Indoor Positioning System Using Wifi Fingerprint

Name:
Tanviraj Jayam PES1UG23CS637
Varshini N PES1UG23CS672

Problem Statement:

GPS doesn't work well inside buildings, making it hard to track the location of people or objects indoors. We need an Indoor Positioning System (IPS) that can accurately determine a person's location using WiFi signals already available in most buildings.

Approach/ Implementation

1.kNN Model

- We experimented with different values of k for kNN and found classification error increased as k grew
- Increasing the training sample size dramatically reduced error in kNN, showing the importance of having more data for effective learning and generalization.
- Testing different numbers of PCA components revealed that the first several principal features captured most of the relevant information, while adding more components beyond a point yielded only marginal gains.

2.SVM Model

- We found that SVM error dropped as sample size increased, with further improvements when using more PCA components (e.g., 150 vs. 50).
- We implemented SVM classifiers with both linear and third-order polynomial kernels on PCA-transformed data. And found higher accuracy when using polynomial kernel
- We found that SVM error dropped as sample size increased, with further improvements when using more PCA components

3. Decision Tree Model

- A single, fully grown Decision Tree provided moderate accuracy but was outperformed by ensemble methods. For instance, with 50 PCA components, it achieved a 95.4% accuracy.
- We replicated the paper's pruning methodology by using cross-validation to find the optimal cost-complexity pruning parameter (`ccp_alpha`).
- The experiment revealed that for the 50-component dataset, the best-performing model was the unpruned tree (`ccp_alpha=0.0`), which achieved a final accuracy of **95.4%**, a result superior to the paper's best pruned tree (92.9% accuracy).

4. Gradient Boosting Model

- As an ensemble method designed to improve upon a single tree, Gradient Boosting proved to be a top performer, as suggested by the paper.
- Using parameters from the paper (`n_estimators=189`, `max_depth=4`), the model achieved an excellent accuracy of **97.1%** on the full dataset with 50 PCA components, slightly outperforming the best k-NN and pruned Decision Tree models in that feature space.
- The model's performance was highly dependent on sample size. Accuracy increased dramatically from a volatile **58%** on 100 samples to a much more stable **89.5%** on 1,000 samples, confirming the paper's conclusion that it is robust with sufficient data volume.

Conclusion

The k-Nearest Neighbors (k-NN) model (with k=1) and the Gradient Boosting model proved to be the most accurate and reliable classifiers, both achieving approximately 97-99% accuracy with an appropriate number of PCA components. While a single Decision Tree provided a strong baseline, the Gradient Boosting ensemble significantly improved upon its accuracy, supporting the paper's primary conclusion.

Ultimately, the experiments validate the paper's final recommendation to use a combination of k-NN for its high performance on large datasets and Gradient Boosting for its robustness and superior predictive power.