

## Итеративный отбор существенных признаков в условиях их мультиколлинеарности при прогнозировании временных рядов

Магнитосфера Земли представляет собой сложную динамическую систему, состояние и поведение которой зависит не только от её текущего состояния и от внешнего воздействия на неё, но и от предыстории. Для прогнозирования её состояния могут использоваться методы машинного обучения. При этом для учёта предыстории следует использовать либо рекуррентные алгоритмы, либо погружение (топологическое вложение) многомерного временного ряда входных переменных, при котором в каждый пример данных включаются, помимо текущих значений входных переменных, также их предыдущие значения с задержками от одного шага временного ряда до величины, называемой глубиной погружения.

Недостатком погружения временного ряда является кратное увеличение количества входных признаков, что повышает требования к количеству примеров в обучающей выборке и может создавать проблемы, связанные с переучиванием моделей. Между тем, получаемые при погружении временного ряда входные признаки характеризуются мультиколлинеарностью и несут во многом сходную информацию. Поэтому для понижения размерности входных данных при прогнозировании правильно использовать алгоритмы отбора входных признаков, которые учитывают мультиколлинеарность.

В данной работе рассматривается задача прогнозирования геомагнитного индекса Dst, а в качестве входных данных используются данные о параметрах солнечного ветра и межпланетного магнитного поля в точке Лагранжа L1 между Солнцем и Землёй, а также предыстория самого прогнозируемого индекса.

Отбор существенных входных признаков осуществляется с помощью метода, основанного на итеративном отборе признаков с наибольшей корреляцией по отношению к целевой переменной и исключении признаков с высокой взаимной корреляцией. Этот метод сравнивается с традиционным методом отбора - кросс-корреляционным фильтром, а также с ограничением глубины погружения каждого входного параметра по величине автокорреляции без использования отбора. Критерием сравнения является среднеквадратичная ошибка построенного на отобранном наборе признаков прогноза Dst-индекса на независимых данных.

Исследование выполнено за счёт гранта Российского научного фонда № 23-21-00237, <https://rscf.ru/project/23-21-00237/>.

### Секция

Космическая погода

**Primary authors:** ЩУРОВ, Николай (Физический факультет МГУ им. М. В. Ломоносова); Mr ИСАЕВ, Игорь (НИИЯФ МГУ); Dr МЯГКОВА, Ирина (НИИЯФ МГУ); Dr БАРИНОВ, Олег (НИИЯФ МГУ); ДОЛЕНКО, Сергей (НИИЯФ МГУ)

**Presenter:** ЩУРОВ, Николай (Физический факультет МГУ им. М. В. Ломоносова)