ДИФФЕРЕНЦИАЛЬНАЯ ПЕРЕНОСИМОСТЬ ПРОСТРАНСТВЕННЫХ И ВРЕМЕННЫХ КОМПОНЕНТ ПРИ КРОСС-ГОРОДСКОМ ПРОГНОЗИРОВАНИИ ПРЕСТУПНОСТИ

Абдирайимов Х.Б.[1], Болтаев Ж.Э.[1],
Марахимов А.Р.[2], д.т.н., профессор, Худайбергенов К.К.[3], PhD
[1] Термезский государственный университет, г. Термез
[2] Ташкентский государственный технический университет им. И. Каримова, г. Ташкент
[3] Международный университет Кимё, г. Ташкент

Ключевые слова: кросс-городской перенос, доменный сдвиг, диссоциация компонент, семантическая инвариантность, нулевой шот, горячие точки преступности.

Проблема. Практически все существующие работы по глубокому прогнозированию преступности оценивают модели в рамках одного города, оставляя вопрос о переносимости обученных представлений при доменном сдвиге нерешённым. Между тем для городов, не располагающих историческими данными достаточного объёма, именно перенос предобученных моделей представляет наибольший практический интерес. Открытым остаётся также вопрос: какие именно компоненты глубокой пространственно-временной модели переносятся между городами, а какие теряют предсказательную силу при смене культурного, регуляторного и экономического контекста.

Предлагаемое решение. Применён строгий протокол нулевого переноса: модель тезиса 1, обученная на данных Лос-Анджелеса, применяется к данным Ташкента без какого-либо переобучения. Все параметры модели заморожены; пересчитываются лишь граф пространственных связей и конформный квантиль на новом калибровочном окне. Пространственное и временное ядра оцениваются раздельно по двум некоррелированным метрикам — средней абсолютной ошибке для точечной точности и индексу прогностической точности PAI@10 для ранжирования горячих точек, — что позволяет выявить покомпонентный характер деградации при доменном сдвиге.

Данные целевого домена. Реальные данные системы «102» Единой службы экстренного реагирования МВД Республики Узбекистан за период 2020–2025 гг., охватывающие все 12 районов города Ташкента. Корпус содержит 610 868 подтверждённых инцидентов по 14 категориям преступлений и правонарушений, агрегированных на гексагональной сетке H3 (800 активных ячеек). Для обеспечения сопоставимости с источниковым доменом сохраняется единая признаковая схема: 27 категорий POI, демография, метеорологические данные ERA5, календарь с учётом местных праздников.

Результаты. Сопоставление специализации ядер в исходном (Лос-Анджелес) и целевом (Ташкент) доменах приведено в таблице 1.

Метрика	LA: врем.	LA: простр.	ТШК: врем.	ТШК: простр.
MAE ↓	0,321 ✓	0,348	1,580 ✓	1,685
PAI@10 ↑	2,80	2,95 ✓	0,674	0,816 ✓
Веса ансамбля	0,524	0,476	0,70	0,30

Таблица 1. Специализация ядер: внутри домена (LA) против кросс-домена (Ташкент).

В обоих доменах временное ядро доминирует по точечной метрике, а пространственное — по ранговой. Однако масштаб различий перераспределяется при доменном сдвиге: преимущество временного ядра по MAE составляет 7,8 % в LA и 6,2 % в Ташкенте (сохраняется), тогда как преимущество пространственного ядра по PAI@10 возрастает с 5,4 % до 21 % — почти четырёхкратное усиление. Общий масштаб доменного сдвига составляет nMAE = 0,85 → 3,78, увеличение в 4,4 раза. Бутстрап-тест подтверждает статистическую значимость усиления ($p < 0{,}001$).

Содержательная интерпретация. Пространственное ядро опирается на POI-плотности и демографические признаки, обладающие

семантической универсальностью между городами: категория «базар» в Ташкенте и «marketplace» в Лос-Анджелесе имеют типологически сходные криминогенные характеристики (высокий пешеходный поток, наличный оборот, ограниченный формальный контроль); категория «школа» в обоих городах представляет место концентрации несовершеннолетних. Эта семантическая близость обеспечивает корректное ранжирование ячеек по уровню риска даже в новом городе. Временное ядро, напротив, обучено на локальных паттернах источника: ритме ночной жизни, специфических праздниках, метеорологических циклах. В Ташкенте эти паттерны качественно иные: другая структура праздников (Навруз, Рамазан), другая алкогольная регуляция, другой режим торговли и досуга, — в результате чего выученные временные представления теряют семантическую нагрузку при переносе.

Практические рекомендации. (1) Для задач распределения патрульных ресурсов, где ключевым критерием является ранговое качество выявления горячих точек, при кросс-городском применении следует отдавать предпочтение пространственному ядру без переобучения. (2) Временное ядро требует адаптации на локальных данных (дообучение календарных признаков, эмбеддингов ячеек, метеорологической компоненты); единой ансамблевой головы с глобальными весами недостаточно. (3) Конформная калибровка робастна к доменному сдвигу: эмпирическое покрытие 0,90 при номинальном 0,90 сохраняется без модификации калибровочной процедуры.

Выводы. (1) Впервые на реальных данных двух географически и культурно различных городов документирована устойчивая диссоциация пространственного и временного ядер ансамблевой модели прогнозирования преступности в режиме нулевого переноса. (2) Установлено эмпирическое правило: POI-укоренённые пространственные представления переносятся между городами, тогда как выученные временные представления локальны и требуют адаптации. (3) На реальных данных системы «102» Ташкента (610 868 подтверждённых инцидентов, 12 районов, 2020–2025) подтверждена практическая полезность нулевого переноса для задачи выявления горячих точек. (4) Результаты определяют принципы построения мультигородских систем оперативного прогнозирования: замораживание пространственных компонент и адаптация временных.

Литература

1. Тезис 1 настоящей работы: Ансамблевая пространственно-временная модель вероятностного прогнозирования городской преступности с конформной калибровкой.
2. Huang Y. et al. CrimeForecaster // ECML-PKDD. 2022.
3. Gasteiger J. et al. Predict then propagate // Proc. ICLR. 2019.
4. Xu C., Xie Y. Conformal prediction for dynamic time-series // Proc. ICML. 2021.