

ПРИМЕНЕНИЕ НЕЧЕТКОГО МОДЕЛИРОВАНИЯ В ЗАДАЧЕ ИМПУТИРОВАНИЯ ДАННЫХ

Лучкова С.В., Перемитина Т.О., Яценко И.Г.

Федеральное государственное бюджетное учреждение науки Институт химии нефти Сибирского отделения Российской академии наук (ИХН СО РАН), г. Томск

Лучкова С.В.
Перемитина Т.О.
Яценко И.Г.
Федеральное
государственное
бюджетное учреждение
науки Институт химии
нефти Сибирского
отделения Российской
академии наук (ИХН СО
РАН)

В настоящее время технология нечеткого моделирования является одной из развивающихся областей обработки данных. Применение нечетких систем в задаче импутирования вызывает ряд вопросов касательно параметров системы: как правильно выбрать параметры построения базы правил, функции принадлежности, метода эволюционной стратегии, оптимизирующей параметры нечеткой системы, и другие. Данные вопросы очень важны, и являются объектами рассмотрения в данной работе.

Многие исследования связаны со сбором и обработкой данных, представленных в виде таблиц наблюдений. Данные из этих таблиц участвуют как в различных задачах анализа, так и в задачах построения моделей прогноза. Но часто, по различным причинам, эти таблицы содержат пропущенные значения.

Большинство методов анализа данных не могут обрабатывать таблицы с пропущенными значениями, так как получаются не адекватные модели, либо модель вообще построить не возможно. Иногда, получается, откинуть свойства с пропущенными значениями или попытаться подставить среднее значение данного показателя по свойству в таблице. Но подобные действия приводят к потере информации и значительно искажают результаты, а данные все равно нуждаются в обработке. Поэтому «восстановление» или «процедура заполнения пропусков» является очень важным моментом в обработке данных.

Самыми известными методами восстановления являются различные статистические модели, основанные на методе безусловных средних, методе ближайшего соседа, методе главных компонент, регрессии, Z-методе, методе, основанном на производных, и т.д. Методы различными способами восстанавливают значения по имеющимся данным, но, что очень важно, они требуют от данных определенных условий, например, нормальность распределения, что зачастую не выполняется, так как данные природного характера не имеют однородной структуры из-за своих свойств, а процедуры нормировки разрушает связи в данных. Поэтому требуется такой метод, который даст максимально точное восстановленное значение пропуска, и при этом не будет требователен к структуре таблицы наблюдений, либо будет наименее требовательным.

В данной статье, для решения задачи восстановления данных, предлагается применить метод, основанный на нечеткой системе, использующий для идентификации своих параметров эволюционные стратегии.

Целью работы является описание задачи восстановления данных в интерпретации нечеткой системы на основе эволюционной стратегии, и сравнение результатов работы алгоритма с аналогами.

В работе предложены алгоритм идентификации нечеткой системы на основе метода эволюционной стратегии для настройки параметров и алгоритм импутирования. На примере использования массива о свойствах вязких парафинистых нефтей были поставлены вычислительные эксперименты с искусственно «пропущенными» данными. Эксперименты показали, что для различных данных оптимальные параметры системы будут разными, и в каждом случае их необходимо подбирать. Поэтому было показано, как следует выбирать параметры нечеткой системы и эволюционной стратегии. Дальнейшие исследования направлены на более детальный подход к обрабатываемым данным, например, возможность задавать количество термов для каждой переменной, что, возможно, сократит количество правил и снизит вычислительное время, а, следовательно, позволит обрабатывать одновременно выборки большего объема данных.

Литература

1. *Ходашинский И.А., Гнездилова В.Ю., Дудин П.А., Лавыгина А.В.* Основанные на производных и метаэвристические методы идентификации параметров нечетких моделей // Труды VIII международной конференции

"Идентификация систем и задачи управления" SICPRO '08. Москва, 2009 г. С.501-528.

2. *S. Hoche, and S. Wrobel.* A Comparative Evaluation of Feature Set Evolution Strategies for Multirelational Boosting // Proc. 13th Int. Conf. on ILP, 2003 г.
3. *Рутковская Д., Пилиньский М., Рутковский Л.* Нейронные сети, генетические алгоритмы и нечеткие системы. М.: Горячая линия – Телеком, 2006.
4. *H.-P. Schwefel.* Numerical Optimization of Computer Models. John Wiley & Sons, 1981
5. *H.-P. Schwefel.* Evolution and Optimum Seeking. New York: John Wiley & Sons, 1995
6. *Дудин П.А.* Применение алгоритма муравьиной колонии для идентификации нечетких моделей // Материалы XLV Международной научной конференции «Студент и научно-технический прогресс» Информационные технологии. НГУ. Новосибирск, 2007. С. 188-189.
7. *Лучкова С.О.* Алгоритмы нечетких систем в задачах импутирования // Материалы III Всероссийской молодежной научной конференции «Современные проблемы математики и механики», г. Томск 2012. С.329-334.
8. *Козин Е.С., Полищук Ю.М., Яценко И.Г.* База данных по физико-химическим свойствам нефтей. Нефть. Газ. Новации. 2011. № 3.С. 13-16
9. *Яценко И.Г., Сваровская Л.И., Перемитина Т.О., Лучкова С.В.* Методы статистического анализа и нечетких систем в исследованиях влияния химического состава и условий залегания нефтей на численность и активность пластовой микрофлоры в задачах повышения нефтеотдачи // Химия нефти и газа: Материалы VIII Международной конференции, 24-28 сентября 2012 г., г. Томск. – Томск: Изд-во Томского государственного университета, 2012.С. 299-302