

Методы интеграции экологических данных при использовании географических эмбедингов в вопрос-ответной информационной системе

Гавенко Ольга Юрьевна

Д.т.н., к.филол.н., г.н.с. ИДСТУ СО РАН, в.н.с. ФИЦ ИВТ

Шашок Наталья Александровна

М.н.с. ИДСТУ СО РАН, аспирант ФИЦ ИВТ

В сфере экологических исследований обработка информации требует интеграции разнородных (гетерогенных) данных: **географических координат, экологических параметров и семантических признаков.**

Географические эмбединги, или векторные представления пространственных данных, являются эффективным инструментом для решения этой задачи, в частности, при построении интеллектуальной вопрос-ответной системы, обрабатывающей экологическую информацию.

Географические эмбединги применяются в задачах анализа пространственных зависимостей, таких как *прогнозирование транспортных потоков, кластеризация локаций* для рекомендательных систем или *семантическая сегментация городских зон*.

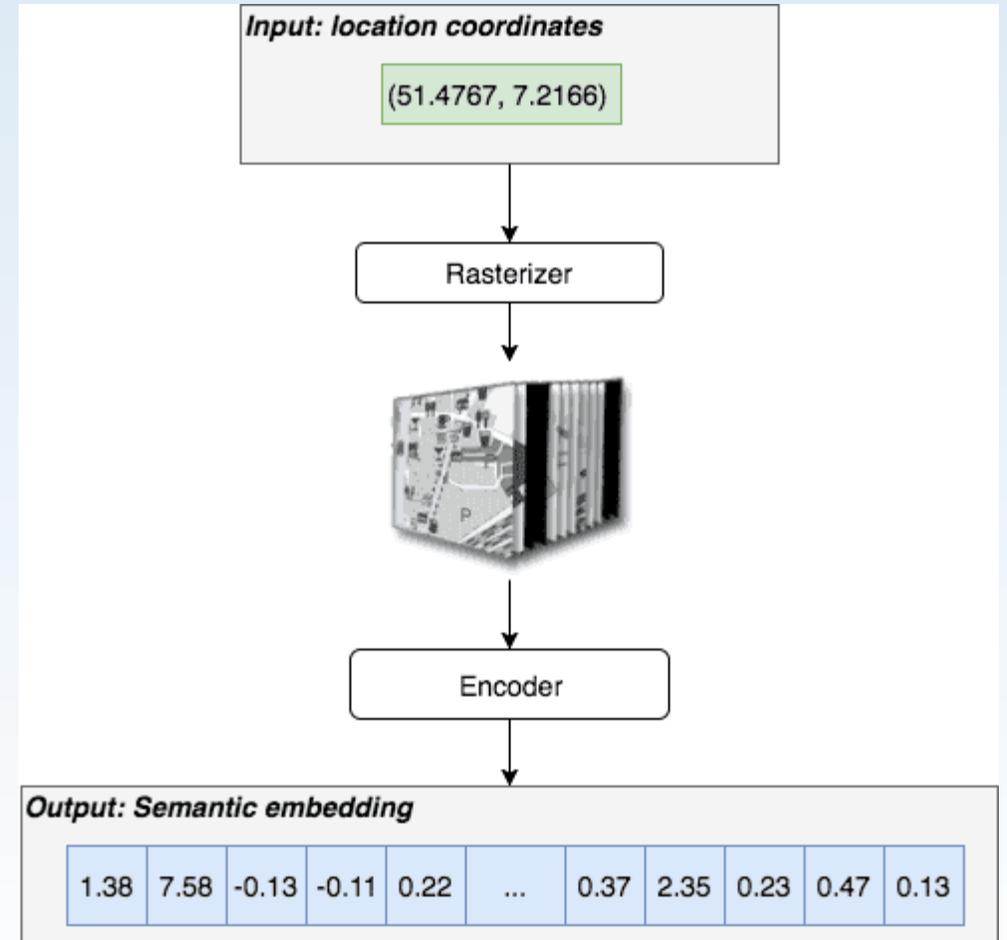
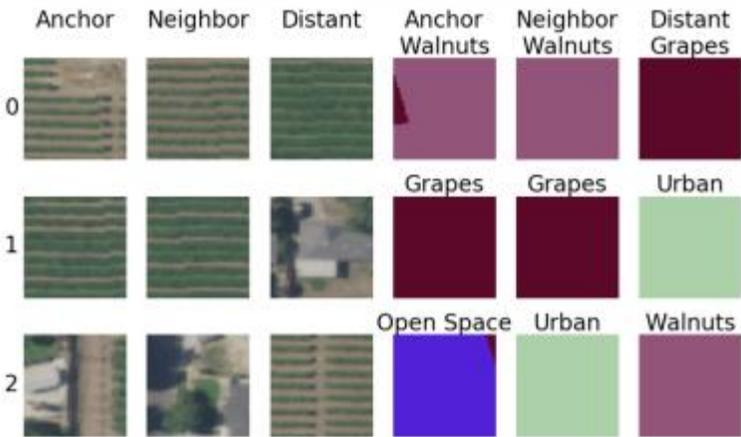
Их использование в вопрос-ответных системах позволяет обогащать ответы географическим контекстом, что критично для запросов, связанных с локализацией объектов, маршрутизацией или анализом пространственных закономерностей.

Методы генерации географических эмбеддингов зависят от целевых задач и ограничений данных.

Такие алгоритмы, такие как **Space2Vec**, **Place2Vec**, **Tile2Vec** и **Loc2Vec**, специализируются на трансформации пространственной информации в векторное пространство.

Метод **Space2Vec** использует модель энкодер-декодер на основе синусоидальных функций с различными частотами для моделирования абсолютных положений и пространственных контекстов, а также механизм многоголового внимания, основанный на контекстных точках.

Place2Vec преобразует координаты точек интереса (POI) в пространственно совмещенные пары POI, определяя их множества, основанные на расстояниях между точками, и не сохраняет информацию о направлении по сторонам света между точками.



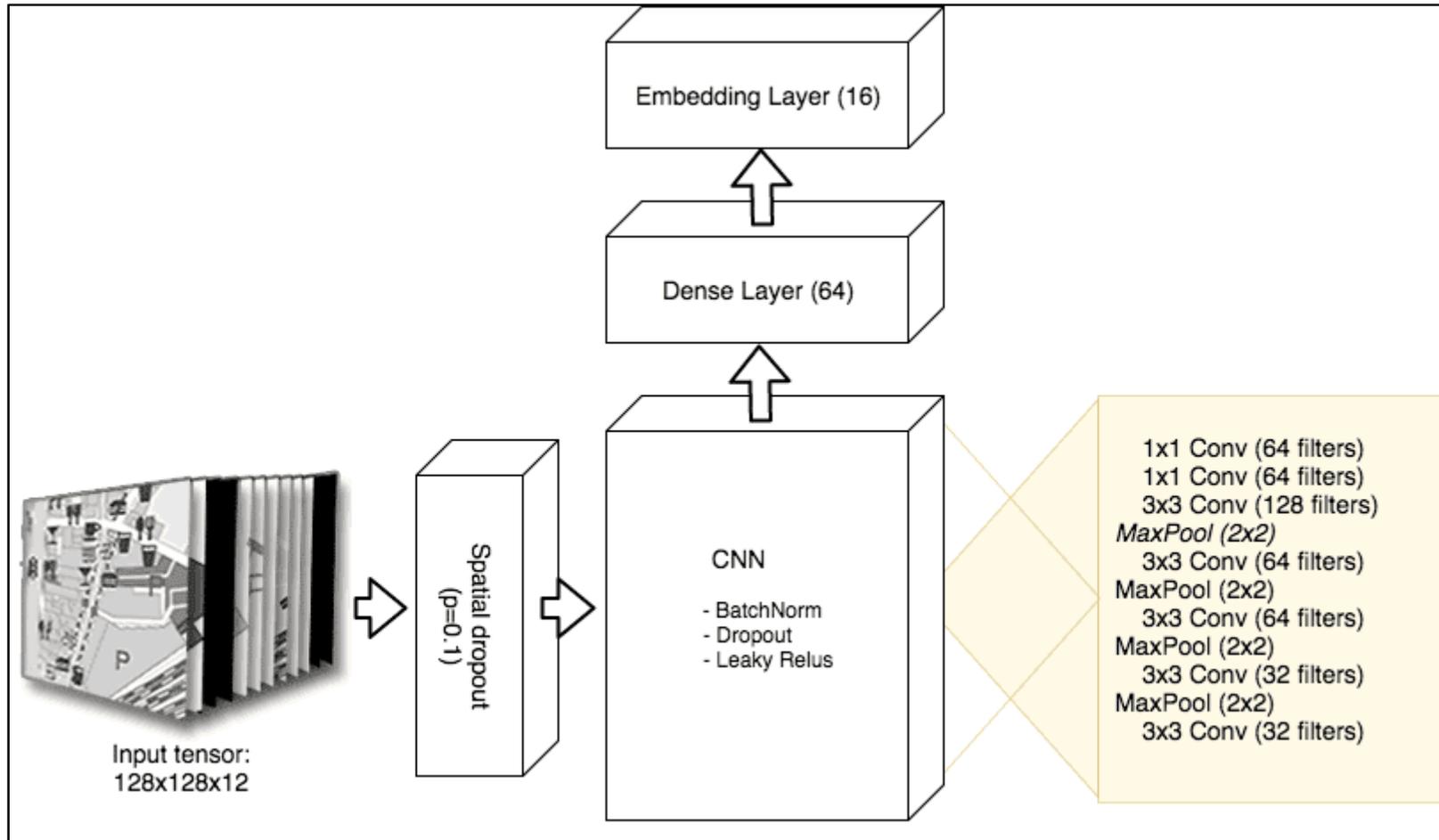
В контексте экологических исследований ключевая задача заключается в эффективном хранении и интеграции гетерогенных данных для создания эмбедингов, указывающих на множество данных, связанных с одной территориальной единицей.

Решению этой задачи могут способствовать подходы, аналогичные **Loc2Vec**, где мультисканальный тензор, представляющий фрагмент карты, кодирует совокупность информационных слоев. К базовым пространственным данным, описываемым авторами этого подхода, относятся дорожные сети, границы водных объектов и береговые линии.

Экологические параметры охватывают более широкий спектр информации: данные рельефа, атмосферы, биосферы, гидросферы и антропогенного воздействия, а также иные данные.

Соответственно, возникает необходимость использования спутниковых снимков, гидрологических замеров или климатических моделей в качестве исходных данных. Это позволяет создавать эмбединги, с помощью которых можно осуществлять поиск данных для решения широкого круга экологических задач.

Для того, чтобы данные, используемые для генерации эмбедингов, отображали полноту экологических характеристик конкретной территории, предполагается возможным дополнительное использование таких практически неизменяемых характеристик рельефа, как высоты над уровнем моря и типы почв, поскольку эти данные определяют температурные градиенты, распределение осадков и стока, а также могут быть использованы для оценки эрозийных рисков и устойчивости склонов.



Таким образом, использование географических эмбедингов представляет собой эффективный подход для хранения, интеграции и анализа гетерогенных экологических данных в рамках вопрос-ответных систем при учете специфики поставленных задач принятия решений и предоставления ответов на основе имеющихся данных. Методы их генерации, основанные на интеграции информации из множества каналов, позволяют значительно повысить релевантность и контекстную осведомленность ответов системы.

Перспективы развития лежат в области решения задач интеграции динамических данных, совершенствования алгоритмов для специализированных экологических сценариев и исследования возможностей искусственного интеллекта для комплексного пространственно-экологического прогнозирования.

Спасибо за внимание!

Контакты:

1. olga.yu.gavenko@mail.ru
2. n.shashok@alumni.nsu.ru