

## **Индекс УДК 004.09** Прикладные информационные (компьютерные) технологии

Технологии повышения производительности расчетов в задачах классификации на основе электронных медицинских данных

Ефимов Е.О.

(СПб НИУ ИТМО, Санкт-Петербург)

Научный руководитель – Ковальчук С.В.

(СПб НИУ ИТМО, Санкт-Петербург)

**Введение.** Применение электронных медицинских карт привело к изменениям в исследованиях в области здравоохранения, поскольку они содержат текстовые, графические и числовые данные, в которых содержатся результаты лабораторных исследований, заключения визуализационных и инструментальных методов диагностики, показатели комплексного клинического осмотра, представленные в виде историй болезни, анамнезов, эпикризов. Несмотря на это, данные представлены в слабоструктурированном виде. Актуальной задачей остаются поиск клинически значимой информации в диагнозе заболевания, оценка рисков пациента, исследования воздействия препаратов, классификация и категоризация. Опираясь на полученную информацию, возможно спрогнозировать исходы заболеваний пациентов, а также выбрать наиболее подходящую тактику лечения. Интеллектуальный анализ текста (text mining) позволяет проводить анализ ЭМК и извлекать ценную информацию, что несомненно имеет высокую ценность для лечения пациентов и системы здравоохранения в целом. Проблема заключается в том, что векторы, полученные в результате факторизации разнородных медицинских записей имеют значительную размерность, так как процесс лечения, отраженный в электронных медицинских картах комплексен: он состоит из многих этапов с привлечением значительного разнообразия специалистов и персонала. Вследствие этого данные, описывающие лечебный процесс, также комплексны. Более того, имеется тенденция значительного роста накапливаемых медицинских данных.

**Цель работы:** разработать методы анализа разнородных медицинских записей для производительности расчетов при применении методов интеллектуального анализа текстов и данных.

**Базовые положения исследования:** расчеты в процессе анализа данных ЭМК обычными средствами требуют значительное количество времени. Интерактивность анализа данных требует адекватного отклика в получении результатов на запросы пользователей. Значимой становится задача повышения производительности расчетов в условиях специфики предметной области, решаемых задач, а также специфики данных, описывающих клинические процессы.

**Промежуточные результаты:** модифицирован метод повышения производительности расчетов текстовых векторов значительной размерности при решении задач классификации, разработанный метод реализован при использовании технологий вычисления на GPU и CPU, проведен вычислительный эксперимент по апробации и сравнению разработанного методов на производительность на ретроспективных данных МИС о лечении пациентов с ишемической болезнью сердца.