条形图、表格与地图等，从而构成多视图联动界面。

4.3. 数据更新与异步执行

平台通过继承 QThread 创建 Worker 线程，在后台生成图表文件，避免阻塞主界面；在用户触发检索或切换模块时，通过重新生成图表并加载最新 HTML 实现刷新。

5. 实验与结果分析

5.1. 运行效率与可用性

平台在 CPU 环境下运行，典型任务响应时间为分钟级别，整体效果良好，主要耗时集中在文本预处理、TF-IDF 计算、图表渲染及人脸识别训练及推理等步骤，采用离线渲染 HTML 并加载的方式降低了 GUI 绘图复杂度，提升了整体可维护性。

5.2. 识别精度分析

“见表 1”所示的实验结果，各模型在人脸识别任务中的性能存在明显差异。整体而言，基于线性分类器和支撑向量机的模型表现较为稳定，最佳模型是基于 PCA 降维与 RBF-SVM 模型，其各项评估指标都是最优，准确度约为 83%，Macro-F1 为 0.8378，Kappa 为 0.804，MCC 为 0.8072，ROC-AUC 值达到 0.9818。表明传统机器学习方法在中小规模人脸数据集上仍具有较好的实用价值。建议后续研究采用卷积神经网络 CNN 等深度学习方法进一步提升精度[8]。平台给出了训练与测试的评估输出与可视化混淆矩阵生成过程，能够支持对不同数据集条件下的精度波动进行复现实验与对比分析。

5.3. 实战适配性

平台围绕舆情分析、警情信息检索与人脸识别三类典型任务，建立从数据处理到可视化展示的链路，模块分工清晰，图表形态覆盖表格、统计图与地图等多种表达方式，适合教学实训与原型验证。以下就

是警务数据分析可视化平台展示图。

Table 1. Model evaluation metrics results
表 1. 模型评估指标结果

模型	Accuracy	Macro-F1	Kappa	MCC	ROC-AUC
LinearSVM	0.8289	0.833	0.8038	0.807	0.971
PCA + LinearSVM	0.7632	0.7507	0.7281	0.7315	0.9606
PCA + KNN	0.4079	0.319	0.3182	0.3437	0.7468
PCA + LogReg	0.8289	0.8163	0.8035	0.8058	0.9718
PCA + RBF-SVM	0.8289	0.8378	0.804	0.8072	0.9818



Figure 1. Public opinion text analysis and visualization module

图 1. 舆情文本分析与可视化模块



Figure 2. Text information retrieval and machine-learning-based face recognition module (PCA + SVM)

图 2. 文本信息搜索及基于机器学习 PCA + SVM 模型的人脸识别模块

6. 讨论与改进方向

数据接入方面：本文平台定位为面向教学实训与科研演示的桌面端原型系统，现阶段实验数据规模相对较小，主要用于验证“数据导入-清洗处理-建模分析-可视化展示”的端到端流程与功能闭环，而非面向真实业务环境的大规模在线处理。

算法能力方面：人脸识别模块目前基于传统机器学习流程，具备可解释与易部署优势，但在复杂光照、姿态变化条件下存在精度上限。可在保留 PCA 与 SVM 基线模型的基础上，引入深度学习特征提取网络并采用模型压缩与加速技术以适配边缘部署。

可视化联动方面：当前采用多 HTML 视图并行展示，后续可进一步引入跨图联动选择、高亮与钻取分析机制，实现从词项到文档到地域的关联追踪，提升研判效率。

本文实验数据由两部分构成：1) 公开可获取的开源数据集(微博情感数据集、人脸 LFW 数据集)；2) 教学用途的模拟警情文本数据(不包含真实接处警系统数据、不包含个人敏感信息)。因此本文不涉及涉密警务数据处理与对外发布风险。若平台后续用于真实警务业务数据，须在单位安全管理制度下完成数据脱敏、权限控制、日志审计、网络隔离等安全加固，并通过相应的保密审查流程后方可部署使用。

7. 结论

本文设计并实现了一套面向警务业务需求的数据可视化建模处理平台。平台以 PyQt5 作为桌面端交互框架，以 pyecharts 作为交互式可视化引擎，通过 QtWebEngine 实现图表网页与 GUI 的融合展示，构建了网络舆情分析、警情文本检索与机器学习人脸识别三类模块，实现从数据预处理、建模分析到可视化呈现的完整流程。实验与应用分析表明，该平台具有良好可用性与扩展潜力，能够为警务数据可视化处理与相关教学科研提供工程化参考。后续将围绕实时数据接入、深度学习识别与可视化联动分析进一步完善。

基金项目

新疆维吾尔自治区自然科学基金资助项目，项目编号：2025D01B94；2024 年度自治区高校本科教育教学研究和改革项目普通教改项目，项目编号：XJGXJGPTB-2024077；2024 年新疆警察学院校级科研创新团队项目(数据建模与智能计算)，项目编号：XJKYKJTD202401。

参考文献

- [1] 胡轲鑫. 业务数据可视化平台的设计与实现[D]: [硕士学位论文]. 北京: 北京交通大学, 2023.
- [2] 何巍. 警务数据可视化分析[J]. 福建电脑, 2020, 36(6): 81-83.
- [3] 朱博伦, 李凌崑, 王永健, 等. 人工智能时代警务数据建设及前景展望[J]. 法制博览, 2025(30): 85-87.
- [4] 吴春生, 穆日磊, 佟晖. 公安院校警务数据建模实践教学新路径探析——以北京警察学院为例[J]. 北京警察学院学报, 2025(2): 108-115.
- [5] 刘继, 李磊. 大数据背景下网络舆情智能预警机制分析[J]. 情报杂志, 2019, 38(12): 92-97.
- [6] 李卫红, 童昊昕. 针对非平衡警情数据改进的 K-Means-Boosting-BP 模型[J]. 中国图象图形学报, 2017, 22(9): 1314-1324.
- [7] 杨长盛, 陶亮. 几种机器学习方法在人脸识别中的性能比较[J]. 计算机工程与应用, 2009, 45(4): 169-172.
- [8] 王信, 汪友生. 基于深度学习与传统机器学习的人脸表情识别综述[J]. 应用科技, 2018, 45(1): 65-72.