

ТЕХНОЛОГИЯ ОЦЕНКИ И УМЕНЬШЕНИЯ ПОГРЕШНОСТЕЙ
ПРИ ПРОГНОЗИРОВАНИИ ЭПИЗООТИЙ ЧУМЫ
НА ОСНОВЕ ИСКУССТВЕННЫХ ИММУННЫХ СИСТЕМ

Г.А. Самигулина, С.В. Чебейко

Институт проблем информатики и управления МОН РК, Казахстан, г. Алматы

Данная работа посвящена вопросам оценки энергетических погрешностей Искусственной иммунной Системы (ИИС) прогнозирования эпизоотий чумы и интенсивности распространения инфекции. Обработка многомерной совокупности данных с помощью ИИС приводит к серьезным проблемам, влияющим на достоверность прогноза. Оценка погрешностей ИИС осуществляется на основе свойств гомологичных белков.

Природные очаги чумы занимают на территории Казахстана около 39%, несут постоянную угрозу распространения эпизоотий чумы и являются опасным источником заражения людей. Данное заболевание отнесено к особо опасным (карантинным) инфекциям ввиду высокой способности к быстрому распространению и большой смертности [1]. У бубонной чумы смертность около 60%, у легочной формы почти 100%. Слово «чума» происходит от арабского слова «джума», то есть «шарик», который возникает в воспаленных лимфатических узлах. Жизнедеятельность людей способствует распространению чумы из первичных очагов по всему миру. До сих пор Всемирная Организация Здравоохранения регистрирует в мире ежегодно десятки тысяч больных и сотни случаев с летальным исходом. Известны вспышки чумы в 2001г. в Казахстане, в 2004г. в Туркменистане и Китае, в 2005г. в Тибете. Большим природным очагом чумы является Афганистан. В Америке есть очаги в Колорадо, Калифорнии и Техасе. Актуальность проблемы предсказания эпизоотий чумы и интенсивности распространения инфекции очевидна не только для Казахстана, но и для всего мира.

Сложные механизмы взаимодействия между членами эпизоотической триады возбудителем чумы - чумным микробом, переносчиком и носителем до сих пор недостаточно хорошо изучены. Для осуществления достоверного прогноза начала эпизо-

отий чумы и интенсивности распространения инфекции разработана интеллектуальная информационная технология прогнозирования на основе перспективного подхода Искусственных Иммунных Систем (ИИС). Иммунные Сети - это адаптивные системы для обработки и анализа данных, которые представляют собой математическую структуру, имитирующую некоторые функции иммунной системы человека и обладающие способностью к обучению, к прогнозированию на основе уже имеющихся временных рядов и принятию решения в незнакомой ситуации. ИИС построены на биологических принципах обработки информации молекулами белка. В основе подхода ИИС лежит идея взаимодействия между белками (пептидами) иммунной системы человека и чужеродными антигенами, то есть в возможности произвольного связывания (молекулярного узнавания) посредством определения минимальной энергии связи между формальными пептидами [2].

Входной информацией для ИИС служат пространственно - временные ряды информативных признаков состояния чумного микроба, носителя, переносчика, различных биотических и абиотических факторов, сформированные по результатам обследования долины Акдала Прибалхашского природного очага чумы [3]. Информация предоставлена Казахским Научным Центром Карантинных и Зоонозных Инфекций им. М. Айкимбаева (KSCQZD) по гран-

ту INCO COPERNICUS «STEPICA» ICA2-ST-2000-10048 «Чума в Центральной Азии - эпидемиологическое исследование, основанное на пространственно - временной динамике». Достоинством данного подхода ИИС является возможность комплексной оценки состояний чумного микроба. Количество анализируемых признаков может быть достаточно большим. Прогностическое значение особенностей популяции чумного микроба связывают с комплексом свойств микроорганизма [4]. У возбудителя чумы обнаружен ряд признаков, определяющих степень вирулентности для носителя чумного микроба и возможность формирования блока преджелудка у блох, что обеспечивает микроорганизму существование на двух экологических этапах, связанных со сменой условий обитания в организмах столь далеко отстоящих друг от друга. Культивирование чумного микроба на искусственных питательных средах дает возможность изучить и определить ряд типичных признаков характерных для всех подвидов, а также выявить ряд особенностей, позволяющих классифицировать по географическому распространению. Изменчивость микроба проявляется в его ферментативной активности по отношению к глицерину, триптофану, рамнозе; в формировании пестицина, лейцина; способности к сорбции железа; формированию окрашенных в черный цвет (P+) колоний и др. Признаки, коррелирующие с проявлением вирулентности, обозначаются как «детерминанты вирулентности»: способность микробов продуцировать Фракцию 1, которая уменьшает эффективность захвата чумного микроба макрофагами, снижает степень внутриклеточного переваривания и является видовым признаком возбудителя чумы; пуринозависимость, определяющая способность проявлять резистентность к фагоцитозу; кальцийзависимость, связанная с формированием VW – антигенов, которые способ-

ствуют устойчивости микроба к фагоцитозу моноцитами; фибринолизинкоагулазная активность, усиливающая проницаемость возбудителя чумы; наличие мышинового токсина и др. Для определения эпидемической сущности штаммов важны не отдельные составляющие, а сочетание всех признаков, влияющих на вирулентность возбудителя чумного микроба.

Выходной характеристикой ИИС является график зараженности секторов по годам. Под зараженным сектором понимается сектор, в котором обнаружен либо зараженный носитель, либо зараженный переносчик, либо зараженная нора.

Однако обработка многомерной совокупности данных на основе ИИС приводит к ряду серьезных проблем, которые влияют на достоверность прогноза. Погрешности энергетических оценок ИИС возникают за счет существующих проблем неполноты данных, их корреляции и погрешности измерения. Определение нативной (функциональной) укладки белковой цепи, соответствующей минимуму энергии затруднено из-за различного рода погрешностей. Известно, что нативная укладка белковой цепи соответствует самому нижнему уровню в спектре энергий белковой цепи и является наиболее стабильной структурой, однако вследствие погрешностей энергетических оценок энергия нативной структуры может быть выше, чем энергия каких-либо других структур и поэтому никак не может быть опознана, в качестве нативной.

Постановка задачи формулируется следующим образом: необходимо разработать технологию оценки и уменьшения погрешностей (ошибок обобщения) при прогнозировании эпизоотий чумы на основе подхода ИИС [5].

Данная задача решается в два этапа. На первом этапе предварительной обработки многомерной совокупности данных погреш-

ности устраняются с помощью методов факторного анализа [6] на основе вращения собственного вектора (метода главных компонент) и развития варимаксного подхода, которые предназначены для перевода данных в наиболее информативное для поставленной задачи представление. Осуществляется это путем уменьшения размерности исходных данных (сжатия информации) вследствие редукции малоинформативных признаков, которые несут в себе ошибки, влияющие на качество обучения ИИС и скорость обучения. Скорость обучения является очень важным показателем интеллектуальных систем, особенно при прогнозировании в реальном масштабе времени. От качества обучения ИИС напрямую зависит достоверность полученного прогноза.

Переход от набора исходных признаков a_1, \dots, a_n к новому набору признаков Z_1, \dots, Z_m осуществляется с наименьшими потерями информации. Подбор новых показателей $Z(A) = (z_1(a), \dots, z_m(a))$ основан на максимизации некоторой меры информативности $I_m(Z(A))$. Конкретный выбор функционала $I_m(Z(A))$ нацелен на максимальное сохранение информации, содержащейся в исходной матрице данных. В результате переходим от множества исходных переменных к существенно меньшему числу новых переменных (факторов). При этом в один фактор объединяются переменные, сильно коррелирующие между собой.

До сих пор нет четкой картины относительно того, где и как сохраняется чумный микроб в период депрессии эпизоотий. С помощью факторного анализа осуществляется исследование структуры взаимосвязей признаков и выявление скрытых (латентных) переменных. Определение основных факторов [7], которые влияют на патогенность штаммов чумного микроба и взаимодействие этих факторов относительно чумной триады осуществляется с помощью вращения собственного вектора. Для этого

необходимо повернуть все оси n -мерной совокупности данных таким образом, чтобы первая ось соответствовала направлению наибольшей дисперсии в данных, а каждая последующая представляла максимум остаточной дисперсии. Это дает возможность произвести сокращение данных, то есть отбрасывание факторов, которые не содержат значимой информации о данных. К полученному факторному решению применена процедура вращения. Вращение осуществляется с помощью варимаксного подхода. Сущность варимаксного вращения заключается в том, чтобы максимизировать дисперсию каждого вектора. В результате ортогонального преобразования увеличивается число переменных с очень большими, или с очень малыми нагрузками. При этом уменьшается число переменных со средними нагрузками, что существенно облегчает задачу выделения малоинформативных признаков. Согласно разработанному алгоритму преобразования исходных данных получены результаты, которые позволяют уменьшить размерность анализируемого признакового пространства и построить оптимальную структуру иммунной сети [8] по весовым коэффициентам информативных признаков.

На втором этапе после решения задачи распознавания образов оценка энергетических погрешностей осуществляется [9, 10] с использованием свойств гомологов. Гомологичные белки (белки, имеющие одинаковое происхождение) имеют примерно одинаковые структуры, несмотря на многочисленные мутации в аминокислотной последовательности. Нативная укладка цепи может быть найдена в процессе сворачивания цепи, где потенциал каждого взаимодействия усреднен по гомологам.

Разработан пакет прикладных программ «PERTID», который предназначен для реализации поставленной задачи с помощью искусственных иммунных сетей [11, 12]. Данное программное обеспечение реализо-

вано на языке программирования DELPHI 7.0. В состав предложенного программного продукта входит девять взаимосвязанных

подпрограмм, выполняющих определенные задачи.



Рисунок 1 - Структурная схема пакета прикладных программ «PEPTID»

Входными характеристиками ИИС являются временные ряды, составленные из

различных параметров описывающих рассматриваемую систему. Вся необходимая

информация заносится в базы данных (БД) и базы знаний (БЗ). Для построения иммунной сети в качестве признаков берутся поля таблиц БД. Структурная схема пакета прикладных программ «PERTID» обработки информации искусственной иммунной системой приведена на рисунке 1.

Для нахождения минимума энергии связи необходимо осуществить сингулярное разложение матриц. Так как идет обработка многомерных данных, то для осуществления поставленной задачи используется пакет прикладных программ MATLAB 7.1. Базы данных матриц эталонов и матриц образов хранятся в EXCEL, поэтому необходимо подключить специальное программное обеспечение, которое обеспечивает возможность доступа данных EXCEL для обработки в MATLAB. После сингулярного разложения матриц в MATLAB данные экспортируются обратно в базу данных EXCEL, где согласно разработанным алгоритмам находятся энергии связи. Полученные результаты отображаются в подпрограмме MIN_ENERG, где создается база данных энергий связи.

Подпрограмма ENERG_ERRORS служит для оценки энергетических ошибок при решении задачи распознавания образов и обработки многомерной информации на основе анализа свойств гомологичных пептидов. Для выбранной базы данных энергий связи между пептидами вводится рассматриваемый класс. Затем определяется количество гомологичных белков для данного класса. Вычисляются среднее значение по гомологам, среднее значение стандартных отклонений по гомологам, определяется значение Z- факторов для всех гомологов.

В подпрограмме K_PREDIC осуществляется расчет коэффициентов риска прогнозирования на основе Z – факторов. Чем выше коэффициент риска прогнозирования, тем меньше достоверность прогнозирования на основе ИИС. На основе полученных результатов осуществляется корректировка прогноза и выбор правильного решения по управлению процессом.

Заключительная подпрограмма GRAFIC позволяет с помощью пакета прикладных программ MATLAB 7.1 построить график энергий связи в трехмерном пространстве, то есть сгруппировать данные для визуального представления и наглядно выделить зоны неоднозначного решения для проведения дополнительных процедур.

Получены авторские свидетельства на разработанные программы [13, 14].

Эпизоотический процесс при чуме в природном очаге Акдала Республики Казахстан характеризуется сложными взаимодействиями между членами чумной триады. Поэтому изучение механизмов взаимодействия чумного микроба с носителем и переносчиком на основе нового подхода ИИС для прогноза начала эпизоотий и интенсивности распространения инфекции, применение методов факторного анализа позволяют специалистам глубоко проанализировать скрытые (латентные) взаимодействия признаков, различные факторы и их сочетания в зависимости от фаз эпизоотического процесса. На основе этих данных возможно рассмотрение совершенно новых гипотез и методов противоэпидемиологической борьбы с чумой.

Литература: [1] Айкимбаев А.М. и др. Введение в эпидемиологическую ситуацию горного очага чумы Тянь-Шаня и Алтая. -Алматы, 1991. – 125с.; [2] Tarakanov A.O. Formal peptide as a basic of agent of immune networks: from natural prototype to mathematical theory and applications // Proc. of the I Int. workshop of central and eastern Europe on Multi-Agent Systems.-1999.; [3] Sokolova S.P., ..., Samigulina G.A. Artificial Immune System for the gerbil natural plagues focus. /Edited by A.O. Tarakanov. –Almaty, 2002. -180p. ; [4] Соколова С.П., Темиральева Г.А., Самигулина Г.А. и др. Комплексная оценка состояний чумного микроба // Труды Междунар. конф. «Проблемы управления и информатики». – Бишкек, 2000. –С.426-428; [5] Самигулина Г.А. Разработка интеллектуальных экспертных систем управления на основе технологии искусственных иммунных систем. –Алматы: ИПИУ МОН РК, 2008. -137с.; [6] К. Иберла. Факторный анализ. -

М.: Статистика. -1980.; [7] *Самигулина Г.А., Чебейко С.В.* Выделение информативных признаков штаммов чумного микроба для прогнозирования эпизоотий чумы //Проблемы нелинейного анализа в инженерных системах. Казанский государственный технический университет им. А.Т.Туполева. –Казань, 2004 .-Вып. №1(20). –Т.10. -С.83-95.; [8] *Самигулина Г.А.* Моделирование оптимальной структуры иммунной сети в интеллектуальных экспертных системах поддержки принятия решений //Информационные технологии моделирования и управления. –Воронеж, 2008. –Вып.4(47). –С.313-317.; [9] *Samigulina G.A., Chebeiko S.V.* Technology of elimination errors the energy estimations of Artificial Immune Systems of the forecasting plague //Proceedings on the sixth international conference on Computational Intelligence and Natural Computation. -Cary, North Carolina, USA. -2003. -P.P. 1693-1696.; [10] *Самигулина Г.А., Чебейко С.В.* Процедура оценки энергетической погрешности Искусственной Иммунной Системы прогнозирования чумы на основе гомологов //Вестник КазНУ им. Аль-Фараби. -Алматы. -2003. -N3(38). -С. 61-66.; [11] *Самигулина Г.А., Самигулина З.И.* Разработка программного обеспечения для интеллектуальной системы прогнозирования на основе Искусственных Иммунных Систем //Тезисы V Всесибирского конгресса женщин-математиков. –Красноярск, 2008, http://www.sfu-kras.ru/liles/spis_t.doc.; [12] *Самигулина Г.А., Самигулина З.И.* Разработка пакета прикладных программ для обработки, анализа и прогноза данных Искусственными Иммунными Сетями //Вестник Национальной Инженерной Академии Республики Казахстан. – Алматы, 2008. -№ 1(27). –С.39-44.; [13] *Самигулина Г.А.* Предварительная обработка данных AIS (программа для ЭВМ). Свидетельство о государственной регистрации объекта интеллектуальной собственности №286. Зарегистрировано в Комитете по правам интеллектуальной собственности Министерства юстиции Республики Казахстан. –Астана, 2008.; [14] *Самигулина Г.А., Самигулина З.И.* Оценка энергетических погрешностей искусственной иммунной системы по гомологам (программа для ЭВМ). Свидетельство о государственной регистрации объекта интеллектуальной собственности № 396. Зарегистрировано в Комитете по правам интеллектуальной собственности Министерства юстиции Республики Казахстан. – Астана, 2008.

Принято в печать 10.06.2009

УДК 550.36+577.31

ТЕХНОЛОГИЯ ОЦЕНКИ И УМЕНЬШЕНИЯ ПОГРЕШНОСТЕЙ ПРИ ПРОГНОЗИРОВАНИИ ЭПИЗОТИЙ ЧУМЫ НА ОСНОВЕ ИСКУССТВЕННЫХ ИММУННЫХ СИСТЕМ

Г.А.Самигулина, С.В.Чебейко

Институт проблем информатики и управления МОН РК, Казахстан, г. Алматы
050026, ул. Богенбай батыра, д.221, ком. 341, раб. тел. 254-73-73 E-mail: galinasamigulina@mail.ru

TECHNOLOGY OF ESTIMATION AND DECREASE THE ERRORS AT PREDICTION EPIZOOTIC OF PLAGUE ON THE BASIS OF ARTIFICIAL IMMUNE SYSTEMS

Galina A. Samigulina, Svetlana V. Chebeiko

Institute of Informatics and Control Problems, Kazakhstan, Almaty,
050026, Bogenbay-batira, h.221, ap. 341, tel. 254-73-73 E-mail: galinasamigulina@mail.ru

The given work to the question of an estimation of the energy of the error of Artificial Immune System (AIS) forecasting the epizootic plagues and the intensity of spreading infection is devoted. Processing of the multidimensional data set with help AIS to results in the number of the serious problems influencing on the reliability of the forecast. The estimation of errors AIS is carrier out the basis of properties homologous proteins.

ЖАСАНДЫ ИММУНДЫҚ ЖҮЙЕ НЕГІЗІНДЕ ОБАНЫҢ ЭПИЗООТТІН БОЛЖАУ НЕЗІНДЕ КЕТЕТІН ҚАТЕЛЕРДІ АЗАЙТУ ЖӘНЕ БАҒАЛАУ ТЕХНОЛОГИЯСЫ

Г.А.Самигулина, С.В.Чебейко

Информатика және басқару проблемалары институты
050026, Алматы қ., Богенбай батыр к., 221. Тел:254- 73- 73. Факс:254 73 09
E-mail: galinasamigulina@mail.ru

Ұсынылып отырған жұмыс обаның эпизооті мен індеттің жайылу қарқындылығын болжаушы Жасанды Иммундық Жүйенің (ЖИЖ) энергетикалық катесін бағалау сұрақтарына арналған. Көпмәлшерлі мағұлыматтар жиынын ЖИЖ көмегімен өңдеу, ақиқатты болжамға әсер ететін көптеген проблемаларға әкеліп соғады. ЖИЖ кателігін бағалау гомологиялы ақуыздың қасиеттері негізінде жасалады.