

**УДК 004.8**

**АЛГОРИТМ ВЫБОРА ПРИЗНАКОВ ДЛЯ ВЫСОКОРАЗМЕРНЫХ ВРЕМЕННЫХ РЯДОВ С ПРИМЕНЕНИЕМ СКРЫТЫХ МАРКОВСКИХ МОДЕЛЕЙ**

**Довжик Л.И.** (Университет ИТМО, Факультет Информационных Технологий и Программирования),

**Научный руководитель – к.т.н., доцент ФИТиП Сметанников И.Б.**  
(Университет ИТМО, Факультет Информационных Технологий и Программирования)

В данном докладе рассматривается алгоритм выбора признаков для высокоразмерных временных рядов, основанный на скрытых марковских моделях.

Среди проблем машинного обучения и анализа данных проклятие размерности одна из ключевых, поскольку большинство наборов данных имеют большое число различных признаков, что затрудняет работу моделей машинного обучения и с точки зрения скорости обучения и качества работы этих моделей. В связи с чем алгоритмы понижения размерности очень важны и часто применяются при решении задач машинного обучения. Существует несколько видов алгоритмов понижения размерности: фильтры, обертки, ансамбли, встраиваемые и гибридные методы. Каждый из видов имеет свои преимущества и недостатки, в связи с чем выбор алгоритма понижения размерности зависит от постановки задачи и от самих данных. Например, основной проблемой в решении генетических задач является то, что упрощенные модели, работают с конкретной точкой во времени, предполагая отсутствие изменений в признаках или генах, что может отличаться от реальности из-за возможности их изменения с течением времени и условий проведения испытаний. На данный момент нет библиотек позволяющих разнообразно решать задачу понижения размерности.

Для решения этой задачи разрабатывается библиотека ИТМО\_FS, в которой будут собраны различные алгоритмы и методы выбора признаков, что облегчит будущие исследования в рамках машинного обучения. Один из алгоритмов, достаточно эффективно решающих данную проблему для временных рядов, базируется на параметрической скрытой марковской модели из двух состояний – изменение или отсутствие изменений. Для решения задачи выбора признаков с применением скрытых марковских цепей необходимо научиться находить наиболее подходящий набор скрытых состояний, который может объяснить набор наблюдений. Для этого можно использовать алгоритм Витерби: в качестве наблюдений используются последовательности значений признаков во времени для исследуемых представителей.

На основе статьи реализовывается алгоритм выбора признаков, расширяющий функционал библиотеки ИТМО\_FS, в которой в данный момент слабо представлены алгоритмы, работающие с временными рядами. Ибо качество подобных алгоритмов может быть значительно выше по сравнению с алгоритмами, оперирующими данными в конкретный момент времени.

Довжик Л.И. (автор)

Подпись

Сметанников И.Б. (научный руководитель)

Подпись