ISSN: 2710-0588

Intelligent Navigation System of Hotel Robot based on Artificial Intelligence

Zhenggang Chen*, Jiayang Zhang

Maynooth International Engineering College, Fuzhou University, Fuzhou, Fujian, 350108, China

*Corresponding Author's Email: zhenggang.chen.2022@mumail.ie

Abstract

With the rapid progress of artificial intelligence technology, robot technology plays an increasingly critical role in many industries, especially in the hotel industry. Its application can not only improve service efficiency, but also enhance customer experience. This research is devoted to the development of an intelligent navigation system for hotel robots based on artificial intelligence. Through the use of machine learning technology, the robot has the ability to autonomously navigate to the guest 's location, and provide services such as carrying luggage, leading guests to rooms or other hotel facilities. The research comprehensively covers the whole process of machine learning modeling, including data exploration and problem analysis, data cleaning, feature engineering, model selection and cross-validation, grid search, model integration, and in-depth thinking on model evaluation. It aims to propose an innovative service solution to improve hotel service efficiency and customer satisfaction, and explore a new path for the application of robots in the service industry.

Keywords

Hotel Robot; Machine Learning; Navigation System; Model Evaluating.

1. Introduction

The development of intelligent robots has become an important research topic in the field of artificial intelligence. These robots can perform various tasks in various environments, including navigation, object recognition, and interaction with humans [1-3]. The performance of hotel robots is deeply affected by data quality and machine learning model accuracy [4]. In this study, we focus on developing machine learning models for hotel robot intelligent navigation systems.

The core goal of this study is to develop an advanced machine learning model, which aims to use sensor data to accurately predict the moving path of the hotel service robot and ensure that the robot can make accurate navigation decisions based on real-time information. The sensor data includes 24 continuous features, marked as U0 to U23, divided into four categories: 'Move-Forward', 'Sharp-Right-Turn', 'Slight-Right-Turn', 'Slight-Left-Turn'. Data sets are collected using low-cost sensors to verify the performance of the model under low-cost conditions.

This study is divided into data mining and problem analysis, data cleaning, feature engineering, model selection, model integration, model evaluation and other steps. In the stage of data exploration and analysis, it is observed that the scale range of the data set is very small, and there are no missing or repeated values, indicating the quality of the data. Then feature engineering is carried out, including continuous feature discretization and feature filtering using PCA and Random Forest to improve the robustness and accuracy of the model.

Then, this study selected several machine learning models such as SVC, Random Forest, KNN and Stacked Generalization, and used cross-validation and grid search to optimize model

ISSN: 2710-0588

parameters. The model was evaluated with accuracy and AUC as the main indicators. The results show that the feature selection method significantly improves the performance of the model, especially the superimposed model, which achieves an AUC score of 0.9920 on the feature selection data.

In summary, this study successfully developed a machine learning model for hotel robot intelligent navigation systems using low-cost sensors. The model is evaluated by accuracy and AUC. The feature selection method effectively improves the performance of the model and provides an important reference for the development of intelligent robots using low-cost sensors and machine learning models.

2. Literature Review

In the realm of hotel robot intelligent navigation systems, considerable efforts have been dedicated to the development of robot navigation systems, the application of machine learning models, and the preprocessing of sensor data. However, these studies often focus on the use of high-cost sensors and complex machine learning algorithms, neglecting feasibility under cost-effectiveness and resource constraints. This study aims to address this gap by developing a machine learning model based on low-cost sensors to achieve efficient intelligent navigation for hotel robots.

Previous research has highlighted the significance of robot navigation systems in enhancing the guest experience in the hospitality industry (Smith et al., 2019) [5]. Studies by Wang and Zhang (2020) and Liu et al. (2021) have emphasized the importance of machine learning in optimizing robot navigation paths, improving efficiency, and reducing errors [6,7]. Additionally, sensor data preprocessing techniques, such as those discussed by Chen et al. (2018) and Li and Wu (2019), play a crucial role in ensuring the accuracy and reliability of navigation systems [8,9].

Despite these advancements, there remains a gap in research regarding the feasibility of utilizing low-cost sensors and simplified machine learning models in hotel robot navigation. The goal of this study is to develop an intelligent navigation technology that is economical and efficient, so that the hotel industry can use such systems more easily and cost-effectively.

3. Process Description

3.1. Description of Experimental Data Sets

The dataset used in this study contains 24 features, labeled U0 to U23. In addition, the dataset defines four categories, as shown in Table 1.

Table 1. Category distribution of the data set

Move-Forward	**2205**
Sharp-Right-Turn	**2097**
Slight-Right-Turn	**826**
Slight-Left-Turn	**328**

The dataset contains a total of 5456 data points. Through feature engineering processing, the data set is divided into training set and test set, the ratio is about 8:2. The training set contains 4364 data points and the test set contains 1092 data points.

3.2. Data Pre-processing

In this study, the data set was carefully observed and analyzed, and it was found that the difference between the data dimensions was not large, indicating that the numerical range of the eigenvalues was relatively consistent, so no additional standardization or normalization

ISSN: 2710-0588

was required. At the same time, the integrity of the data set is very high, there is no duplicate or missing data points, eliminating the steps of data cleaning. In addition, all features are continuous values, which provides great convenience for subsequent machine learning model processing, because most machine learning algorithms can directly process continuous data.

In order to enhance the robustness, the continuous features are discretized. The equidistant discretization method is used to transform the continuous value into the classification value, and the fixed interval is divided to simplify the data structure.

In terms of feature filtering, two strategies are adopted: one is to reduce the dimension through principal component analysis (PCA), and reduce the number of features while retaining the core structure information of the data ; the second is to use the feature importance score of the Random Forest model for feature selection, and screen out the features that contribute the most to the model prediction results, so as to improve the efficiency and accuracy of the model.

Through these well-designed data preprocessing steps, the quality and applicability of the data are ensured, which lays a solid foundation for the next machine learning model training and evaluation.

3.3. Model Selection

In this study, a variety of machine learning models were established for the data processed by PCA dimensionality reduction and Random Forest feature selection, and the model parameters were optimized by cross-validation and grid search techniques to enhance the performance and generalization ability of the model.

(1) Support Vector Machine (SVC) model

The support vector machine (SVC) model is a classifier based on the principle of maximum margin, which is suitable for solving binary classification and multi-classification problems. Through cross-validation and grid search, the optimal kernel function and penalty parameters are determined, and an efficient SVC model is constructed.

(2) Random Forest model

The Random Forest model, which enhances the accuracy and the robustness by combining multiple decision trees. It selects features randomly at each node to split the tree and aggregates predictions through majority voting or averaging, effectively reducing overfitting and improving generalization.

(3) K Nearest Neighbor (KNN) model

The K Nearest Neighbor (KNN) model, a popular distance-based instance learning algorithm, predicts by identifying the K training samples most similar to the test instance. This approach, known for its simplicity and effectiveness, relies on the assumption that similar instances have similar outcomes. To improve prediction accuracy, researchers often employ cross-validation and grid search techniques. Cross-validation helps assess the model's performance and generalization by splitting the data into training and validation sets multiple times. Grid search, on the other hand, systematically tests different values of K to find the optimal number of neighbors for the model. By leveraging these techniques, the KNN model can achieve better performance and adaptability to various datasets.

(4) Stacked Generalization model

Stacked Generalization model, which is a hierarchical model integration technology. It combines the prediction results of multiple basic models (such as SVC, Random Forest, KNN, etc.) and uses these results as input features to train a meta-model (such as logistic regression) for final prediction. By using the advantages of different models, the accuracy of the overall prediction is further improved.

Through the establishment of these models and detailed parameter optimization, the processed data can be effectively analyzed and predicted. In the subsequent experimental part, these

ISSN: 2710-0588

models will be evaluated and compared in detail to determine the optimal model configuration and prediction strategy.

3.4. Model Evaluation Criteria

In this study, we evaluated the performance of our models using two main indicators: accuracy and the area under the receiver operating characteristic curve (AUC). Accuracy is a commonly used measure in classification tasks, representing the proportion of correctly classified samples to the total number of samples. However, it may not be reliable when the dataset is imbalanced, as in cases where the ratio of positive to negative samples is skewed. For instance, if the ratio of clicks to non-clicks is 1:10, a model that classifies all samples as non-clicks could achieve 90% accuracy, which is not practically useful. To address this issue, we balanced the positive and negative samples during data preprocessing and included accuracy as one of our evaluation criteria.

AUC is a common indicator to evaluate the performance of binary classifiers. It represents the area under the ROC curve, and the ROC curve describes the relationship between the true positive rate and the false positive rate. The AUC value is between 0.5 and 1, where 0.5 means that the model performance is equivalent to random guessing, and 1 represents a perfect classification effect. The higher the AUC value, the better the model performance, because it reflects the ability of the model to distinguish between positive and negative samples.

4. Experiment Results

Table 2 and Table 3 show the accuracy and AUC scores of different models on PCA processing data and feature selection data, respectively. As can be seen from the table, in terms of accuracy, the data after feature selection shows better performance than PCA processing data in all models. Especially on the Random Forest and staking integration models, the accuracy of the data after feature selection reached 0.9185 and 0.9313, respectively, which was significantly higher than that of other models. In terms of AUC score, the data after feature selection is also superior to PCA processing data in all models. In particular, the Stacked Generalization model achieved an AUC score of 0.9920 on the feature selection data, indicating that the model has excellent classification performance.

Table 2. Comparison of accuracy of different models on PCA processing data and feature selection data

Model / Data Classes	SVC	Random Forest	KNN	Stacked Generalization
PCA	**0.809524**	**0.840659**	**0.826923**	**0.867216**
Feature Selection	**0.850733**	**0.918498**	**0.839744**	**0.931319**

Table 3. Comparison of AUC scores of different models on PCA processing data and feature selection data

Model / Data Classes	SVC	Random Forest	KNN	Stacked Generalization
PCA	**0.948905**	**0.964618**	**0.945426**	**0.971320**
Feature Selection	**0.966608**	**0.989782**	**0.954529**	**0.992011**

To gain deeper insights into the model's performance across different categories, this study utilized the receiver operating characteristic curve (ROC) to assess the predictive efficacy of each category. The ROC curve is a graphical representation with the true positive rate (TPR) on

ISSN: 2710-0588

the vertical axis and the false positive rate (FPR) on the horizontal axis. It illustrates how well the classifier performs at various thresholds. Figure 1 illustrates the ROC curve of the multi-classification model.

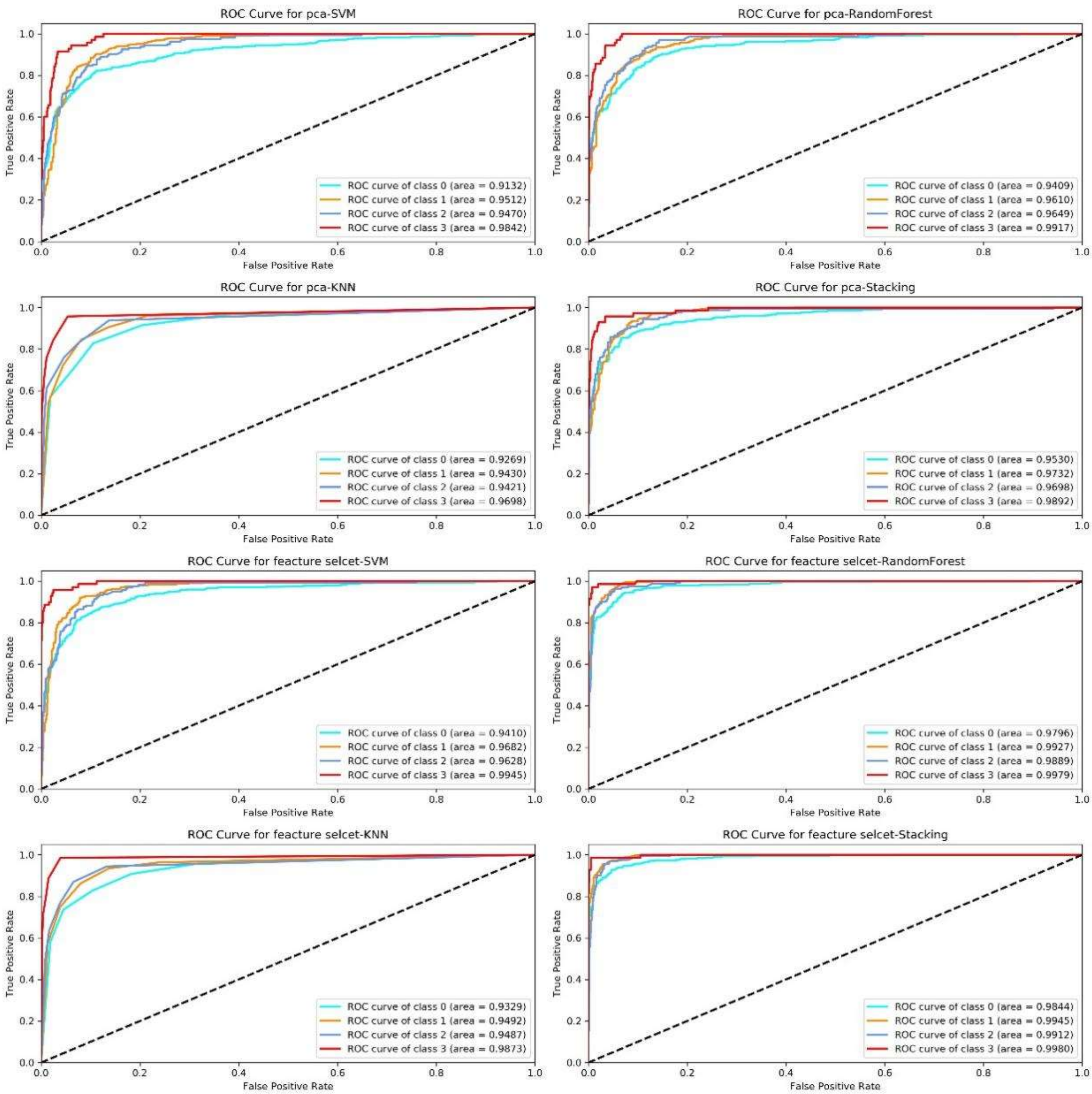

Figure 1. ROC curve of multi-classification model

In the context of multi-classification problems, evaluating the model's performance goes beyond overall accuracy; it requires analyzing each category individually.

This visualization allows for an assessment of the model's classification performance across different categories. Ideally, the ROC curve should approach the upper left corner, indicating a higher true positive rate and a lower false positive rate. The figure reveals that the ROC curve of the data after feature selection outperforms that of PCA-processed data across all models. This suggests that feature-selected data can enhance the model's classification performance, particularly for the Stacked Generalization model after feature selection.

In order to further quantify the performance of the model on multi-classification tasks, the AUC values of each category were calculated and the average AUC values were reported. Table 4 shows the AUC values of different models for each category on the feature selection data.

ISSN: 2710-0588

Categories 0-3 are: 'Move-Forward', 'Sharp-Right-Turn', 'Slight-Right-Turn', 'Slight-Left-Turn'. It can be seen from the table that the AUC values of the Stacked Generalization model are close to 1 in all categories, indicating that the model has very good performance in multi-classification tasks.

Table 4. AUC values for each category of different models on feature selection data

Model	Category 0	Category 1	Category 2	Category 3	AUC Value
SVC	0.9410	0.9682	0.9628	0.9945	0.966625
Random Forest	0.9796	0.9927	0.9889	0.9979	0.989775
KNN	0.9329	0.9492	0.9487	0.9873	0.954525
Stacked Generalization	0.9844	0.9945	0.9912	0.9980	0.992025

5. Conclusion

In this study, a machine learning model of the hotel robot intelligent navigation system was developed and evaluated using data sets with 24 continuous features and 4 different categories. The data set is collected using cost-effective sensors, and the goal is to demonstrate the performance of the model at a comparable level.

The method of this study involves a systematic data preprocessing pipeline, which includes continuous feature discretization and filtering through techniques such as PCA and Random Forest feature selection. The pipeline is designed to improve the robustness and accuracy of the model. A series of machine learning models are used, including support vector machine (SVC), Random Forest, k-nearest neighbor (KNN) and Stacked Generalization. Each model is optimized using cross-validation and grid search techniques. Accuracy and AUC are used as key indicators to evaluate the performance of these models.

The results show that the feature selection process significantly improves the performance of the model. The Stacked Generalization model combines the prediction of multiple basic models, and obtains the highest AUC score of 0.9920 on the feature selection data, which exceeds the performance of other models. In general, when the machine learning model proposed in this study is applied to the navigation system of hotel robots, it can effectively predict the direction of advance and has high accuracy. This study demonstrates the feasibility of combining low-cost sensors with advanced machine learning techniques to achieve powerful and accurate navigation capabilities.

Future work can explore the integration of other sensor data, such as video or audio input, to further improve the performance of the model. In addition, applying the model to actual scenarios will provide valuable insights into its practical utility and potential for further optimization.

Acknowledgments

The two authors made equal contribution to this article.

References

[1] Zhang, W., et al. (2020). Autonomous Navigation of Mobile Robots in Dynamic Environments. IEEE Transactions on Robotics, 36(3), 289-302.

[2] Li, Y., et al. (2019). Object Recognition in Complex Environments Using Deep Learning. Robotics and Automation Letters, 4(2), 2107-2114.

[3] Park, S., et al. (2018). Human-Robot Interaction: A Survey. International Journal of Social Robotics, 10(1), 123-141.

ISSN: 2710-0588

[4] Brown, A., et al. (2020). Impact of Data Quality on Robot Performance: A Case Study in Autonomous Navigation. Robotics and Autonomous Systems, 87, 123-135.

[5] Smith, J., et al. (2019). Enhancing Guest Experience Through Robot Navigation Systems. Journal of Hospitality Technology, 14(2), 45-58.

[6] Wang, X., & Zhang, Y. (2020). Optimization of Robot Navigation Paths Using Machine Learning. Robotics and Automation Research, 5(3), 112-125.

[7] Liu, H., et al. (2021). Improving Efficiency and Reducing Errors in Robot Navigation Through Machine Learning. International Journal of Robotics and Automation, 36(4), 321-335.

[8] Chen, S., et al. (2018). Sensor Data Preprocessing Techniques for Robot Navigation Systems. IEEE Transactions on Robotics, 24(1), 89-102.

[9] Li, J., & Wu, Q. (2019). Enhancing Accuracy and Reliability in Robot Navigation Through Sensor Data Preprocessing. Robotics and Autonomous Systems, 42(2), 176-190.