

**УДК:** 004.021

**Название:** Идентификация устойчивой структуры модели непрерывного процесса на основе гидрометеорологических данных.

**Авторы:**

Масляев М.А., Университет ИТМО, г. Санкт-Петербург;

Калюжная А.В., Университет ИТМО, г. Санкт-Петербург;

**Научный руководитель:** Калюжная А.В., Университет ИТМО, г. Санкт-Петербург;

**Контакты:** [miklemas@list.ru](mailto:miklemas@list.ru), +79151459725;

**Тезис доклада:**

Управляемые данными алгоритмы являются распространенными механизмами построения моделей явлений, недостаток знаний о которых не позволяет строить их аналитические описания. Частными случаями такого подхода являются управляемые данными алгоритмы вывода уравнений в частных производных, которые могут выявлять структуру дифференциального уравнения, описывающего динамическую систему на основе полей физических величин. Управляемые данными методы могут применяться во многих сферах науки, где по тем или иным причинам невозможно аналитически вывести описывающее систему уравнение, но можно получить ряды наблюдений в исследуемой области. Подобные алгоритмы в большинстве случаев основываются на применении методов машинного обучения и, в частности, на использовании разреженной регрессии.

Целью данной работы является создание управляемой данными системы вывода дифференциальных уравнений. Для детального анализа способности алгоритма описывать динамическую систему, валидация может быть проведена на результатах решения уже исследованного уравнения в частных производных. Использование искусственных входных данных для валидации алгоритма дает большую гибкость по сравнению с данными полноценных измерений, которые будут использоваться при дальнейшей работе метода.

Разработанный в данной работе метод основывается на сочетании L1 – регуляризованной линейной регрессии и генетического алгоритма для воссоздания структуры уравнения в частных производных, которая наиболее соответствует входным данным. В качестве входных данных алгоритм использует нормированную по времени матрицу решения уравнения, на основе которых вычисляются производные по времени и пространству. Для получения производных в случае данных с низким шумом рекомендуется использовать метод конечных разностей, в то время как при большой дисперсии шума необходимо применять более точные алгоритмы, например, дифференцирование построенных на решении интерполяционных полиномов.

Использование L1 – регуляризации при линейной регрессии на нормированных данных позволяет отбирать идентифицировать слагаемые, которые будут иметь ненулевые веса на итоговом уравнении и определять знак соответствующих коэффициентов. В большинстве существующих методов предполагается использовать регрессию на основе полного набора потенциальных слагаемых, однако это существенно ограничивает класс получаемых уравнений. Эта проблема решается приложением генетического алгоритма, подбирающего наиболее подходящую структуру через построение различных наборов случайно полученных слагаемых. После нахождения структуры уравнения для получения истинных значений коэффициентов используется линейная регрессия на ненормированных данных.

Созданный алгоритм успешно идентифицирует структуру волнового уравнения и уравнений Бюргерса и Кортевега-де Вриза на данных без шума. Зашумленные данные, существенно ограничивают качество результата на решениях уравнения волны и Кортевега-де Вриза в силу их свойств; структура уравнения Бюргерса успешно определяется даже при высоких значениях дисперсии шума.

Автор	_____	/	_____	/
Научный руководитель	_____	/	<u>Бухановский А.В.</u>	/
Директор МФ ТИИТ	_____	/	<u>Бухановский А.В.</u>	/

