

## ОТЗЫВ

официального оппонента

на диссертационную работу Натальи Владимировны Сапкиной «Восстановление закономерностей на основе нечетких регрессионных моделей», представленную на соискание ученой степени кандидата технических наук по специальности 05.13.17 – «Теоретические основы информатики»

### 1. Актуальность темы диссертационной работы

Диссертационная работа Сапкиной Н.В. посвящена теме восстановления закономерностей на основе нечетких линейных регрессионных моделей. В настоящее время все более востребованным становится создание систем интеллектуального анализа данных, позволяющих выявить полезную скрытую информацию и восстановить закономерности, которые содержатся в исходных данных. При этом получение достоверных результатов и выводов на основе исследования осуществляется за счет применения различных подходов, основными из которых являются: классификация, поиск ассоциативных правил, кластеризация и др. Определение зависимости некоторого показателя от различных признаков и отображение их взаимосвязи в форме модели осуществляется в ходе проведения регрессионного анализа.

Актуальность темы обусловлена тем, что в современных вычислительных алгоритмах анализ данных характеризуется наличием слабоструктурированной, неполной, неточной и нечеткой информации, возникающей вследствие неопределенности, присущей моделям сложных процессов, а также необходимостью решения задач в тех областях, где существенная роль принадлежит суждениям и знаниям экспертов. Эта специфика значительно усложняет процесс построения моделей и обуславливает необходимость разработки специальных методов повышения их адекватности. С тех пор, как в работах L. Zadeh для формализации неопределенности стала использоваться теория нечетких множеств, интерес к анализу приближенной и, в частности, нечеткой информации значительно вырос.

Одним из возможных способов определения зависимостей между признаками в нечеткой среде является нечеткий регрессионный анализ, который позволяет решать различные задачи в ситуациях, когда традиционные методы неэффективны или совсем неприменимы из-за отсутствия достаточно точных сведений об исследуемых объектах. Построение нечетких регрессионных моделей опирается на математический аппарат, включающий определение арифметических операций над нечеткими числами и их

сравнение. Построение нечетких регрессионных моделей требует, чтобы результаты операций с нечеткими числами представляли также нечеткие числа.

Усовершенствование существующих методов нечеткого регрессионного моделирования за счет применения нечетких чисел  $L$ - $R$ -типа и использования моделей, в которых учитываются различные типы данных в выборке, обуславливает актуальность диссертационного исследования.

Диссертационная работа выполнена в соответствии с одним из основных научных направлений Воронежского государственного университета «Математическое моделирование, программное и информационное обеспечение, методы вычислительной и прикладной математики и их применение к фундаментальным исследованиям в естественных науках».

## **2. Достоверность и новизна основных выводов и результатов диссертации**

Обоснованность научных результатов подтверждается применением современного аппарата и современного программного обеспечения. Диссертационное исследование проведено с использованием хорошо проработанного и верифицированного математического аппарата. Весь математический аппарат, в том числе и обладающий научной новизной, использован для решения поставленных задач вполне корректно.

В диссертационной работе Сапкиной Н.В. представлен комплексный подход к реализации отдельных этапов нечеткого линейного регрессионного моделирования, рассмотрены различные постановки задач, учитывающих нечеткость исходных данных и/или параметров модели. Исследование отличается логичностью и последовательностью изложения. Научной новизной на наш взгляд обладают следующие основные результаты диссертационного исследования:

- 1) Исследованы свойства арифметических операций над нечеткими числами  $L$ - $R$ -типа: коммутативность, ассоциативность, дистрибутивность, наличие нейтрального и обратного элементов. Установлено существование алгебр с операцией сложения или умножения. Для группоидов нечетких чисел с одной бинарной операцией найдены нейтральные элементы и определено, что каждый из них образует моноид. В работе также доказано, что группоид с умножением образует абелеву группу, а группоид с операцией сложения не имеет обратного элемента, ввиду чего алгебраические структуры с двумя операциями не могут быть выявлены. Кроме того, в ходе исследования получены формулы для нахождения инволюции, произведения нечеткого

числа на нечеткий ноль и единицу, формула квадрата нечеткого числа, а также произведения суммы двух нечетких чисел и произведения квадрата нечеткого числа на обратное ему число, которые позволяют осуществлять вычисления при проведении нечеткого регрессионного моделирования. Все перечисленные результаты являются новыми.

2) Проведена оценка параметров нечеткой парной и множественной линейной регрессионной модели с помощью метода наименьших квадратов. При этом были рассмотрены различные постановки задач, учитывающих нечеткость исходных данных и параметров модели. Отличительной особенностью подхода является возможность нахождения функционального взаимодействия между зависимыми и независимыми переменными в нечеткой среде, что позволяет решать различные задачи в ситуациях, когда традиционные методы неэффективны или совсем неприменимы из-за отсутствия достаточно точных сведений об исследуемых объектах или ввиду недостаточности объема выборочной совокупности.

3) Осуществлена оценка адекватности построенных регрессионных моделей и исследована последовательность остатков (ошибок) вычисления, а также определена точность моделей путем вычисления средней относительной ошибки. При этом предложены различные процедуры дефазификации нечеткой остаточной последовательности и нечеткого коэффициента корреляции, что дает возможность применять классические методы для их анализа.

4) Предложены альтернативные подходы к выявлению множества существенных переменных в рамках нечеткого регрессионного моделирования, основанные на нечетком коэффициенте корреляции, стандартизированном уравнении нечеткой множественной линейной регрессии и применении автоассоциативных нейронных сетей, «работающих» с приближенной информацией. Все это в совокупности позволяет обеспечить комплексность анализа данных на различных этапах процесса выявления закономерностей в данных.

5) Разработана информационная система интеллектуального анализа данных с помощью нечеткого регрессионного моделирования, основанная на технологии информационного хранилища и способная обрабатывать большие массивы данных, осуществлять автоматизированный поиск ранее неизвестных закономерностей и скрытых и неочевидных правил в данных.

Перечисленные научные результаты получены Н.В. Сапкиной самостоятельно, опубликованы в печати, докладывались и обсуждались на различных международных и российских конференциях.

При проведении исследований корректно были использованы методы математической статистики, в частности, линейные модели регрессионного анализа, алгебры, нечеткой арифметики, теории информационных систем. Теоретические положения и выводы диссертации строго обоснованы и подтверждены результатами вычислительных экспериментов. Диссертация содержит значительное количество иллюстративных примеров, поясняющих особенности нечеткого регрессионного анализа. Достоверность научных результатов диссертации также подтверждена созданием информационной системы интеллектуального анализа данных, работоспособность которой проверена в рамках вычислительного эксперимента, что позволило соискателю сравнить подходы к восстановлению закономерностей в данных на основе классического и нечеткого линейного регрессионного моделирования. Полученные выводы экспериментального исследования полностью согласуются с теоретическими положениями.

Таким образом, научные результаты, выводы и основные положения диссертационной работы Н.В. Сапкиной обоснованы и достоверны.

### **3. Ценность диссертационной работы для науки и практики**

Рассмотрение теоретических основ диссертации приводит к выводу о высоком научном уровне проведенных доказательств основных положений диссертации. Важными теоретическими результатами, оказывающими существенное влияние на дальнейшее развитие математического аппарата теории нечетких чисел, следует признать все заявленные в диссертации основные теоретические разработки, которые позволяют использовать и развивать сформированные в диссертации подходы, связанные с нечеткими регрессионными моделями.

Ценность результатов диссертационной работы Сапкиной Н.В. для науки заключается в следующем:

- получена совокупность теоретических результатов, касающихся свойств арифметических операций над нечеткими числами  $L$ - $R$ -типа и существования алгебраических структур, что позволяет осуществлять вычисления при построении нечетких регрессионных моделей;

- предложены методы восстановления закономерностей на основе нечетких линейных регрессионных моделей, в которых рассмотрены различные постановки задач, учитывающих нечеткость исходных данных и параметров модели. Использование нечетких концепций в задачах регрессионного анализа может служить альтернативой классическому

статистическому подходу. В рамках этого подхода в значительной мере снимается проблема оценки распределений эндогенных переменных.

Ценность результатов диссертационной работы для практики обусловлена созданием программного обеспечения информационной системы интеллектуального анализа данных, предназначенного для автоматизированного поиска закономерностей, которые позволяют выявить скрытые правила в базе данных. Информационная система может быть использована в качестве подсистемы в экспертных системах и системах поддержки принятия решений на этапе предварительного анализа больших объемов информации; выявления взаимосвязи между различными показателями, информация о которых содержится в базе данных (сокращение размерности, установление приоритетов); формирования знаний в виде правил о предметной области.

Научные результаты диссертационной работы Н.В. Сапкиной используются в учебном процесс Воронежского государственного университета и, кроме того, внедрены в практическую деятельность ЗАО ЛЦ «АВС Фарбен» (г. Воронеж), что подтверждено соответствующими актами.

В дальнейшем, результаты, полученные в диссертации, целесообразно использовать в вузах и организациях, занимающихся нечетким моделированием и вопросами разработки новых методов и алгоритмов восстановления закономерностей в данных.

#### 4. Замечания по диссертации

К замечаниям, в целом хорошо написанной работы, всё же необходимо отнести следующие:

1. При отборе регрессоров в классическом уравнении регрессии (с.20) оценивается не столько величина коэффициента корреляции, сколько его значимость.

2. Для оценки качества модели нечеткой регрессии (с.88), на наш взгляд, более корректным с методической точки зрения было бы, использование коэффициента детерминации  $R^2$ . Это дает возможность распространить этот подход, как на случай множественной нечеткой регрессии, так и использовать  $R^2$  скорректированный на число степеней свободы в качестве критерия отбора информативности регрессоров.

3. В оформлении текста диссертации допущены некоторые небрежности: математическое ожидание обозначается то как  $E(y/X)$  (с.16), то как  $M_x$  (с.18). В обозначения не всегда можно отличить случайную величину, например  $\varepsilon_i$ , от её значений  $e_i$ . Написание слова «дефазификация» в

диссертации и автореферате различаются.

Перечисленные замечания не влияют на общую положительную оценку работы. Оформление диссертации соответствует требованиям ВАК, работа содержит 151 страницу текста, отличается логичностью построения, изложена ясным и понятным языком. Автореферат полностью соответствует содержанию диссертации. Сапкиной Н.В. опубликовано 14 научных работ, из них 5 – из перечня изданий, рекомендованных ВАК РФ, 7 опубликованы в материалах конференций, имеется 2 свидетельства о государственной регистрации программ.

### 5. Заключение

Высказанные замечания по работе не снижают ее научной ценности и практической значимости.

Диссертация написана грамотным научным языком, что свидетельствует о достаточно высокой профессиональной и общей культуре автора. В работе достаточно полно рассмотрены и решены проблемы исследования, поставленная цель достигнута.

Диссертационная работа Сапкиной Натальи Владимировны «Восстановление закономерностей на основе нечетких регрессионных моделей» является законченной научно-квалификационной работой, посвященной усовершенствованию методов обработки информации, содержит новые научные результаты в этой области, удовлетворяет требованиям «Положения о присуждении ученых степеней», предъявляемым к кандидатским диссертациям. Автор диссертации, Сапкина Наталья Владимировна, заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.17 – «Теоретические основы информатики».

### Официальный оппонент

Профессор кафедры прикладной математики и  
математических методов в экономике  
Воронежского государственного аграрного университета  
имени Императора Петра I  
докт. техн. наук, доцент

Буховец Алексей Георгиевич

Контактная информация:

394087, г. Воронеж, ул. Мичурина, 1,  
тел.: +7 (473) 253-73-69, e-mail: abuhovets@mail.ru.



09.06.2014