

УДК 519.254

Многофакторный контроль заболеваемости клещевым энцефалитом

Скударев А. Г., студент

*Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана,
кафедра «Медико-технические и информационные системы»
skudarevag@student.bmstu.ru*

Научный руководитель: Котин В.В., к.ф.-м.н., доцент

*Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана
кафедра «Медико-технические и информационные системы»
skudarevag@student.bmstu.ru*

Одним из направлений, обеспечивающим «интеллектуальную» поддержку специалистов по решению новых задач анализа временных рядов (ВР), является интеллектуальный анализ данных, в котором анализ поведения и тенденций развития процессов может быть рассмотрен как интеллектуальный анализ временных рядов.

По данным ФБУН Центральный НИИ Эпидемиологии Роспотребнадзора заболеваемость клещевым энцефалитом в период с 2005 по 2012 годы на территории Российской Федерации составляла порядка 2000 – 3000 тысяч человек (см. рис. 1), что является значительной цифрой с учётом специфичности и тяжести болезни.

Клещевой энцефалит - природно-очаговая вирусная инфекция, характеризующаяся лихорадкой, интоксикацией и поражением серого вещества головного мозга. Заболевание может привести к стойким неврологическим и психиатрическим осложнениям и даже к смерти больного. Для заболевания характерна строгая весенне-летняя сезонность заболевания, соответствующая активности клещей.



Рис. 1. Заболеваемость клещевым энцефалитом

Немаловажной особенностью при анализе эпидемиологических временных рядов является учет регионального фактора – зависимость заболеваемости от территориального деления по федеральным округам и субъектам РФ, соответственно, была разработана общая блок-схема обработки данных, представленная на рис. 2.

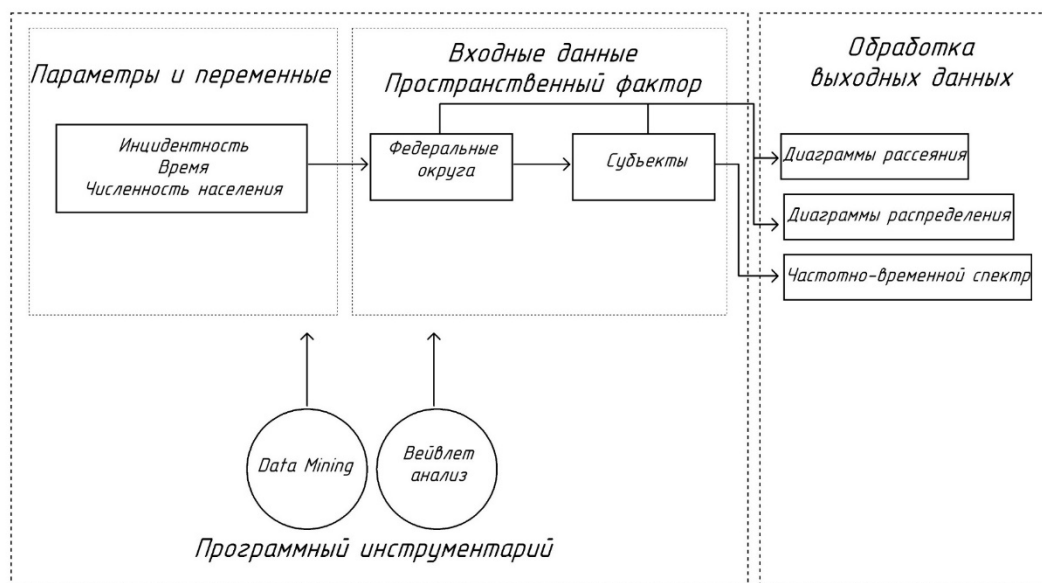


Рис. 2. Общая блок-схема обработки данных

Структурно блок-схема состоит из трех основных частей.

Первая часть - блок параметров и переменных, где загружаются входные данные и устанавливаются входы системы. В работе использовались данные по инцидентности [1] и численности населения в заданный период времени.

Вторая часть – блок разделения входных данных с учетом пространственного фактора, а именно, крупного разделения на федеральные округа и последующего более мелкого на субъекты РФ.

Разделение проводилось в математическом пакете *Statistica* с использованием инструмента *Интерактивное бурение (Interactive Drill Down)*. Задача интерактивного бурения состоит в том, чтобы предоставить пользователю инструмент анализа, сочетающий графические и разведочные методы, который позволит быстро определять распределения переменных и связи между ними, а также определять наблюдения, принадлежащие неким специфическим группам данных [3, 7]. Можно проводить следующие виды анализа на выбранном подмножестве наблюдений: диаграммы размаха, дающие представление о распределении непрерывных переменных; матричные диаграммы рассеяния, дающие представление о связях между непрерывными переменными.

В результате интерактивного бурения рассмотрены зависимости инцидентности от времени и численности населения для всех ФО РФ, результаты представлены на рис. 3.

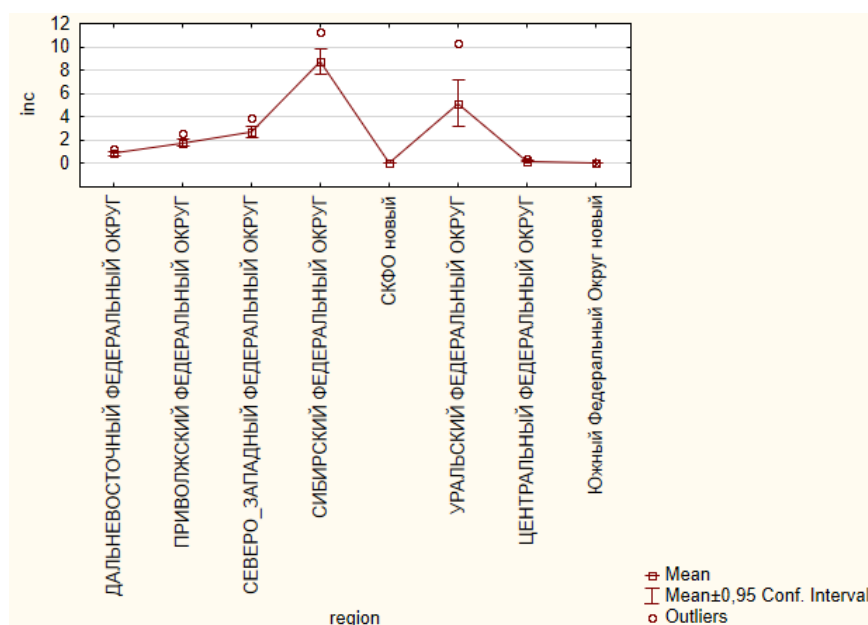


Рис. 3. Распределение инцидентности по ФО: inc – показатель инцидентности; Mean – среднее значение; Conf. Interval – доверительный интервал; Outliers – значения, выходящие за 95% доверительный интервал

В дальнейшем авторы исследовали субъекты Сибирского ФО, как наиболее эндемичного района по результатам интерактивного бурения.

Анализ Сибирского ФО проводился с использованием вейвлетов. В настоящее время вейвлет-анализ нашёл широкое применение в экономике, геологии, астрономии, анализе медицинских данных. В частности, в работе Анятин А.П., Морозов Д.С. «О вейвлет-спектрограммах рядов Кондратьева» [4] исследуются вейвлет-спектрограммы ряда Кондратьева полученные с помощью вейвлетов различного типа. Применение технологии вейвлет-анализа в обработке эпидемиологических данных является новым направлением.

В ходе вейвлет-анализа сначала был выбран «материнский» вейвлет – Gabor wavelet, как наиболее полно отражающий особенности сигнала. На рис. 4 показан результат анализа на примере Сибирского ФО.

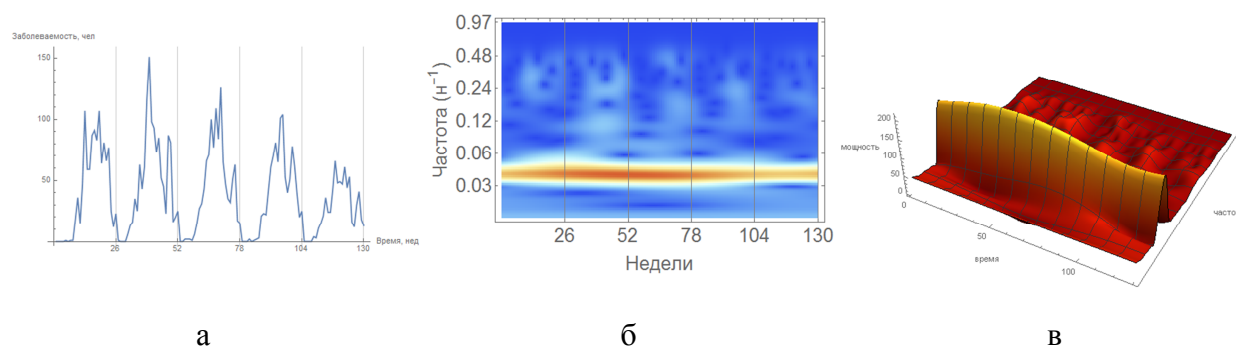


Рис. 4. Вейвлет-анализ Сибирского ФО: а) временной ряд заболеваемости; б) 2D-скалограмма временного ряда; в) 3D-скалограмма временного ряда.

Для временных рядов заболеваемости по субъектам были построены скалограммы. Вместо пространственного изображения поверхности часто используются её проекция на плоскость с изображением линий уровня (изоуровней с применением цветной или чёрно-белой шкалы), которые позволяют проследить изменения интенсивностей коэффициентов вейвлет-преобразования на разных временных масштабах.

Для сравнения заболеваемости в субъектах со всем регионом в целом была найдена разность в скалограммах, как модуль результата попиксельного вычитания значений яркости между ними. Результаты представлены на рис. 5.

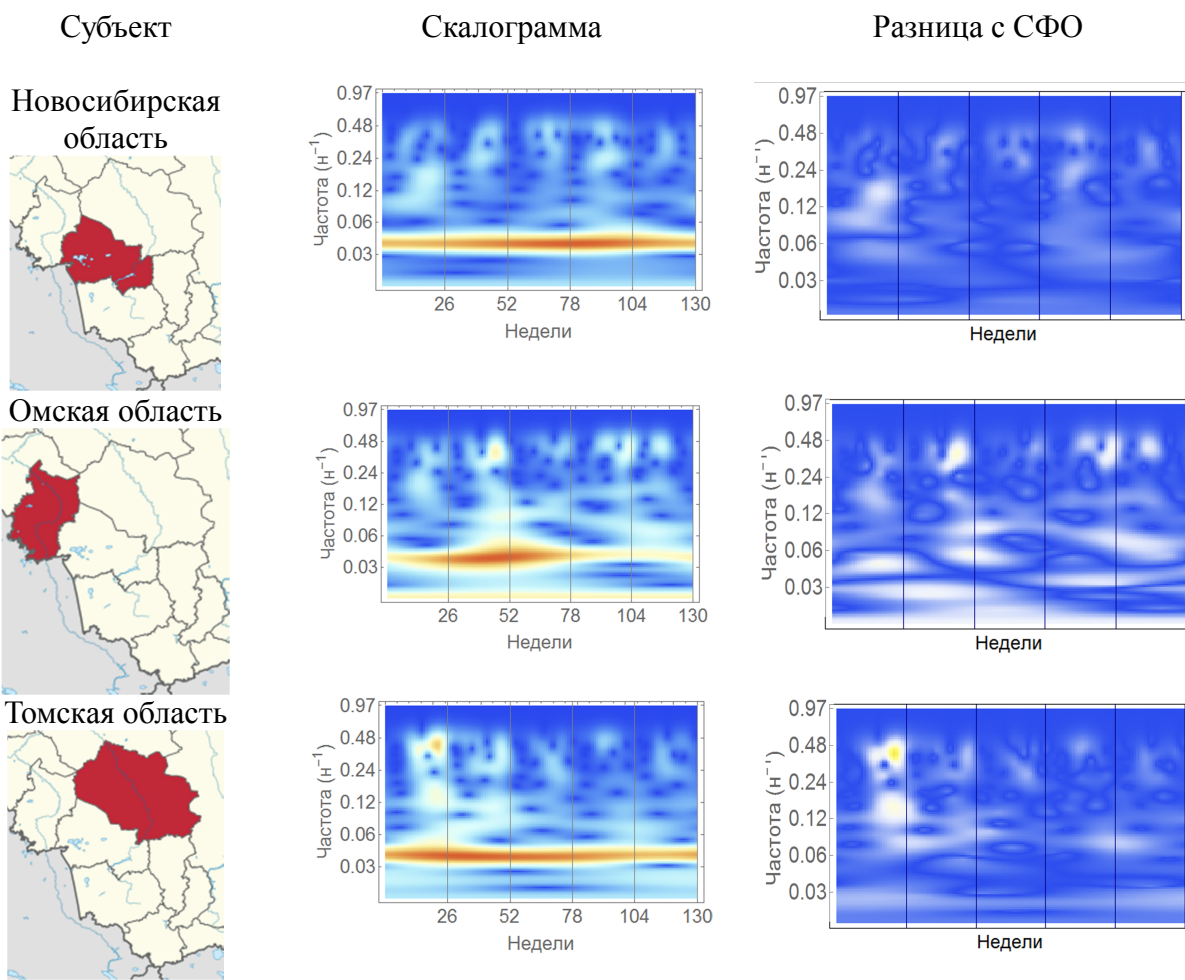


Рис. 5. Анализ скалограмм

Анализируя результат, можно сказать, что в исследуемых временных рядах ярко выражена сезонная периодичность, также наблюдается двухгодичный период и внутригодовая динамика. Данные представленные в таком виде помогут экспертам в области эпидемиологии увидеть скрытые особенности временных рядов.

Третья часть – блок обработки выходных данных, по сути является графическим представлением результатов анализа второй части, и, в то же время, входным блоком для дальнейшей работы, перспективным направлением которой авторы видят в использовании системы нечеткого вывода.

Для моделирования, проектирования и анализа получили распространение экспертные системы, оперирующие экспертными оценками, основанными на опыте и знаниях эксперта (или группы экспертов).

При сравнении понятий измерения и оценки, можно отметить следующее:

экспертная оценка – это измерение, основанное на знаниях эксперта;

экспертная оценка – это интерпретированное количество;

экспертная оценка – это нечеткое отношение по своей природе, выраженное лингвистически.

При анализе ВР эксперт представляет свои суждения с помощью нечетких оценок, относящихся ко многим объектам [8]:

временные области: интервалы времени (несколько дней), абсолютная или относительная позиция на временной шкале (близкое будущее), периодические или сезонные интервалы;

ранг значений ВР (высокая инцидентность, низкая инцидентность);

набор паттернов ВР (быстро растущий, слегка выпуклый);

набор ВР, их атрибутов, как элементов системы (зависимость инцидентности от времени при некотором заболевании);

набор отношений между ВР, атрибутами или элементами (тесно связанный);

множество значений возможности или вероятности (непохоже, очень возможно).

Экспертная оценка некоторой величины является по существу качественной интервальной оценкой локального состояния. Упорядоченная во времени последовательность таких оценок состояний системы и объектов представляет собой временной ряд экспертных оценок. Характерной особенностью такого временного ряда является нечеткость его значений, вытекающая из природы экспертных оценок, поэтому такой временной ряд относится к классу нечетких временных рядов [9].

Модель описания поведения систем на естественном (или близком к естественному) языке в виде приближенных рассуждений в теории нечетких множеств и нечеткой логики, основанная на композиционном правиле вывода, называется системой нечеткого логического вывода [10], структура которой отображена на рис. 6.

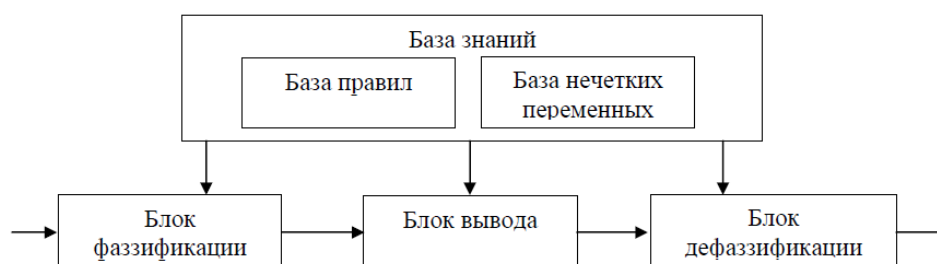


Рис. 6. Система нечеткого вывода

База правил хранит множество логических правил вывода, а также их порядок (иерархическую структуру) применения. База нечетких переменных содержит названия

лингвистических термов и параметры их функций принадлежности. База правил вместе с базой нечетких переменных образуют базу знаний системы нечеткого вывода.

Заключение

Неопределенность в структуре модели и её параметров – это задача которая до сих пор полностью не решена. Кроме того, гибридные модели, сочетающие нечеткую логику, вейвлет анализ и распределенную обработку данных (так называемые мягкие вычисления) имеют огромный потенциал для практического применения. Математические задачи, связанные с процессом нечеткого моделирования, однако, по-прежнему очень серьезные препятствия, которые необходимо преодолеть.

Список литературы

1. Джон М. Ласт. Эпидемиологический словарь. М.: Глобус, 2009. 316 с.
2. Покровский В.И. Общая эпидемиология с основами доказательной медицины. М.: ГЭОТАР-Медиа, 2008. 400 с.
3. Боровиков В.П. STATISTICA. Искусство анализа данных на компьютере. М.: Питер, 2001. 688 с.
4. Аняутин А.П., Морозов Д.С. О вейвлет-спектрограммах рядов Кондратьева // Вестник Российского нового университета. 2012. №4. С. 42–45
5. Витязев В.В. Вейвлет-анализ временных рядов: учебное пособие. СПб.: Изд-во С.-Петербург. ун-та, 2001. 58 с.
6. Скворцов С.П. Основы применения вейвлет-преобразования для фильтрации и сжатия биомедицинских данных: учебное пособие. М.: Изд-во МГТУ им. Н.Э. Баумана, 2012. 67 с.
7. Oded Maimon, Lior Rokach Data mining and knowledge discovery handbook. NY: Springer Science, 2005. 1383 p.
8. Батыршин И.З., Шереметов Л.Б. Модели и методы перцептивного дата майнинга временных рядов для систем поддержки принятия решений // Нечеткие системы и мягкие вычисления. 2007. Т. 2. № 1. С. 37–44
9. Ярушкина Н.Г. Основы теории нечетких и гибридных систем: учебное пособие. М.: Финансы и статистика, 2004. 320 с.
10. Ярушкина Н.Г., Афанасьева Т.В., Перфильева И.Г. Интеллектуальный анализ временных рядов: учебное пособие. Ульяновск: УлГТУ, 2010. 320 с.

11. Рутковская Д., Пилиньский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы: пер. с польск. И. Д. Рудинского. М.: Горячая линия - Телеком, 2006. 452 с. [D. Rutkowska, M. Pilinski, L. Rutkowski Neural networks, genetic algorithms and fuzzy systems. Warsaw: PWN, 1997.].