

DOI:10.20033/j.1003-7241.(2026)02-0170-04

# 基于 KNN 算法的问答气象服务技术研究与应用

陈冲,张锋,李建,陶慧青  
(浙江省气象服务中心,浙江杭州 310057)

**摘要:**由浙江省气象部门推出的“智慧气象”App,依托权威气象数据源与精细化数值预报产品,构建了专业化的气象服务体系,自发布以来持续获得良好的用户口碑与行业认可。随着社会对气象服务的需求向精准化、个性化方向不断深化与细分,叠加气象科学技术的快速发展,为有效回应用户在不同场景下多样化的气象信息获取诉求,本研究设计并实现了一种基于 K 近邻算法(K-nearest neighbors, KNN)的气象问答技术。该技术在保留原有 App 功能架构的基础上,创新性地整合了以 KNN 为核心的气象问句分类与相似匹配机制,实现了对用户自然语言问句的精准理解与自动化应答。该技术通过构建历史训练语料的特征向量空间,依据问句语义近邻实现高效分类与答案推送,显著提升了气象服务的交互性与智能化水平。实证研究及实际应用反馈表明,引入 KNN 问答模型后,系统在响应准确率、服务响应效率及用户综合满意度等方面均实现一定提升,为气象服务智慧化升级提供了有效技术支撑,具备良好的推广价值与应用前景。

**关键词:** App; 气象服务; KNN; 问答; 分类; 智能

**中图分类号:** TP399; P409

**文献标志码:** A

**文章编号:** 1003-7241(2026)02-0170-04

## Research and application of question answering meteorological service technology based on KNN algorithm

CHEN Chong, ZHANG Feng, LI Jian, Tao Huiqing

( Zhejiang Meteorological Service Center, Hangzhou 310057, Zhejiang, China )

**Abstract:** Developed by the Zhejiang Meteorological Department, the "Smart Meteorology" App utilizes authoritative data sources and refined numerical forecasting products to deliver the professional meteorological service. Since its launch, it is earned widespread user acclaim and industry recognition. In response to growing and diversifying user needs for scenario-specific information, this study introduces a meteorological QA system based on the K-nearest neighbors (KNN) algorithm. Without altering the original App framework, the model integrates a KNN-driven mechanism for the classification and matching of weather-related questions, enabling precise understanding and automated replies to natural language queries. By mapping historical training data into a feature vector space and leveraging semantic similarity, the method enhances both interactivity and intelligence of the service. Experiments and practical applications demonstrate that the KNN-based QA model significantly improves response accuracy, efficiency, and user satisfaction. It supports meteorological service intelligent upgrading effectively, proving its value for broader promotion and application.

**Keywords:** App; meteorological service; K-nearest neighbors; QA; classification; intelligence

随着社会发展和科技进步,人们对气象信息的需求日益多样化和个性化。传统的气象服务主要以简单的天气监测与预报形式呈现,难以满足当下人们在复杂场景下的服务需求<sup>[1-5]</sup>。因此,问答式气象服务应运而生,其强大的互动性能更精准地满足人们个性化需求,在不同场景下提供更具针对性的气象服务。

文献[6]提出了一种基于 LLM 的台风灾害知识问答技术,融合领域微调和检索增强生成两种技术范式,以提高灾害信息检索的精度。文献[7]利用 5G RCS 设计气象服务的分级结构,开发了适用于特定场景、能理解交互意

图的聊天机器人。文献[8]对智能问答系统的发展进行介绍,对数据集、评测指标和各类问答系统的主流技术进行阐述。文献[9]提出了一种基于动态规划算法的中文问答检索方法,将多步骤问答过程统一,把形式语言中的状态机思想应用于实体识别和问答检索过程并完成建模。文献[10]通过对智能问答机器人现状分析,开发了基于知识库的智能机器人问答系统。文献[11]通过问答系统与传统搜索引擎优劣对比,阐述了基于深度学习的自动问答系统研究进展。

上述研究为问答气象服务技术的研究与实现提供了

收稿日期:2025-01-03;录用日期:2025-02-08

基金项目:浙江省气象局科技计划项目(2023YB07);浙江省科技厅项目(LGF21D050001);浙江省科技厅项目(LZJMY24D050004)

作者简介:陈冲(1990—),男,硕士,工程师,研究方向:气象信息技术。

引用本文:陈冲,张锋,李建,等. 基于 KNN 算法的问答气象服务技术研究与应用[J]. 自动化技术与应用, 2026, 45(2): 170-173. (CHEN Chong, ZHANG Feng, LI Jian, et al. Research and application of question answering meteorological service technology based on KNN algorithm[J]. Techniques of Automation and Applications, 2026, 45(2): 170-173.)

借鉴。本文提出基于 KNN 算法的问答气象服务技术,通过分析气象用户对话内容,提取对话内容特征,快速分类用户意图,理解用户需求并做出应答,为用户提供更加精准、实用的气象服务。

### 1 资料来源

#### 1.1 用户数据来源

数据来源于浙江省气象服务中心对外服务出口的用户信息,包括用户所在位置,手机号,用户标签以及用户与“智慧气象”APP 语音助手对话信息,并对其进行脱敏处理初步形成基础用户信息储备库。

#### 1.2 气象数据来源

气象数据包括实况监测与预报数据。监测数据为多要

素的自动气象站信息插值网格,其中分钟监测数据更新频次 1 次/10 min,小时监测数据更新频次 1 次/h,网格大小 5 km×5 km。预报数据采用智能网格预报的多要素资料,预报更新频次 24 h 内为 1 次/h,24~72 h 为 1 次/6 h,72~168 h 为 1 次/12 h,网格大小 5 km×5 km。

### 2 技术实现

#### 2.1 技术架构

基于 KNN 算法的问答气象服务技术由数据层、算法层和应用层组成。数据层处理用户和气象数据;算法层基于 KNN 算法实现对用户问题的分类处理与回答形成;应用层为用户提供交互界面,接收用户的问题并返回答案,如图 1。

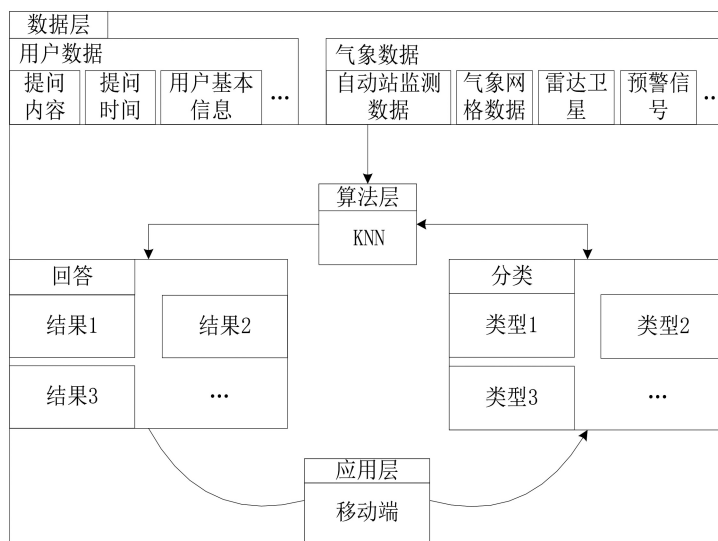


图 1 技术整体架构

Fig. 1 Overall technical architecture

#### 2.2 具体实现

##### 2.2.1 数据准备

收集和整理数据,包括用户提问内容、提问时间与基础信息等用户数据,监测与预报网格、雷达卫星与预警信号等气象数据。对数据进行预处理,去除异常值与数据标准化等,以保证数据的质量和可靠性,然后进行数据脱敏。将数据划分为训练集和测试集,用于算法的训练和评估,通过实践经验确定具体训练集,确保训练集能代表各种可能的气象情况和问题类型,涵盖不同的气象参数组合。

##### 2.2.2 特征提取

通过前期研究气象类问题文本,利用分词和去除停用词等手法,从用户与气象数据中提取特征,构建词汇表为

$$V = \{w_1, w_2, w_3, \dots, w_n\} \quad (1)$$

式中,  $w$  表示特征词汇,例如天气、温度、湿度、风速与降水量等的气象特征词汇以及地理信息与时间等常规类特征词汇。

##### 2.2.3 距离度量

选择合适的距离度量方法能提升计算问题之间相似效率<sup>[12-15]</sup>,鉴于气象类问题文本特征的实际情况以及

即时性问答的现实需求,本文将欧氏距离作为距离度量主要方法。

##### 2.2.4 问题分类

当用户提出问题时,将问题转换为特征向量表示;计算用户问题与训练集中每个问题的距离;选择距离最近的  $K$  个问题,最佳  $K$  值由气象类问题的泛化情况,以及权衡算法的性能和准确性确定;根据最近邻问题的类别分布与多数投票法,最终确定用户问题的类别,流程如图 2。

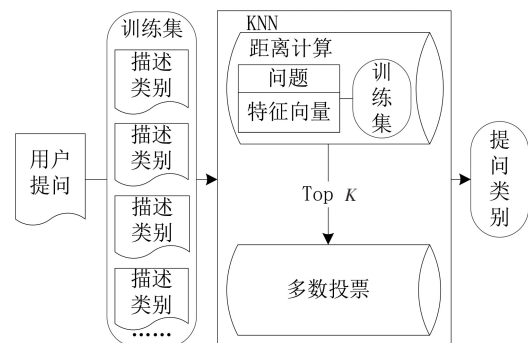


图 2 问题分类整体流程

Fig. 2 Overall process of question classification

对于问题文本  $Q = (q_1, q_2, \dots, q_m)$ , 利用 TF-IDF 算法<sup>[16]</sup>, 词  $q_i$  的词频为

$$tf_{q_i} = \frac{\text{count}(q_i)}{\sum_{j=1}^m \text{count}(q_j)} \quad (2)$$

式中, 包含词  $w$  的问题数为  $n_w$ , 词  $w$  的逆文档频率  $\text{idf}_w = \log \frac{N}{n_w}$ , 则问题中词  $q_i$  在特征向量中的权重为  $tf_{q_i} \times \text{idf}_{q_i}$ , 最终特征向量  $\mathbf{X} = (x_1, x_2, \dots, x_n)$ , 其中  $x_i = \sum_{q_j \in Q, q_j = w_i} tf_{q_j} \times \text{idf}_{q_j}$ 。

若接收到的用户问题特征向量为  $\mathbf{Y}$ , 训练集中某问题的特征向量为  $\mathbf{Z} = (z_1, z_2, \dots, z_n)$ , 通过欧氏距离计算方法, 距离为

$$d(\mathbf{Y}, \mathbf{Z}) = \sqrt{\sum_{i=1}^n (y_i - z_i)^2} \quad (3)$$

根据相似度计算结果, 选择相似度最高的  $K$  个问题作为最近邻。最后, 统计这  $K$  个最近邻中属于每个类别的问题数量, 将用户问题分类为数量最多的那个类别。

研究将用户问题分为三类: 天气状况类、气象预警类和气象科普类, 如表 1。

表 1 训练集样例

Tab. 1 Training set examples

问题描述	类别
特征向量 $\mathbf{Z}_1$	天气状况类
特征向量 $\mathbf{Z}_2$	气象预警类
特征向量 $\mathbf{Z}_3$	气象科普类
.....	.....

具体步骤:

**Step 1** 接收到用户提出新的问题, 例如问题“明天会不会下雨?”, 设为  $\mathbf{Y}_{\text{new}}$ 。

**Step 2** 计算  $\mathbf{Y}_{\text{new}}$  与训练集  $\mathbf{Z}_i$  中每个问题的欧氏距离, 计算公式为

$$d(\mathbf{Y}_{\text{new}}, \mathbf{Z}_i) = \sqrt{\sum_{j=1}^n (y_{\text{new},j} - z_{ij})^2} \quad (4)$$

**Step 3** 选择距离最近的  $K$  个训练样本, 并找到的这  $K$  个最近样本中分别属于天气预报类、气象灾害预警类及气象科普类个数。

**Step 4** 统计这  $K$  个样本中所属类别出现的频率。例如天气预报类频率最高, 则将新问题“明天会不会下雨?” 归类为天气预报类。

### 2.2.5 答案生成

通过以上小节步骤, 实现了对用户问题的具体分类。根据用户问题的类别, 从前期已完成分类标签的气象知识库中快速检索出相关答案, 从而提高问答气象服务的检索效率和准确性。

## 3 实验与分析

实验与分析围绕用户问题分类的准确性与系统应用

效果进行评估, 用户对话问题数量约为 100 000 条。

### 3.1 准确性评估

本文采用抽样调查法, 选取对话问题总量 15% 作为样本, 通过 2.2 节计算得到了分类结果集合, 抽稀后分类效果如图 3, 横坐标为样本数, 纵坐标为分类编号, 其中零散点为未正确归类样本, 计算有 86.1% 的训练样本被正确分类, 分类结果如图 4, 由此初步证明训练集的可靠性以及 KNN 算法作为用户气象问答分类手段的效果良好。

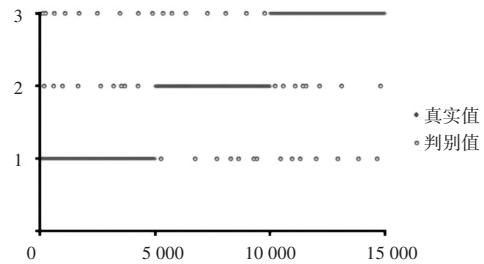


图 3 用户分类分布

Fig. 3 User classification distribution

真实值	预测组成员信息			总计	
	1	2	3		
原始	计数	4 278	213	509	5 000
		334	4 450	216	5 000
		303	513	4 184	5 000
%		85.6	4.3	10.2	100.0
		6.7	89.0	4.3	100.0
		6.1	10.3	83.7	100.0

a. 正确地对 86.1% 个原始已分组个案进行了分类。

图 4 分类结果

Fig. 4 Classification results

然后利用上述分类办法对剩余 75 000 条对话问题批量分类, 通过计算得到这部分用户的分类结果, 在分类结果中随机再选取 10% 的样本, 共计 7 500 条样本进行可靠性验证, 验证方法为人工判别是否正确回答并赋值, 将验证结果绘制成 ROC 曲线 (如图 5), 计算 AUC 达 84%, 处于 70%~90% 区间, 即对抽取的样本分类效果较好。

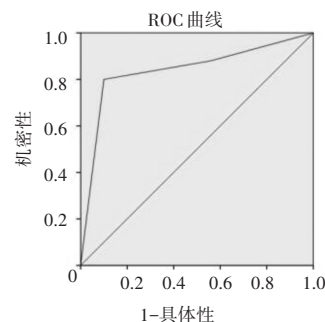


图 5 ROC 曲线

Fig. 5 ROC curve

综上, 基于 KNN 算法的问答气象服务技术作为用户对话问题分类的手段基本是可靠的。

### 3.2 应用效果评估

该技术于 2023 年 1 月应用于浙江省气象服务中心开发的“智慧气象”APP 中, 截至 2024 年 12 月, “智慧气象”

APP 月均活跃度约为 46 万,较以往有明显提升,对话服务效果如图 6。

根据服务出口的用户活跃度数据库查询,集成了基于 KNN 算法的问答气象服务技术的“智慧气象”APP,自

2022 年至今,用户活跃度总体稳步提升,尤其在 2023 年应用该技术后用户活跃度提升显著(如图 7),由此可见,应用基于 KNN 算法的问答气象服务技术对增进气象服务出口用户粘性与用户活跃度产生了积极影响。



图 6 智能问答服务效果图

Fig. 6 Effect diagram of intelligent QA service

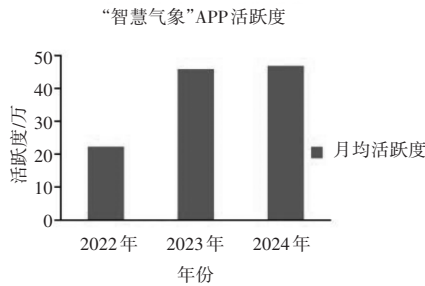


图 7 “智慧气象”App 活跃度

Fig. 7 Activity of the "Smart Meteorology" App

### 4 结论

本文研究了基于 KNN 算法的问答气象服务技术,利用 KNN 算法优势对气象类问题数据进行处理,完成了个性化气象问答服务,整体采用数据获取、特征提取、问题分类及答案生成四步走方案,初步实现了气象服务信息的快速交互。实践表明,应用该技术的系统取得了较好的社会效益。未来,将持续扩大训练集,优化算法和系统功能,进一步提升气象问答服务的准确性与可靠性,为用户提供更优质的气象服务。

### 参考文献

[1] SHI W, WEN S M, ZHANG J, et al. Extreme weather as a window: exploring the seek and supply of climate change information during meteorological disasters in China [J]. *Advances in Climate Change Research*, 2023, 14(4): 615-623.

[2] BRUNET G, PARSONS D B, IVANOV D, et al. Advancing weather and climate forecasting for our changing world [J]. *Bulletin of the American Meteorological Society*, 2023, 104(4): 909-927.

[3] 李燕. 基于局域网的气象数据智能监控系统[J]. *自动化技术与应用*, 2024, 43(2): 76-80.

[4] 陈冲, 张锋, 朱潜, 等. 基于 CFAP 架构的气象服务系统设计与实现 [J]. *微型电脑应用*, 2024, 40(1): 32-35.

[5] 李建, 张锋, 陈冲, 等. 基于 AI 智能语音交互气象服务的研究与应用 [J]. *计算机技术与发展*, 2020, 30(12): 197-202.

[6] XIA Y, HUANG Y, QIU Q, et al. A question and answering service of typhoon disasters based on the T5 large language model [J]. *ISPRS International Journal of Geo-Information*, 2024, 13(5): 165.

[7] QU H, TANG W, LI Y, et al. Weather China: A 5G RCS solution for meteorological service [C]//2022 3rd Asia Conference on Computers and Communications (ACCC). Shanghai, China: IEEE, 2022: 84-92.

[8] 姚元杰, 龚毅光, 刘佳, 等. 基于深度学习的智能问答系统综述 [J]. *计算机系统应用*, 2023, 32(4): 1-15.

[9] SHOUYU L, WEI J, YUANWEI Y, et al. Question answering retrieval method for knowledge-based with dynamic programming algorithm [C]. 2023 2nd International Conference for Innovation in Technology (INOCON). Bangalore, India: IEEE, 2023: 1-5.

[10] LU J, LI S. Design of question answering system of intelligent customer service robot based on knowledge base [J]. *Computer Science and Application*, 2019, 9(11): 2098-2104.

[11] ZHAO S, JIN Z. Research progress of automatic question answering system based on deep learning [J]. *Open Access Library Journal*, 2020, 7(6): 1-6.

[12] PRISCILA S S, RAJEST S S, REGIN R, et al. Classification of satellite photographs utilizing the K-nearest neighbor algorithm [J]. *Central Asian Journal of Mathematical Theory and Computer Sciences*, 2023, 4(6): 53-71.

[13] 吴浩, 罗少辉, 李颖昕, 等. 基于聚类挖掘的科技数据价值动态监测方法 [J]. *自动化技术与应用*, 2024, 43(2): 81-84, 106.

[14] 王婷, 张震宇, 任腾云, 等. 基于聚类算法的多维数据库一致性检测与恢复方法研究 [J]. *自动化技术与应用*, 2024, 43(3): 108-111, 119.

[15] ZHANG S, LI J, LI Y. Reachable distance function for KNN classification [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2022, 35(7): 7382-7396.

[16] 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.