

3-6-2020

APPLICATION OF IMMUNE ALGORITHM IN INTELLECTUAL PROCESSING OF INFORMATION

D. T. Muxamedieva

Scientific-innovative center of information and communication technologies at the Tashkent University of Information Technologies named after Muhammad al-Khorazmiy

Z Juraev

Scientific-innovative center of information and communication technologies at the Tashkent University of Information Technologies named after Muhammad al-Khorazmiy

Follow this and additional works at: <https://uzjournals.edu.uz/ferpi>

Recommended Citation

Muxamedieva, D. T. and Juraev, Z (2020) "APPLICATION OF IMMUNE ALGORITHM IN INTELLECTUAL PROCESSING OF INFORMATION," *Scientific-technical journal*: Vol. 3 : Iss. 1 , Article 2.

Available at: <https://uzjournals.edu.uz/ferpi/vol3/iss1/2>

This Article is brought to you for free and open access by 2030 Uzbekistan Research Online. It has been accepted for inclusion in Scientific-technical journal by an authorized editor of 2030 Uzbekistan Research Online. For more information, please contact sh.erkinov@edu.uz.

APPLICATION OF IMMUNE ALGORITHM IN INTELLECTUAL PROCESSING OF INFORMATION**Muxamedieva D.T., Juraev Z.**

Scientific-innovative center of information and communication technologies at the Tashkent University of Information Technologies named after Muhammad al-Khorazmiy

ПРИМЕНЕНИЕ ИММУННОГО АЛГОРИТМА ПРИ ИНТЕЛЛЕКТУАЛЬНОЙ ОБРАБОТКЕ ИНФОРМАЦИИ**Мухамедиева Д.Т., Жўраев З.**

Научно-инновационный центр информационно-коммуникационных технологий при Ташкентском университете информационных технологий имени Мухаммада ал-Хоразмий

МАЪЛУМОТЛАРНИ ИНТЕЛЛЕКТУАЛ ҚАЙТА ИШЛАШДА ИММУН АЛГОРИТМИНИ ҚЎЛЛАНИЛИШИ**Мухамедиева Д.Т., Жўраев З.**

Мухаммада ал-Хоразмий номидаги Тошкент Ахборот технологиялари Университети ҳузуридаги Ахборот-коммуникация технологиялари илмий-инновацион маркази

Abstract. *Immune algorithms are widely used in various fields of intellectual information processing. The properties of artificial immune systems, such as recognition, diversity, learning, memory, allow using immune principles to solve such problems as pattern recognition, data search, computer security, error detection, classification, optimization, etc. The article discusses the use of an immune algorithm to solve a problem assessing the state of complex processes.*

Keywords. Immune algorithms, evaluation, data mining, complex process, classification.

Аннотация. *Иммунные алгоритмы широко используются в различных областях интеллектуальной обработки информации. Свойства искусственных иммунных систем, такие как распознавание, разнообразие, обучение, память позволяют использовать иммунные принципы для решения таких задач как распознавание образов, поиск данных, компьютерная безопасность, обнаружение ошибок, классификация, оптимизация и др. В статье рассматривается применение иммунного алгоритма для решения задачи оценки состояния сложных процессов.*

Ключевые слова. Иммунные алгоритмы, оценка, интеллектуальный анализ данных, сложный процесс, классификация.

Аннотация. *Маълумотларни интеллектуал таҳлил қилишининг турли соҳаларида иммун алгоритмлар кенг қўлланилмоқда. Сунъий иммун тизимларининг таниб олиш, турли хиллик, ўқитиш, хотира каби хусусиятлари иммун тамойилларини белгиларни таниб олиш, маълумотларни ахтариш, компьютер хавфсизлиги, хатоларни аниқлаш, синфлаштириш, оптималлаштириш каби масалаларни ечишда қўллашга имкон беради. Мақолада иммун алгоритми мураккаб жараёнларни баҳолаш масалаларини ечишга қўллаш қўрилган.*

Таянч сўзлар. Иммун алгоритмлари, баҳолаш, маълумотларни интеллектуал таҳлил қилиш, синфлаштириш.

ENERGETICS, THE ELECTRICAL ENGINEERING, ELECTRONIC DEVICES AND INFORMATION TECHNOLOGIES

1. Введение. В искусственных иммунных системах (ИИС) используется способность естественной иммунной системы вырабатывать новые типы антител и отбирать наиболее подходящие из них для взаимодействия с попавшими в организм антигенами [1]. Для объяснения иммунологических механизмов существуют различные теории – теория иммунной сети, принципы клонального и негативного отбора.

Естественная иммунная система состоит из большого количества защитных элементов, молекул и органических соединений, поддерживающих организм в здоровом состоянии, борясь с болезнями, служащими причиной заболеваний. Защитные элементы, используемые в естественной иммунной системе, называются лимфоцитами (белые кровяные тельца), главная задача которых – борьба с антигенами, молекулами, принадлежащими чужеродным телам, таким, как бактерии или вирусы, которые внедрились в организм.

Защитная реакция организма в борьбе с болезнью состоит в том, что он начинает вырабатывать клетки (антитела), способные распознать и нейтрализовать антигены. Клетки, полученные в результате мутационного процесса, имеют большое сходство с антигенами, имеют большее время жизни и остаются в организме на случай, если в будущем атака повторится (память клетки). Пока процент клонирования прямо пропорционален сходству с антигенами, процент мутации обратно пропорционален такой схожести, поэтому, чем ближе клетка к антигену, тем меньше процент ее мутации. С другой стороны, если сходство антигена и клетки очень низкое, высокий процент мутации применяется в надежде повысить значение сходства [1].

Для эволюционных алгоритмов известно [2], что сходимость к глобальному оптимуму в задаче оптимизации достигается в том случае, если есть уверенность в том, что алгоритм находит решение за конечное число шагов, и если такое решение будет оставаться в дальнейшем в популяции. Поскольку состояния переходов эволюционных алгоритмов имеют стохастический характер, детерминированная концепция сходимости не может быть использована для определения срока действия таких алгоритмов. Существуют две широко используемые меры стохастической сходимости эволюционных алгоритмов – это полное совпадение и совпадение по значению [3].

В вычислительных процедурах иммунокомпьютинга в качестве аналога расстояния используется понятие энергии связи, основанное на сингулярном разложении матрицы. Энергия связи между объектами A и M представляется следующим образом:

$$\omega_i = -U_i^T M V_i, \quad U_i^T U_i = 1, \quad V_i^T V_i = 1, \quad i = 1, 2, \dots, r,$$

где U_i, V_i – соответственно, правые и левые сингулярные векторы матрицы A , r – ранг матрицы.

2. Алгоритм вычислительной процедуры. Алгоритм вычислительной процедуры обучения с экспертом состоит из следующих шагов:

Шаг 1. Сворачивание вектора в матрицу. Заданный вектор X размерности $(n \times 1)$ сворачиваем в матрицу M размерности $n_U \times n_V = n$.

Шаг 2. Формируем матрицы A_1, A_2, \dots, A_k для эталонных классов $c = 1, \dots, k$ и вычисляем их сингулярные векторы:

$$\{U_1, V_1\} - \text{для } A_1, \{U_2, V_2\} - \text{для } A_2, \{U_k, V_k\} - \text{для } A_k.$$

Шаг 3. Распознавание. Для каждого входного образа M вычисляем k значений энергии связи между каждой парой сингулярных векторов:

$$\omega_1 = -U_1^T M V_1, \dots, \omega_k = -U_k^T M V_k.$$

Шаг 4. Определяем класс, к которому принадлежит входной образ M . Минимальное значение энергии связи ω^* определяет этот класс:

$$c = \omega^* = \min_c \{\omega_c\}.$$

ENERGETICS, THE ELECTRICAL ENGINEERING, ELECTRONIC DEVICES AND INFORMATION TECHNOLOGIES

Основные вычислительные процедуры оценки состояния слабоформализуемых процессов основаны на свойствах сингулярного разложения произвольных матриц и математическом аппарате иммунокомпьютинга.

Как известно, для произвольной матрицы A размерности $(m \times n)$ существует так называемое сингулярное разложение, т.е. представление матрицы в виде

$$A = USV^T, \quad (1)$$

где $U - (m \times m)$ и $V - (n \times n)$ – ортогональные квадратные матрицы, удовлетворяющие критерию ортогональности:

$$VV^T = V^T V = E_{m \times m}, \quad UU^T = U^T U = E_{n \times n}, \text{ где } E - \text{единичные матрицы соответствующих размерностей.}$$

Матрица S состоит из квадратного диагонального блока размерности $r \times r$ ($r = \min(m, n)$) с неотрицательными элементами на главной диагонали и, если $n \neq m$, из дополнительных нулевых строк или столбцов

$$S = [S'; 0], \text{ если } m < n, \quad S = [S'; 0]^T, \text{ если } m > n,$$

$$S = S', \text{ если } m = n, \quad S' = \text{diag}\{s_1, s_2, \dots, s_r, s_1 \geq s_2 \geq \dots \geq s_r\}.$$

Числа $s_i, i = 1, 2, \dots, r$ называются сингулярными числами матрицы A , которые определяются матрицей A однозначно.

Сингулярное разложение вещественной прямоугольной матрицы A в покомпонентной форме имеет следующее представление:

$$A = s_1 U_1 V_1^T + s_2 U_2 V_2^T + \dots + s_r U_r V_r^T, \quad (2)$$

где s_i – сингулярные числа матрицы A , U_i, V_i – соответственно, правые и левые сингулярные векторы, r – ранг матрицы. Эти сингулярные числа и сингулярные векторы удовлетворяют следующим соотношениям:

$$s_1 \geq s_2 \geq \dots \geq s_r \geq 0, \quad s_i = U_i^T A V_i, \quad U_i^T U_i = 1, \quad V_i^T V_i = 1, \quad i = 1, \dots, r. \quad (3)$$

Известно, что процессы сингулярного разложения для любой вещественной матрицы A обладают весьма полезными свойствами для теории и приложений, а именно, каждая матрица над полем вещественных чисел имеет вещественные сингулярные числа и векторы. Кроме того, сингулярное разложение матриц устойчиво к малым возмущениям матриц, т.е. сингулярное разложение каждой матрицы является хорошо обусловленной процедурой.

Относительно практических аспектов, сингулярное разложение матрицы в общем случае может быть получено по достаточно простой и надежной схеме:

$$\begin{aligned} V_{(k+1)}^T &= U_{(k)}^T A, \\ V_{(k+1)} &= V_{(k+1)} / |V_{(k+1)}| \\ U_{(k+1)} &= A V_{(k+1)}, \quad U_{(k+1)} = U_{(k+1)} / |U_{(k+1)}| \\ s_k &= U_k^T A V_k, \quad |s_{k+1} - s_k| \leq \varepsilon, \end{aligned} \quad (4)$$

где $k=0, 1, 2, \dots$ – номер итерации, $|U_{(k+1)}|$ – любая векторная норма, ε – заданная точность вычисления. Можно показать, что для произвольных начальных векторов $U_{(0)}, V_{(0)}$ итерации по схеме (4) сходятся в общем случае к сингулярным векторам U, V , соответствующим максимальному сингулярному числу $s_{\max} = U^T A V$.

Следует отметить, что такие свойства не свойственны спектральному разложению, которое в действительности формирует основу для многомерного статистического анализа. В отличие от сингулярного разложения матриц, собственные числа и собственные векторы спектрального разложения являются вещественными только для вещественных

ENERGETICS, THE ELECTRICAL ENGINEERING, ELECTRONIC DEVICES AND INFORMATION TECHNOLOGIES

симметрических матриц, в общем случае не симметрические вещественные матрицы обладают комплексным спектром и определить его не просто.

С использованием вышеприведенного итеративного алгоритма (4) сингулярное разложение матрицы A , представленное в форме (2),(3) может быть получено с использованием метода исчерпывания.

Сущность этого метода заключается в следующем:

- максимальное сингулярное число и соответствующие ему правый и левый сингулярный векторы матрицы A вычисляются с помощью итеративного алгоритма (4). Формируется матричная компонента $A_1 = s_1 U_1 V_1^T$;

- формируется матрица невязки

$$A_2 = A - A_1 = A - s_1 U_1 V_1^T \quad (5)$$

для которой максимальное сингулярное число и соответствующие ему правый и левый сингулярный векторы матрицы A_2 вычисляются с помощью итеративного алгоритма (4) и т.д.

4. Решения задачи классификации на основе применения искусственных иммунных систем.

Новая вычислительная процедура, называемая искусственные иммунные системы, основанная на принципах иммунной системы, обладает способностью обучаться новой информации, запоминать ранее полученную информацию и осуществлять распознавание образов и анализ данных на основе принципов биомолекулярного узнавания в высоко распределенной манере. Эти системы предлагают мощные и робастные возможности обработки больших массивов информации для решения сложных задач. Строгий математический базис ИИС основан на биологическом прототипе иммунной сети и понятиях формального протеина и формальной иммунной сети (ФИС).

Предварительные операции включают в себя создание баз данных на основе статистических временных рядов, характеризующих рассматриваемую систему; создание баз знаний на основе мнений экспертов; нормирование входных параметров; удаление пробелов в таблицах данных; операции по считыванию информации из баз данных, баз знаний и т.д.

Основные операции заключаются в создании иммунной сети, обучении иммунной сети с учителем и т.д.

Вспомогательные операции состоят из таких как: выделение информационно - ценных признаков и снижение размерности анализируемого пространства признаков; создание оптимальной структуры иммунной сети; тестирование; оценка энергетических погрешностей ИИС при распознавании образов и т.д.

База знаний содержит сведения, которые отражают закономерности, существующие в рассматриваемой предметной области, позволяют выводить новые знания и прогнозировать потенциально возможные состояния исследуемой области; сведения о структуре и содержании базы данных; сведения по языку общения; метазнания, определяющие способы представления и переработки знаний. В базу знаний помещаются как общедоступные данные, так и знания эксперта в данной предметной области, вычислительные алгоритмы реализации процедур иммунокомпьютинга, результаты группировки и автоматической классификации, а также интерпретация результатов вычислений.

Нечеткие логические уравнения задачи классификации (типа если <вход>, то <выход>) вместе с функциями принадлежности нечетких термов позволяют принимать решение с использованием следующего алгоритма:

1. Фиксируются значения параметров состояния объекта:

$$X = [x_1, \dots, x_n].$$

2. Определяются значения функций принадлежности $\mu^j(x_i^*)$, при фиксированных значениях параметров $x_i^*, i = \overline{1, n}$.

ENERGETICS, THE ELECTRICAL ENGINEERING, ELECTRONIC DEVICES AND INFORMATION TECHNOLOGIES

3. Используя логические уравнения, вычисляются значения функций принадлежности $\mu^{r_j}(x_1^*, x_2^*, \dots, x_n^*)$ при векторе состояния $X^* = (x_1^*, x_2^*, \dots, x_n^*)$.

4. Определяется решение r_j^* , для которого:

$$\mu^{r_j^*}(x_1^*, x_2^*, \dots, x_n^*) = \max_{j=1, n} [\mu^{r_j}(x_1^*, x_2^*, \dots, x_n^*)].$$

Суть обучения состоит в подборе таких параметров функций принадлежности, которые минимизируют различие между результатами нечеткой аппроксимации и реальным поведением объекта.

Определим задачу классификации как отображение $f(X) \rightarrow \{1, \dots, c\}$ любого образа X в одно из целых чисел $1, \dots, c$, которые представляют классы.

Задача классификации может быть сформулирована следующим способом.

Дано:

- число классов c ;

- набор из m обучающих образов: X_1, X_2, \dots, X_m ;

- класс любого обучающего образа: $f(X_1) = c_1, \dots, f(X_m) = c_m$;

- произвольный n -мерный вектор Z .

Найти:

Класс вектора Z : $f(Z) = ?$

Процесс обучения состоит из следующих этапов:

1. Сформировать обучающую матрицу $A = [X_1, \dots, X_m]^T$ размерности $m \times n$.

2. Вычислить максимальное сингулярное число s , а также левый и правый сингулярные векторы L и R обучающей матрицы по следующей итеративной (эволюционной) схеме:

$$L_{(0)} = [1, \dots, 1]^T,$$

$$R^T = L_{(k-1)}^T A, \quad R_{(k)} = R / |R|, \quad \text{где } |R| = \sqrt{r_1^2 + \dots + r_n^2},$$

$$L = AR_{(k)}, \quad L_{(k)} = L / |L|, \quad \text{где } |L| = \sqrt{l_1^2 + \dots + l_n^2},$$

$$s_{(k)} = L_{(k)}^T AR_{(k)}, \quad k = 1, 2, \dots,$$

до выполнения условия

$$|s_{(k)} - s_{(k-1)}| < \varepsilon,$$

$$s = s_{(k)}, \quad L = L_{(k)}, \quad R = R_{(k)}.$$

3. Хранить сингулярное число s .

4. Хранить правый сингулярный вектор R (как «антитело-пробу»).

5. Для всякого $i=1, \dots, m$ хранить компоненту l_i левого сингулярного вектора L и класс c_i соответствующий обучающему образу X_i .

6. Для всякого n -мерного образа Z вычислить его энергию связи с R :

$$w(z) = Z^T R / s.$$

7. Выбрать l_i , которая имеет минимальное расстояние d с w :

$$d = \min_i |w - l_i|, \quad i=1, \dots, m,$$

и считать класс c_i искомым классом образа Z .

В общем виде один шаг работы иммунного алгоритма можно представить следующим образом [4]:

$$\forall i \in \{1, \dots, m\} : x_i' = \text{mut}(\text{clon}(x_1, \dots, x_n))$$

ENERGETICS, THE ELECTRICAL ENGINEERING, ELECTRONIC DEVICES AND INFORMATION TECHNOLOGIES

$$\begin{aligned}(x_1'', \dots, x_k'') &= \text{aging}(x_1, \dots, x_n, x_1', \dots, x_m') \\ (y_1, \dots, y_n) &= \text{sel}(x_1'', \dots, x_k'')\end{aligned}\quad (6)$$

где $(x_1, \dots, x_n) \in X^n$ – текущая популяция антител; (x_1', \dots, x_m') – популяция антител, возникающая в результате клонирования и мутации; (x_1'', \dots, x_k'') – антитела, которые удаляются из популяции; (y_1, \dots, y_n) – антитела, добавляемые в текущую популяцию. Как следует из (6), ИА присущи следующие операторы: клонирование, мутация, старение и селекция. Рассмотрим подробнее эти операторы.

Оператор клонирования генерирует новое поколение копий антител в будущей популяции.

Оператор мутации действует в зависимости от имеющейся популяции клонов, применяя к каждому антителу определенное количество одиночных мутаций, осуществляемых случайным образом. Можно выделить следующие способы мутации [5]:

1. Статическая мутация. Количество мутаций зависит от минимизируемой функции f , поэтому антитело в каждый момент времени будет подвергаться определенному числу мутаций.

2. Пропорциональная мутация. Количество мутаций антитела пропорционально соответствующему значению.

3. Обратная пропорциональная мутация. Количество мутаций антитела обратно пропорционально соответствующему значению.

4. Круглая мутация. Каждое антитело подвергается мутации круглого сочетания.

Оператор старения устраняет старые особи. Статический оператор старения использует возрастной параметр для максимального количества поколений антител, которым разрешено оставаться в популяции. Когда антитело старше, оно удаляется из системы, даже если оно может оказаться вполне пригодным на последующих итерациях.

При колониальной экспансии клонированное антитело наследует возраст его родителя. После этапа мутации только те антитела, которые получили высшее значение аффинной, получают возраст, равный 0. Элитный вариант этого оператора получается путем взятия наилучших антител популяции в поколение с возрастом, равным 0.

Элитный вариант этого оператора получается взятием лучшего антитела из популяции в поколение.

Оператор селекции заменяет наихудшие антитела в популяции новыми случайными антителами.

Заключение. В работе рассмотрены обобщенные условия сходимости ИА в зависимости от используемых иммунных операторов. Анализ сходимости ИА основан на выполнении двух условий:

1) в результате мутации можно достичь оптимальное состояние из неоптимального за один шаг;

2) как только оптимальное состояние будет найдено, оно сохранится в популяции и не будет утеряно. Показано, что только операторы мутации и селекции, которые могут вносить изменения в антитела, способствует поиску оптимума.

Данный алгоритм иммунокомпьютинга может быть рассмотрен как «иммунный» алгоритм, так как любой образ может быть представлен как частный случай формального протеина и его распознавание основывается на энергии связи с антителом формального протеина.

References:

- [1]. Dasgupta D. Искусственные иммунные системы и их применение.- Fizmatlit.- 2006.-344 s.
- [2]. Wang X. Hybrid nature-inspired computation method for optimization / Doctoral Dissertation. Helsinki University of Technology, TKK Dissertations, Espoo 2009, 161 p.

ENERGETICS, THE ELECTRICAL ENGINEERING, ELECTRONIC DEVICES AND INFORMATION TECHNOLOGIES

- [3]. Eiben A. E., Michalewicz Z., Schoenauer M., Smith J. E. Parameter Control in Evolutionary Algorithms / Parameter Setting in Evolutionary Algorithms.- Springer Verlag, 2007, pp. 19-46.
- [4]. Zadeh L.A. What is Soft Computing? Soft Computing 1 (1997).
- [5]. Kondrashina Ye.Yu., Litvintseva L.V., Pospelov D.A. Predstavlenie znaniy o vremeni i prostranstve v intellektualnix sistemax. Pod red. D.A. Pospelova. Seriya «Problemy iskusstvennogo intellekta», -Vip. 6. – M.: Nauka. Fizmatgiz, 1989. -328 S.
- [6]. Muxamediyeva D.T. Model of estimation of success of geological exploration perspective // International Journal of Mechanical and production engineering research and development (IJMPERD) ISSN(P): 2249-6890; ISSN(E): 2249-8001 Vol. 8, Issue 2, USA. 2018, 527-538 pp. Impact Factor (JCC): 6.8765. Muxamediyeva D.T. Structure of fuzzy control module with neural network //International Journal of Mechanical and Production Engineering Research and Development (IJMPERD) ISSN (P): 2249-6890; ISSN (E): 2249-8001 Vol. 9, Issue 2, Apr 2019, pp.649-658.

Литература

- [1]. Дасгупта Д. Искусственные иммунные системы и их применение.- Физматлит.- 2006.-344 с.
- [2]. Wang X. Hybrid nature-inspired computation method for optimization / Doctoral Dissertation. Helsinki University of Technology, TKK Dissertations, Espoo 2009, 161 p.
- [3]. Eiben A. E., Michalewicz Z., Schoenauer M., Smith J. E. Parameter Control in Evolutionary Algorithms / Parameter Setting in Evolutionary Algorithms.- Springer Verlag, 2007, pp. 19-46.
- [4]. Zadeh L.A. What is Soft Computing? Soft Computing 1 (1997).
- [5]. Кондрашина Е.Ю., Литвинцева Л.В., Пospelov Д.А. Представление знаний о времени и пространстве в интеллектуальных системах. Под ред. Д.А. Пospelova. Серия «Проблемы искусственного интеллекта», -Вып. 6. – М.: Наука. Физматгиз, 1989. -328 С.
- [6]. Muxamediyeva D.T. Model of estimation of success of geological exploration perspective // International Journal of Mechanical and production engineering research and development (IJMPERD) ISSN(P): 2249-6890; ISSN(E): 2249-8001 Vol. 8, Issue 2, USA. 2018, 527-538 pp. Impact Factor (JCC): 6.8765. Muxamediyeva D.T. Structure of fuzzy control module with neural network //International Journal of Mechanical and Production Engineering Research and Development (IJMPERD) ISSN (P): 2249-6890; ISSN (E): 2249-8001 Vol. 9, Issue 2, Apr 2019, pp.649-658.