

Рецензия на статью Кривонос А.В. Методы верификации для кластеризации временных рядов

Рецензент: Кожанов И. Р

- **Краткое содержание**

Настоящая статья исследует значимость группировки (кластеризации) временных рядов темпов роста COVID-19 по разным странам и регионам России. При оценке сходства временных рядов используется корреляционный анализ Пирсона с подбором оптимального временного лага для достижения максимума корреляции. Авторы применяют статистический тест на основе перестановок для проверки гипотезы о том, что наблюдаемые совпадения между временными рядами могут быть случайными. Эти перестановки выполняют на матрице мер сходства, случайно меняя местами её элементы и сохраняя симметрию.

На примере эпидемиологических данных COVID-19 авторы показывают, как можно классифицировать страны по четырем крупным кластерам на основе сходства эпидемических кривых роста заболеваемости. Они также вычисляют коэффициент "силуэта", который позволяет оценить качество разделения стран на кластеры.

Анализ показывает, что кластеры сформировались не случайно, и имеют определённые отличительные особенности. Выводы и методы авторов могут использоваться для дальнейшего исследования связей между характеристиками регионов и течением эпидемических процессов.

- **Достоинства**

1. Актуальность темы: Исследование актуально из-за оценки статистической значимости кластеризации временных рядов по теме эпидемиологии, в частности Covid-19, что имеет важное значение для понимания различий в распространении болезни по странам и регионам.

2. Оригинальность подхода: Работает над новым методом оценки значимости кластеризаций, сравнивая с использованием перестановочного теста и анализа матриц сходства.

3. Практическое применение: Проведенная кластеризация стран по эпидемическим кривым может служить основой для анализа и прогнозирования тенденций эпидемий.

4. Наглядность результатов: Использование графических материалов помогает лучше визуализировать результаты кластеризации.

5. Описаны этапы работы: Четко изложены шаги выполнения эксперимента и методы работы с данными.

6. Ссылки на предыдущие исследования: Статья содержит ссылки на другие работы, что позволяет сопоставить используемые методы с текущими научными подходами.

- **Недостатки**

1. Короткий вывод: Вывод статьи представлен кратко и не полностью отражает значимость результатов исследования. Желательно было бы более подробно обсудить влияние результатов на понимание эпидемических процессов и их применение.
2. Ограниченная библиография: В статье отсутствует подробная библиография, включая полные ссылки на цитируемые работы. Для научной статьи крайне важно иметь развернутый список литературы.
3. Неполное объяснение методологии: Хотя в тексте описан альтернативный подход к кластеризации, отсутствует подробное описание используемых методов и статистических инструментов.
4. Не рассматриваются ограничения исследования: В статье не указаны возможные ограничения методов кластеризации, которые могут повлиять на интерпретацию результатов.