Prediction of Crime Types in Chicago Using Deep Learning

Wendi Zhao
wendi2@illinois.edu
University of Illinois Urbana-Champaign
Champaign, Illinois, USA

Hongxiao Chen
hc39@illinois.edu
University of Illinois Urbana-Champaign
Champaign, Illinois, USA

Vincent Lyu
vlyu2@illinois.edu
University of Illinois Urbana-Champaign
Champaign, Illinois, USA

Chongbin Yao*
cy5@illinois.edu
University of Illinois Urbana-Champaign
Champaign, Illinois, USA

Abstract

This paper presents a deep learning approach to predict crime patterns in Chicago, utilizing an LSTM-based model trained on historical crime data, weather information, and temporal features. The study addresses challenges in crime type classification through data preprocessing and model optimization, achieving improved prediction accuracy compared to baseline methods. Our findings highlight the significance of location-based features in crime prediction and provide insights for law enforcement resource allocation.

1 Introduction

The rising trend in shooting incidents in Chicago has highlighted the need for predictive policing tools. This research aims to develop a machine learning model capable of predicting crime types based on historical data and environmental factors, with the ultimate goal of optimizing law enforcement resource allocation.

2 Data Collection and Processing

2.1 Data Sources

The study utilized three primary data sources:

- Chicago Crime Data (2001-Present) from the City of Chicago's public safety database
- Weather data from the National Oceanic and Atmospheric Administration (NOAA)
- Holiday information for temporal context

2.2 Preprocessing Methodology

Initial data analysis revealed significant class imbalance across 31 crime types. To address this:

- Selected the top 10 most frequent crime types
- Grouped remaining types into an "Others" category
- Implemented undersampling/oversampling techniques
- Converted datetime information to pandas timestamps
- Encoded non-numerical data for model compatibility

*All authors contributed equally to this project.

3 Model Architecture

3.1 Baseline Model

Initial experiments with logistic regression revealed limitations:

- Achieved only 23% accuracy
- Demonstrated bias toward majority class (theft)
- Failed to account for temporal patterns

3.2 LSTM Implementation

The final model architecture incorporated:

- Input layer processing sequences of 24 previous crimes
- Multiple LSTM layers with dropout for temporal pattern recognition
- Batch normalization for training stability
- Output layer for 11-class prediction

3.3 Optimization

Comparative analysis of optimizers revealed:

- Adam: Best overall performance with consistent loss reduction
- SGD: Fastest training but poor convergence
- RMSprop: Comparable to Adam but slightly lower performance

4 Results and Analysis

4.1 Feature Importance

Permutation importance analysis identified key predictive factors:

1. Location-based features (District, Ward, Community Area)
2. Temporal patterns (hour of day)
3. Environmental factors (minimal impact)

4.2 Model Performance

The LSTM model demonstrated:

- Improved accuracy over baseline
- Better handling of class imbalance
- Effective capture of temporal dependencies

5 Future Work

Future development will focus on:

1. Model enhancement through fine-tuning and feature engineering
2. Dataset expansion to include other urban areas
3. Development of practical deployment interfaces
4. Integration of real-time prediction capabilities

6 Conclusion

This research successfully developed a deep learning approach to crime prediction in Chicago. The LSTM-based model, enhanced by careful data preprocessing and optimization, provides a foundation for practical law enforcement applications. While the current implementation shows promising results, continued refinement and expansion will further enhance its utility across different urban contexts.

Acknowledgments

We thank the University of Illinois at Urbana-Champaign for supporting this research, and the City of Chicago for providing access to their crime database.