Research on Key Technologies of Public Opinion Monitoring and Early Warning of COVID-19 Prevention and Control in the Post-epidemic Era

Fanrong Meng[1]; Zhan Wen*[1,2]; Lanyun Chen[1]; Rui Jiang[1]; Dehao Ren[1]

[1]School of Communication Engineering, Chengdu University of Information Technology, Chengdu, 610225, China
[2]Meteorological Information and Signal Processing Key Laboratory of Sichuan Higher Education Institutes of Chengdu University of Information Technology, Chengdu, 610225, China

Corresponding Author: Zhan Wen*[1,2]

Publication Date: 2025/03/08

Abstract: In the post-epidemic era, public opinion monitoring and early warning for COVID-19 prevention and control have become vital research areas. This study aims to explore key technologies to enhance the capability of monitoring and early warning for public opinion regarding epidemic prevention and control. First, the study focuses on the collection and processing of public opinion data. By constructing a large-scale data collection system that integrates social media, news platforms, forums, and other channels, this study enables the real-time acquisition of public opinion information related to the epidemic. At the same time, natural language processing and text mining techniques are employed to clean, classify, and analyze the sentiment of large-scale text data, facilitating the extraction of valuable insights. Second, to enhance public opinion monitoring, this study introduces an emotion classification model based on deep learning. The model is compared with traditional machine learning approaches to evaluate its effectiveness in distinguishing the emotional tone of public opinion texts and analyzing individuals' attitudes and emotional responses toward the epidemic. To optimize performance, an early stopping mechanism is implemented during training to prevent overfitting, halting the process when validation loss ceases to improve after a specified number of iterations. Additionally, hyperparameter optimization is conducted using a grid search, systematically exploring various parameter combinations to identify the optimal configuration. Data balance is carefully maintained to enhance the model's predictive accuracy and robustness, ensuring reliable and high-quality results.

Keywords: *Public Opinion Monitoring, Emotion Classification, Deep Learning, Hyperparameter Optimization.*

How to Cite: Fanrong Meng; Zhan Wen; Lanyun Chen; Rui Jiang; Dehao Ren (2025). Research on Key Technologies of Public Opinion Monitoring and Early Warning of COVID-19 Prevention and Control in the Post-epidemic Era. *International Journal of Innovative Science and Research Technology*, 10(2), 1734-1741. https://doi.org/10.5281/zenodo.14979414

I. INTRODUCTION

This project studies the key technologies of public opinion monitoring and early warning for new crown prevention and control 0in the post-epidemic era, and first conducts social research to obtain hot topics. Then, data collection was carried out, and Python was used to crawl the raw data of each platform about the "post-epidemic". These raw data are preprocessed, and Python is used to remove stop words, segmentation, and labeling to divide the text into positive, negative, neutral and other operations to obtain the preprocessed data. Then, the Bert preprocessing model is introduced for vectorization, and the training set, validation set, and test set are divided. Then, the project uses the TF-IDF algorithm to extract keywords, obtain a word cloud map, visualize the public opinion monitoring results, and mine potential topics. After that, the appropriate model was selected for establishment, and the BERTlstm (BTSnet) model was selected for this project, and an early stop mechanism was established in this experiment to avoid overfitting when the model was trained, and the training was stopped when the verification loss was no longer improved after a certain number of times. Define hyperparameter combinations, grid search method, and combine these parameters until the optimal solution is selected, the best model is trained, and the best evaluation indicators are achieved. It was compared with SVM and LSTM in terms of accuracy, F1 score, precision, and recall. The BTSnet model has a huge advantage over SVM and LSTM in all indicators. Finally, the project carries out monitoring and early warning,

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14979414

and monitors the dynamic changes of public opinion according to the results of sentiment analysis. By setting thresholds or establishing early warning models, when public opinion is abnormal or exceeds a certain level, the system can issue early warning information to attract the attention and response of relevant departments or enterprises in a timely manner. On the whole, the project obtains the text through the regular operation of the data collection program, and carries out the above analysis and processing, obtains the focus of attention and emotional trend assessment report of the new crown prevention and control, and gives timely warning.

II. STRUCTURAL DESIGN OF PUBLIC OPINION MONITORING AND EARLY WARNING MODEL FOR NEW CROWN PREVENTION AND CONTROL IN THE POST-EPIDEMIC ERA

The project's scheme design identifies relevant research subjects through social research, then gathers successful rural revitalization cases through methods such as crawler programs. These cases are classified and labeled. Following data processing, the Jieba word segmentation package is utilized to eliminate stop words, followed by PCA to extract keywords and establish a f eature word library. K Nearest Neighbors, logistic regression, and regression forest models are employed for training and testing, ultimately selecting an optimal model design system.

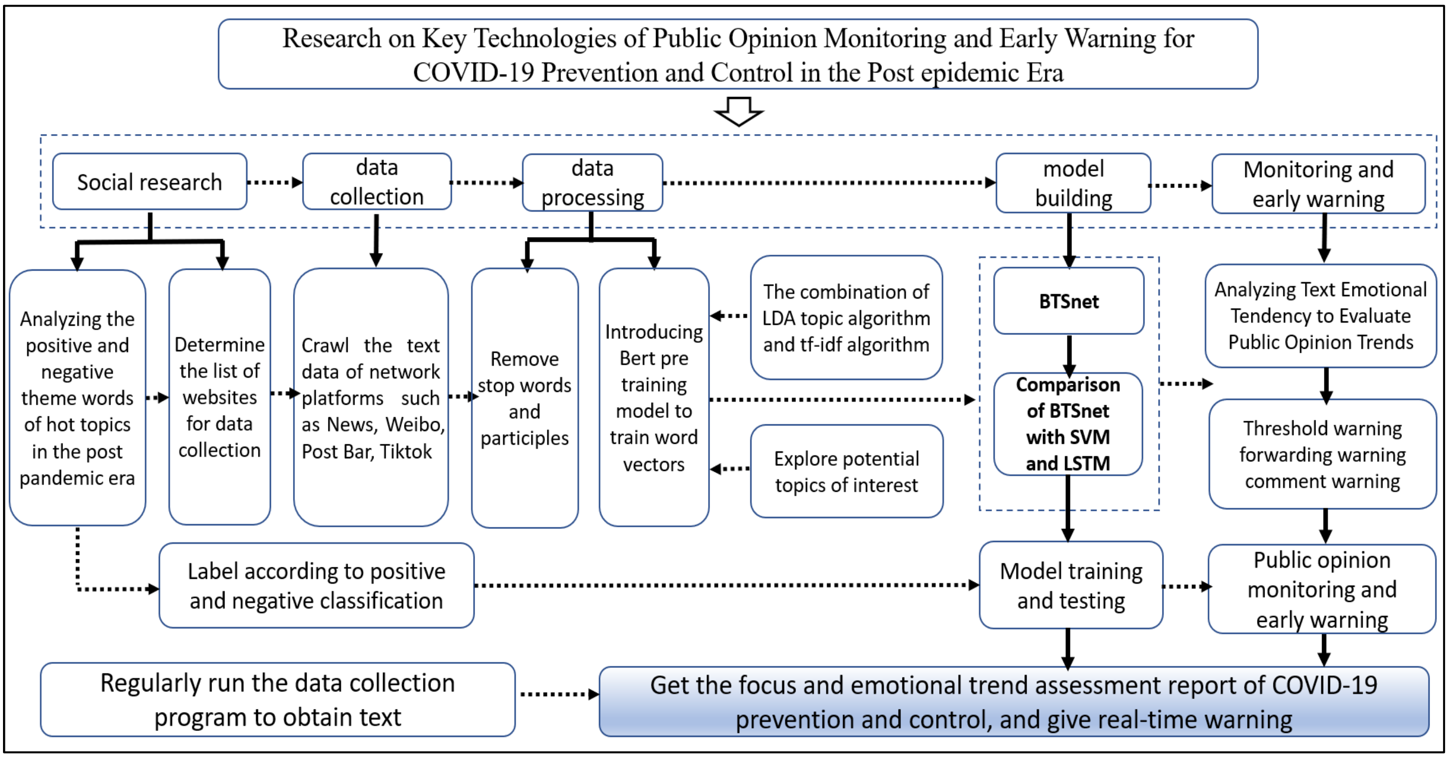

Fig 1 Structure Diagram of the Public Opinion Monitoring and Early Warning model for COVID-19 Prevention and Control in the Post-Epidemic Era

A. *Data Collection*

Through in-depth research on the types of web crawlers, web scraping strategies, and crawler technologies, the data collection was realized by using web crawlers on Weibo, Douban, Tieba, Zhihu, Post Bar, Tiktok and other platforms.

B. *Data Processing*

➤ *Text Cleaning*

In the research of deep learning public opinion monitoring and early warning technology, text data often comes from different channels, such as news, social media, forums, etc., so there are various forms of noise and cluttered information. In order to improve the accuracy and reliability of the model to public opinion, text cleaning is required.

➤ *Remove Stop Words*

The handling of stop words is an important part of text cleaning. Stop words are those words that appear frequently in text but lack actual meaning, such as "of", "is", "and", etc. By establishing a stop word list, these stop words can be removed from the text to reduce distractions and noise.

➤ *Participle*

Tokenization is the process of dividing text into meaningful words or phrases, which can help identify keywords or hot topics in the text. This processing method can extract features with more information and expressive ability, and provide better data support for model training and prediction.

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14979414

缅甸 后 疫情 旅游业 的 挑战 与 希望 并存 缅甸 后 疫情 旅游 人数 暴跌 不到 五万 网友 热议 看 了
近期 文化 和 旅游部 办公厅 发出通知 疫情 过后 恢复 全国 旅行社 及 在线 旅游 企业 经营 中国 公
在 此次 名单 中 缅甸 的 赫然 在 列 引起 了 广大 网友 的 热议 自新冠 疫情 和 政局 的 影响 以来
据 缅甸 媒体报道 疫情 后 的 今年 上半年 缅甸 仅 吸引 了 超过 45 万名 的 国外 游客 这 当中 以 中
缅甸 酒店 和 旅游部 部长 表示 缅甸 期待 中国 游客 重返 的确 中国 是 缅甸 发展 后 疫情 旅游业
一方面 缅甸 的 后 疫情 旅游 基础设施 相对 薄弱 接待 能力 有限 在 接待 外国游客 方面 尤其 是 中
另一方面 缅甸 的 后 疫情 旅游 市场 也 需要 更加 规范 和 有序 一些 不法 商贩 和 旅游 中介 利用
这些 问题 不仅 影响 了 游客 的 后 疫情 旅游 体验 也 损害 了 缅甸 后 疫情 旅游 市场 的 声誉 因此
此外 缅甸 还 需要 加强 与 周边国家 的 合作 共同 推动 区域 后 疫情 旅游 的 发展 同时 缅甸 还 可
缅甸 可以 与 泰国 老挝 柬埔寨 等 国家 进行 合作 打造 跨国 后 疫情 旅游 线路 和 品牌 实现 资源
可以 通过 举办 旅游 展览 文化 节庆 等 活动 展示 缅甸 的 独特 魅力 和 旅游 资源 此外 利用 互联
值得注意 的 是 缅甸 的 后 疫情 旅游 安全 也 是 一个 需要 关注 的 问题 近年来 虽然 缅甸 在 维护
综上所述 缅甸 的 后 疫情 旅游业 虽然 面临 一些 挑战 但 仍 有 很大 的 潜力 和 机遇 通过 加强 基
今天 看到 一个 很 有趣 但 又 有点 后怕 的 热 搜 缅甸 后 疫情 旅游业 苦 等 中国 游客 重返 但 更
有 网友 表示 还 重返 自己 心里 没 点数 啊 多少 的 中国 人 在 苦 等 他们 在 缅甸 的 亲人 回来 疫
再 来 看看 广大 网友 的 评论 如果 可以 任意 提出 一个 可 实现 的 条件 才能 让 你 下定决心 疫情
还有 网友 表示 你 可 拉倒 吧 嘎 腰子 算轻怕 是 去 了 都 回不来 了 苦 等 个 啥 疫情 后 你 缅甸
这年头 真的 要 擦亮 好 自己 的 眼睛 缅甸 到底 有 多 可怕 这 不言而喻 了 女孩子 在外 一定 要 小
疫情 过后 网上 有太多 被 骗 去 缅甸 的 例子 了 对于 自己 身边 的 人 也 要 提高警惕 最 怕 那种
再 加上 疫情 后 的 最近 电影 孤注一掷 的 上映 大家 都 对 缅甸 诈骗 团伙 有 了 一个 全新 的 认识
那些 极为 残忍 的 手段 想想 都 有点 害怕 中国 的 大好河山 那么 多非 要 去 那边 玩 相信 大多数
疫情 之后 又 有 多少 勇士 敢 去 缅甸 旅游 电信 诈骗 和 嘎 腰子 已经 成 了 对 缅甸 的 固有 印象
总之 疫情 后要 时刻 保持警惕 增强 自我 防范 意识 不要 轻易 相信 陌生人 的话 特别 是 涉及 到 人
又 一批 国家 恢复 出境 团体 游 疫情 后 的 8 月 10 日 文化 和 旅游部 办公厅 发出通知 即日起 恢
其实 在此之前 缅甸 已经 对 外国游客 放宽 了 许多 限制 包括 推行 一系列 促进 后 疫情 旅游 发展
据 缅甸 媒体报道 今年 上半年 缅甸 仅 吸引 了 超过 45 万名 的 国外 游客 这 当中 以 中国 和 泰国
疫情 过后 以 缅甸 的 邻国 老挝 为例 今年 中老 铁路 的 开通 加快 了 当地 后 疫情 旅游业 的 复苏
加之 疫情 后 的 当下 缅北 地区 愈演愈烈 的 电信 诈骗 和 嘎 腰子 事件 更是 让 众多 外国游客 对
蒲甘 也 是 缅甸 的 漆器 之 乡 当地 有 很多 手工 漆器 作坊 没 了 外国游客 传统 手艺人 的 生意 也
我们 以前 生意 很 好 的 我家 的 作坊 有 69 个人 在 干活 每天 有 上 百辆 汽车 停 在 我 的 商店 门
不止 游客 减少 不少 当地 靠 后 疫情 旅游业 为生 的 人 工作 内容 也 发生变化 另外 一家 专卖 手工
一名 当地 后 疫情 旅游业 人士 也 透露 他 以前 只 关注 接待 外国游客 的 酒店 和 旅行社 业务 现在
据 日本 经济 新闻报道 疫情 过后 蒲甘 一家 知名度 假 酒店 给 本地 游客 的 房间 价格 已经 降到

Fig 2 Word Segmentation

➢ *Text Tags*

In the research of public opinion monitoring and early warning technology based on deep learning, text tagging refers to the process of labeling or classifying key information in text. Through text tagging, public opinion texts can be classified, sentiment analysis, event identification and other tasks, so as to achieve effective monitoring and early warning of public opinion.

For the sentiment analysis task of our project, each text data can be given a positive, negative, or neutral sentiment label. This can be achieved by manually annotating or training using an existing annotated sentiment dataset. Second, you need to build a labeled dataset. A labeled dataset is a dataset that pairs public opinion text with corresponding labels. The construction of the annotation dataset requires human involvement, and each text is labeled by professional annotators according to the defined set of tags to ensure the accuracy and consistency of the labels. So as to construct a thesaurus of features.

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14979414

1	春江 水暖鸭 先知 后 疫情 旅游 行业 正在 触底 反弹 迎来 结构性 变革 这 两家 后 疫情
1	对于 旅游 行业 几乎 大部分 券商 认为 后 疫情 旅游 行业 已经 吹响 了 复苏 号角 龙头
1	其中 财通 证券 在 2022 年 12 月 13 日 研报 当中 指出 伴随 着 政策 加速 景区 疫后 业
1	首创 证券 则 表示 短期内 国内 跨省 后 疫情 旅游 将 迎来 新一轮 快速 上涨 随着 精准
1	然而 后 疫情 旅游 产业链 行业 众多 如果 从 广义 来说 则 包括 地产 餐饮 酒店 演艺 及
0	公开 资料 显示 疫情 后 在 中国 景区 开发商 及 运营商 都 融为一体 比如 峨眉山 张家界
0	疫情 过后 宋城 演艺 则 为 A股 景区 开发 板块 市值 龙头 在 港股 市场 旅游 平台 主要
1	从 国家 政策 来看 这 三年 一直 在 托底 保稳 和 促 复苏 2020 年 1 月 到 2022 年 10
1	中国 旅游 研究院 数据分析 所 所长 何琼峰 表示 说据 其 介绍 在 广东 江苏 四川 等 传
1	2023 年 后 疫情 旅游 全面 复苏 在 即 全国 各地 的 旅游 目的地 将 迎接 巨大 的 市场
1	虽然 不是 完美 的 数据 但 也 给 疫情 后 的 文旅 行业 打 了 一剂 强心针 饱受 破产 裁
1	展望 疫情 后 的 2023 春节 前及 春节 期间 的 旅游 复苏 星星之火 有望 在 全国 进一步
1	不过 每个 硬币 都 有 另一面 诚然 疫情 结束 后 全国 景区 景点 客流 迎来 大 爆发 受 疫
-1	但 另一方面 国内 后 疫情 旅游 供需 总体 不 平衡 传统 旅游 产品 供过于求 高品质 新
1	新 的 一年 疫情 后 人们 旅游 消费 信心 的 重建 需要 一个 过程 国际 政治经济 的 不断
0	后 疫情 时代 如此 背景 下 国内 后 疫情 旅游 目的地 如何 抓住 2023 旅游 市场 复苏 的
1	自 疫情 防控 新政 发布 后景 域 驴 妈妈 集团 副总裁 旅游 百人会 发起人 执惠 特约 专
1	根据 他 对 中国 各个 旅游 目的地 的 持续 考察 调研 结合 中国 优秀 文旅 企业 在 旅游
1	2022 年 12 月 7 日 国务院 出台 疫情 防控 新 十条 制约 国内 人员 流动 的 最大 的 制
1	疫情 过后 12 月 31 日　白天 新疆 阿勒泰 将军 山 滑雪场 挤满 了 来自 全国 各地 的 滑
1	从后 疫情 旅游 百人会 成员 企业 和文 旅 业界 同仁 朋友圈 中 反馈 2023 年 春节假期
1	疫情 过后 的 1 月 24 日 大年初三 西安 秦始皇陵 兵马俑 大唐 不夜城 安徽 黄山 景区
1	国家 文旅部 1 月 27 日 发布 官方 数据 疫情 之后 的 2023 年 春节假期 全国 国内 旅游
1	当然 疫情 过后 也 并 不是 全国 各地 都 均等 地 享受 到 旅游 复苏 的 红利 从 区域 来
1	也 因为 大 范围 寒潮 热门 后 疫情 旅游 目的地 之外 的 很多 温冷 旅游 目的地 进一步
1	后 疫情 时代 各地 党政 主要 领导 都 在 一心一意 谋发展 全力以赴 抓 经济 其中 拥有
0	2023 年 后 疫情 旅游 全面 复苏 在 即 全国 各地 的 后 疫情 旅游 目的地 将 迎接 巨大
1	天时 不如 地利 地利 不如 人 和 疫情 后 的 旅游 市场 与 疫情 前 的 旅游 市场 已经 发
-1	根据 笔者 旅行 考察 中国 296 个 地 市州 和 1164 个区 县市 的 经历 笔者 发现 疫情
1	如何 避免 成为 井底之蛙 如何 摆脱 后 疫情 旅游 资源 依赖症 首先 必须 认清 后 疫情
0	疫情 过后 把 专家 请进来 用 市场 思维 和 方法 诊疗 资源 依赖症 2022 年 12 月 30 日
1	在 会后 与 新疆自治区 文旅厅 阿勒泰地区 地委 行署 领导 的 单独 座谈 交流 中 笔者
0	疫情 之后 让 人员 走 出去 学习 别人 家 的 先进经验 和 做法 浙江 南湖 文化 旅游 集团

Fig 3 Feature Thesaurus

➢ *Data Segmentation and Collection Construction*

The purpose of data segmentation and collection construction is to ensure the accuracy, diversity and reliability of data in the process of training, validating and testing the model.

Data segmentation usually involves dividing an existing dataset into three parts: the training set, the validation set, and the test set. The specific division ratio can be determined based on the task requirements and the size of the dataset. This is because 70% of the training set and 30% of the test set are the easiest to find partitions in a large number of experiments [1]. Therefore, the division of this project uses 70% of the training set, 15% of the validation set and 15% of the test set. Such a division can ensure that the model can fully learn the data while evaluating the model's generalization ability through the validation set and the test set.

➢ *Word Clouds*

Keyword extraction by Term Frequency-Inverse document frequency (TF-IDF) is used for text information retrieval and mining in many domains, such as news text, social contact text, and medical text [2]. The Term Frequency-Inverse Document Frequency (TF-IDF) algorithm is also a commonly used text feature extraction method to evaluate the importance of a word to a document in a document set or corpus.

In this project, we use the TF-IDF algorithm to extract keywords in the text and generate a visual cloud word map. Through the cloud word map, we can qualitatively analyze and intuitively feel the frequency or importance of the words in the data set.

Fig 4 Word Cloud Diagram

C. Deep Learning Sentiment Classification Model

As prediction models assist policymakers in making decisions based on expected outcomes [3]. So this project establishes a deep learning sentiment classification model. The overall model architecture adopted by the project is shown in Figure 5, which first converts the input text into word vectors using the Bert model, then further extracts high-level features through the transformer structure, and finally uses the softmax activation function to output a classification result.

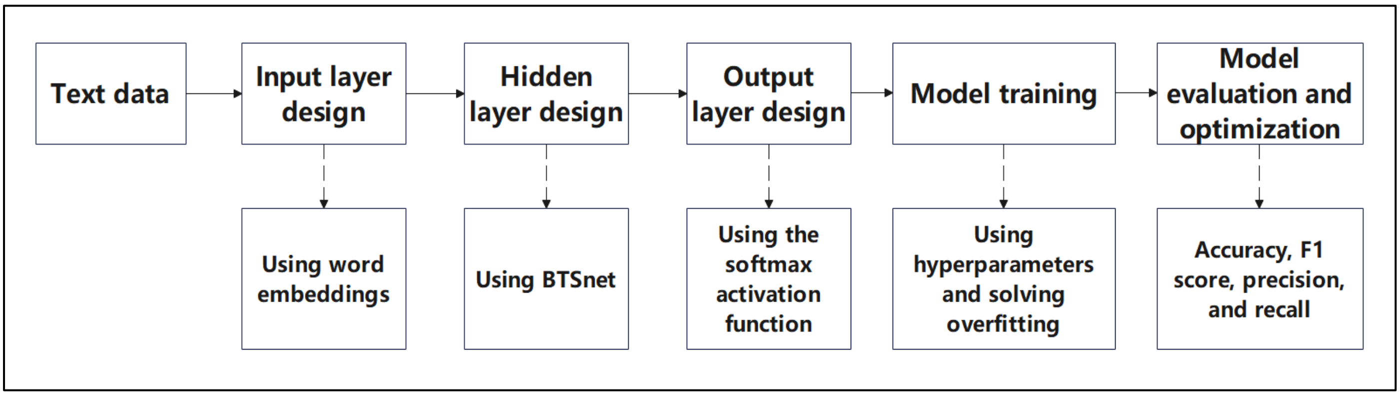

Fig 5 Flow Diagram of a Deep Learning Sentiment Classification Model

➢ *Input Layer Design*

The input layer of the deep learning sentiment analysis model is the first layer in which the model receives text data, and the design depends on the form of the text data and the architecture of the model.

We use Word Embeddings, a technique for mapping words in text as continuous vectors. Diagram is mapped to vectors that can act as input layers to convert words in text into vector form. Word Embedding is a technique that maps words into a vector space to convert discrete text data into continuous numerical representations. Word embeddings can capture the semantics and associations between words, providing a richer representation of features for machine learning models.

➢ *Hidden Layer Design*

The hidden layer of the deep learning sentiment analysis model is the core part of the feature learning and information extraction in the model. Choosing the appropriate network structure is crucial to the task of sentiment classification. The hidden layer design of the model depends on the complexity of the model and the nature of the task. Bert networks and long short-term memory networks (LSTMs) are commonly used neural networks in natural language processing, and Transformer is one of the most advanced neural network architectures. First of all, we can choose the Bert model, which stands for

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14979414

Bidirectional Encoder Representations from Transformers, which is a bidirectional pre-trained language model based on the Transformer architecture. BERT uses a multi-layer Transformer encoder as an infrastructure that enables unsupervised training on a large corpus to obtain high-quality linguistic representations. Bert has achieved SOTA results in a number of natural language processing tasks, and the Bert model is used for vectorization in this project.

In the end, we tried and chose BTSnet. This model combines the advantages of BERT, Transformer and LSTM, using BERT to extract context information, Transformer to extract sequence features, and LSTM to model sequence information. This structure can better capture the semantic and contextual relationships in the text, improving the accuracy of text understanding and prediction.

➢ *Output Layer Design*

The output layer is used to map the features learned from the hidden layer to the sentiment category and output the classification results. Typically, the softmax activation function is used so that the output value of each category is expressed as a probability, and the model selects the class with the highest probability as the prediction outcome.

➢ *Model Training*

- *Environmental Settings*

✓ Configure logs to track the training process and record key information during the training process.
✓ Set up a random seed to guarantee the reproducibility of the experiment, which helps to reproduce the results.

- *Training Process*

Overfitting: Due to lack of data, overfitting ubiquitously exists in real-world applications of deep neural networks (DNNs) [4]. Overfitting describes the phenomenon where a highly predictive model on the training data generalizes poorly to future observations [5]. When overfitting occurs, a model learns so well on the training data that it starts capturing random noise or detail in the data; It's not just the underlying pattern. As a result, such models often do not perform well on new, "unseen data."

Ways to solve overfitting: Dropout is a regularization technique used to prevent overfitting of neural networks. During training, some of the neurons in the network are randomly discarded with a certain probability so that the network does not depend too much on any one neuron. In the code, the dropout rate is set in the range of 0.1 to 0.3. Overfitting can also be solved by using an early stop strategy.

Early stop strategy: Early stop works by evaluating the performance of the model on an independent validation dataset at the end of each training cycle (or epoch). If the performance of the model does not significantly improve or even deteriorates in several consecutive epochs, the training process will be terminated early to prevent overfitting of the model.

Hyperparameters: Hyperparameters are key parameters in machine learning and deep learning models that are set before model training, rather than learned through the training process. The selection of hyperparameters plays a crucial role in the performance and behavior of the model, therefore selecting appropriate hyperparameters is crucial for achieving good model performance. The hyperparameters used in this project include learning rate, batch size, optimization metrics, etc. Optimization of hyperparameters using grid search method.

Learning rate: In the optimizer settings, the learning rate is the parameter that controls how much the model's weights are adjusted. A learning rate that is too high can lead to an unstable training process, while a learning rate that is too low can lead to a slow training speed.

Batch size: Batch size refers to the number of data samples used to calculate gradients and update network weights in each training iteration. The smaller batch size can improve the generalization ability of model training; However, it may increase the noise of the training process. Larger batch sizes can speed up the training process and improve stability, but they can also cause memory usage issues. In the code, the choice of batch size is 32 or 64.

Optimization metrics: For Chinese text sentiment analysis tasks, the model first needs to convert the Chinese text into a form that the model can understand through preprocessing and encoding (for example, encoding through the BERT model), then output the predicted probability of each category. Considering that the traditional softmax cross-entropy focuses on fitting or classifying the training data accurately [6]. Finally, the text uses Cross Entropy Loss to guide model training and improve classification accuracy. Cross-entropy loss measures the difference between the probability distribution predicted by the model and the probability distribution of the true label.

Grid Search: Grid search is a common method for model hyperparameter optimization, mainly to find the best combination of model parameters to improve the performance of the model. It is an exhaustive search technique that is achieved by systematically traversing multiple parameter combinations. Tuning of hyperparameters is performed with a grid search cross-validation approach [3]. A grid-search was used to determine the fastest and most-accurate combination of preprocessing parameters and phase-forecasting algorithms [7].

In this experiment, an early stop mechanism is implemented to avoid overfitting, and the training is stopped when the verification loss does not improve after a certain number of times. Prepare batch data and transfer the data to a device such as a GPU or CPU. Set up hyperparameter search, train, validate, and test models. Define hyperparameter combinations, grid search method, and combine these parameters until the optimal solution is selected, the best model is trained, and the best evaluation indicators are achieved.

ISSN No:-2456-2165 https://doi.org/10.5281/zenodo.14979414

➢ *Model Evaluation and Optimization*

In this experiment, we used the BTSnet deep learning model, and we compared the evaluation indicators of accuracy, F1 score, precision, and recall with SVM and LSTM, respectively.

We find that compared with traditional machine learning algorithms such as SVM (Support Vector Machine), the deep learning model using BTSnet (BERT+Transformer+LSTM) has a significant effect on accuracy, F1 score (F1_score), Precision and recall increased by 5.18, 5.64, 8.54 and 2.25 percentage points, respectively. This is because SVMs are sensitive to noise and overlapping data. If there is noise in the dataset or if there is overlap between classes, the SVM may be overfit or underfitted.

Compared with the deep learning model used for sequential data processing using LSTM (Long Short-Term Memory Network), the accuracy (F1_score), F1 score (), precision and recall rate (recall) were improved by 3.36, 1.77, 6.28 and 1.28 percentage points, respectively. This is because the gating mechanism of the LSTM can selectively forget or retain the information of the input sequence, but in some cases may over-forget important information. This can cause the LSTM to lose critical contextual information when processing some long sequences.

BTSnet makes use of Bert and Transformer models, which are better able to capture semantic relationships in text when handling natural language tasks. Bert learns rich linguistic representations through pre-training and is able to better understand the relationships between words and sentences. The Transformer model can establish a global dependency relationship through the self-attention mechanism and better handle long-distance dependencies. This gives BTSnet an edge in terms of semantic understanding. The LSTM model in BTSnet can effectively model contextual information. Through memory cells and gating mechanisms, LSTMs are able to capture long-term dependencies in input sequences and pass historical information to the current time step. This allows BTSnet to better understand contextual information in the text, which can improve the performance of classification or prediction tasks.

Table 1 Comparison of BTSnet with SVM and LSTM

Model	Evaluating Indicator			
	Accurary	**F1_Score**	**Precision**	**Recall**
SVM	80.36	79.27	78.09	81.68
LSTM	82.18	83.14	80.35	82.65
BTSnet	**85.54**	**84.91**	**86.63**	**83.93**

D. Public Opinion Monitoring and Early Warning System Design

➢ *Design of Public Opinion Monitoring Module*

Public opinion data monitoring is displayed in the form of dynamic charts through multi-dimensional analysis of processed public opinion and public opinion capture. This module is completed with the help of the ECharts charting tool, through which the current hot topic microblog can be analyzed by word cloud diagram and daily public opinion data collection.

To use ECharts to implement data statistics and analysis, you first need to obtain the data that needs to be displayed in the chart in the database in the GetCharts() method in the backend controller, and return it in Json data format. The following takes the statistical line chart of popular topics on Weibo as an example to extract the names of popular topics and the number of netizens who participated in the topic discussion under the hot topic from the database:

var list = db. MyBlog.GetChart;

return Json new total = list. Total, data = list , JsonRequestBehavior.Allow(), Get;

Second, create a div with the specified ID in the front-end page to store the line chart. It should be noted that you need to introduce the downloaded JS and ECharts theme styles into the front-end page, otherwise the page will not be able to find the definition of echart.

Get the name of the Weibo topic and the number of participants that are successfully returned in the background in JS, and the Json type that will be returned

The data is extracted and stored in an array and provided to ECharts for calling.

var topicsName = []; //Create Array 1 to store the Weibo topic name

var topicsCount = []; //Create Array 2 to store the number of topic participants

for (var i = 0; i < topics.data.length; i++){

topicsName[i] = topics.data[i]. Blog; //Extract Weibo topics

topicsCount[i] = topics.data[i]. Count; } //Extract the number of thread entries

On the premise that the DOM element has been defined, the ECHARTS object needs to be initialized first, that is:

echarts.init(document.getElementById ('topicRank'));

Next, you need to configure the basic parameters of the line chart, and you can set the title of the line chart through the "title" parameter - "Weekly Public Opinion Collection Trend"; Add a legend to the line chart via the "legend" property; The type of coordinate axis is determined by the "type" parameter of "x axis", where "value" indicates that the coordinates are numeric types, mainly for continuous data. "category" indicates that the coordinates are categorical, mainly for discrete data, in this case, you need to give "data"

The parameter assignment category. "time" indicates that the coordinates are time-based. In addition, it can be set by the "data" parameter of "series".

The data displayed on the y-axis, in this case, is topicsCount[i]; Set the color properties of the polyline by setting the "color" parameter; grid parameter to set the grid properties. After the above settings are completed, the setOption() function loads the parameters and data for the echarts object: my Chart.set Option option. Figure 10 shows the effect of the line chart of one-week public opinion data statistics.

➤ *Design of Public Opinion Early Warning Module*

The main early warning methods used in this project are keyword early warning and trend early warning, keyword early warning is to send an early warning notice once the early warning keyword appears in the monitored public opinion information, and the trend early warning is issued after the total amount of public opinion information or negative information exceeds the conventional volume range, and the user can customize the early warning type through the front-end page, such as keyword early warning, trend early warning and so on.

In the new alert task page, you can set the alert task name, alert topic keywords, or alert thresholds as needed. After the task is successfully created on the new page, the system automatically obtains the current time as the creation time of the alert keyword, and the task can be modified through the "Edit" button, and the "Modify Time" is also updated immediately. When the system captures public opinion related to keywords, the public opinion information will be classified and displayed according to the type of warning and its early warning keywords, so as to achieve the effect of real-time monitoring of sensitive keywords, which will help the public security and relevant departments to control online speech in real time, monitor target events and keywords, and ensure the healthy development of public opinion of information. The trend warning function is to create an alert name and an alert threshold, and the system will issue an alert to the user by judging whether the set threshold is exceeded. It is mainly divided into forwarding early warning, comment early warning, etc., such as when the system detects that the number of retweets of a microblog exceeds 2000, and the system monitors the participation of a topic exceeds 10000, etc., when the system captures the scene, the early warning public opinion will be classified and displayed on the real-time warning page, so as to timely monitor the heat change trend of network emergencies, which will help the public security department to control the development trend of the situation in a timely manner, achieve emergency response, and prevent public opinion events from causing adverse social impacts.

III. CONCLUSIONS

In the post-epidemic era, the research and application of key technologies for public opinion monitoring and early warning in COVID-19 prevention and control will remain crucial in advancing scientific, precise, and intelligent epidemic management. The continuous innovation and enhancement of these technologies will not only strengthen society’s ability to respond to public health emergencies but also optimize decision-making processes, improve resource allocation, and enhance early intervention strategies. By integrating advanced data analytics, artificial intelligence, and real-time monitoring, these innovations will contribute to a more resilient and adaptive public health system, ultimately safeguarding the well-being of communities and ensuring long-term preparedness for future health crises.

ACKNOWLEDGEMENTS

This work was supported by the Sichuan Science and Technology Program, Soft Science Project (No.2022JDR0076). We also would like to thank the sponsors of Industry-school Cooperative Education Project of Ministry of Education (No. 202002230010, 202101014067, 202101291013, 220500643270506)

REFERENCES

[1]. Xie Erhu. Research on Key Technologies of Data Preprocessing in Data Mining [J]. Science and Technology Bulletin (in Chinese), 2013,29 (12): 211-213. DOI: 10.13774/j.cnki. kjtb.2013.12.066
[2]. Zhuohao W, Dong W, Qing L I. Keyword Extraction from Scientific Research Projects Based on SRP-TF-IDF[J]. Chinese Journal of Electronics, 2021, 30(4): 652-657.
[3]. Sah S, Surendiran B, Dhanalakshmi R, et al. Covid-19 cases prediction using SARIMAX Model by tuning hyperparameter through grid search cross-validation approach[J]. Expert Systems, 2023, 40(5): e13086.
[4]. Je J, Ma Z, Lei J, et al. Advanced dropout: A model-free methodology for bayesian dropout optimization[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(9): 4605-4625.
[5]. Gygi, J. P., Kleinstein, S. H. and Guan, L. (2023) ‘Predictive overfitting in immunological applications: Pitfalls and solutions’, Human Vaccines & Immunotherapeutics, 19(2). doi: 10.1080/21645515.2023.2251830.
[6]. Li X, Chang D, Tian T, et al. Large-margin regularized softmax cross-entropy loss[J]. IEEE access, 2019, 7: 19572-19578.
[7]. Bigoni C, Cadic-Melchior A, Morishita T, et al. Optimization of phase prediction for brain-state dependent stimulation: a grid-search approach[J]. Journal of Neural Engineering, 2023, 20(1): 016039.