

УДК 004.9

## ПРОГНОЗИРОВАНИЕ СОСТОЯНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА С ПРИМЕНЕНИЕМ НЕЧЕТКИХ КОГНИТИВНЫХ КАРТ И АЛГОРИТМОВ НЕЛИНЕЙНОГО ОБУЧЕНИЯ



**Ю.С. Шевнина**  
НИУ МИЭТ, профессор  
института системной и  
программной инженерии и  
информационных  
технологий, д.т.н., доцент,  
yusm@rambler.ru



**А.В. Прокофьев**  
НИУ МИЭТ, аспирант института  
системной и программной  
инженерии и информационных  
технологий,  
alexei.prokofiev2015@yandex.ru



**О.А. Ашарафьянова**  
НИУ МИЭТ, студентка 3-го  
курса института системной  
и программной инженерии и  
информационных технологий,  
ashaolga@mail.ru

### **Ю.С. Шевнина**

*Известный ученый в области современных методов анализа и обработки данных, применения методов предиктивной аналитики и искусственного интеллекта в различных сферах.*

### **А.В. Прокофьев**

*Занимается проектирование и реализацией предиктивных моделей для различных предметных областей.*

### **О.А. Ашарафьянова**

*Студента направления «Прикладная информатика».*

**Аннотация.** В статье рассматривается подход к прогнозированию состояния образовательного процесса с применением мультиреляционных нечетких когнитивных карт и алгоритмов нелинейного обучения Хебба. Описана методика формирования наборов данных для определения текущего состояния образовательного процесса и динамики его изменения на основе цифрового следа обучающихся, включающего академические, системные и социально-психологические показатели. Представлены результаты апробации подхода, подтверждающие высокую точность прогнозирования рисков академической неуспеваемости.

**Ключевые слова:** образовательный процесс, нечеткие когнитивные карты, алгоритм Хебба, прогнозирование, цифровой след, *Advanced Analytics*, *Big Data*.

**Введение.** В настоящее время большое внимание уделяется вопросам прогнозирования в различных предметных областях с целью оптимального планирования и управления, предупреждения внештатных ситуаций и снижения влияния других негативных факторов. Известно, что многие социальные и информационные системы, в частности образовательный процесс (ОП), являются многомасштабными сложными нелинейными системами с множественной гранулярностью (МСНС). Подобные системы характеризуются взаимодействием множества компонентов, имеющих разные масштабы и уровни организации. В таких системах малые изменения в одном компоненте приводят к значительным изменениям в поведении всей системы из-за нелинейных взаимодействий. Следовательно, моделирование и прогнозирование изменения состояния ОП является своевременной и актуальной задачей. Точный прогноз состояния ОП позволит уменьшить неопределенность внешней среды с учетом внутренних характеристик системы, предотвратить академические угрозы и

неуспеваемость, повысить эффективность управления ресурсами и способствовать устойчивому развитию образовательной системы в целом [1].

**Аналитический обзор существующих методов и средств оценки состояния образовательного процесса.** Современная практика оценки состояния образовательного процесса интегрирует следующие методы и средства: индексы академической успешности, многоуровневый анализ данных, использование логов цифрового следа и систем дистанционного обучения. Определены наиболее эффективные методы и средства оценки состояния ОП, выявлены их достоинства и недостатки. Тем не менее, указанные методы и средства не позволяют в полной мере определять состояние образовательного процесса и прогнозировать его изменение с учетом зависимостей между различными факторами.

Как уже отмечалось выше, ОП является сложной нелинейной системой с множественной гранулярностью. Среди современных подходов к прогнозированию изменения состояния нелинейных систем следует выделить аналитические методы, основанные на математическом анализе. Однако они ограничены в использовании для сложных и высокомерных систем.

Наиболее перспективным к прогнозированию изменения состояния сложных нелинейных систем является подход с использованием когнитивных карт, который позволяет выявлять, визуализировать и анализировать взаимосвязи между компонентами системы, в том числе нелинейными. Когнитивная карта представляет собой граф, в котором узлы являются состояниями компонентов системы, а дуги – связями между ними. Нечеткая когнитивная карта (НКК) сочетает в себе аспекты нечеткой логики, нейронных сетей и других методов и подходит для моделирования систем на основе ресурсов данных.

Для прогнозирования изменения состояния ОП предлагается использовать мультиреляционные НКК, в частности, двухуровневые. При этом на низком уровне для каждого массива создается одна НКК, а на высоком уровне – общая НКК для всех массивов данных о состоянии компонентов и системы.

**Математическая модель прогнозирования состояния образовательного процесса.** Рассмотрим примеры состояний образовательного процесса, которые возможно описать с помощью нечетких когнитивных карт. Состояния, характеризующие уровень академической успеваемости с учетом влияния посещаемости и выполнения заданий. Состояния, описывающие качество цифрового следа в ОП с учетом взаимодействия с учебными материалами, которые сложно моделировать строго детерминированными методами. Состояния, связанные с рисками академической задолженности и отчисления.

Нечеткая когнитивная карта  $F$  математически представляет собой кортеж из элементов  $F = \langle C, W, A, f \rangle$ , где  $C = \{C_1, C_2, \dots, C_N\}$  – набор из  $N$  состояний компонентов, образующих узлы графа;  $W : (C_i, C_j) \rightarrow w_{ij}$  – матрица связи между состояниями компонентов образовательного процесса,  $w_{ij}$  – значение веса ребра, направленного от  $C_i$  к  $C_j$ ,  $w_{ij} \in [-1, 1]$ ;  $A : C_i \rightarrow A_i(t)$  – матрица связи между состоянием компонента  $C_i$  с последовательностью активирующих его факторов с учетом степени активации.

Очевидно, если существует положительная причинно-следственная связь между состояниями компонентов  $C_i$  и  $C_j$ , то  $w_{ij} > 0$ , что означает увеличение вероятности возникновения состояния  $C_j$  при увеличении вероятности возникновения состояния  $C_i$ , а уменьшение вероятности возникновения  $C_i$  приведет к уменьшению вероятности возникновения  $C_j$ . При отсутствии связи между состояниями компонентов  $w_{ij} = 0$  (рисунок 1).

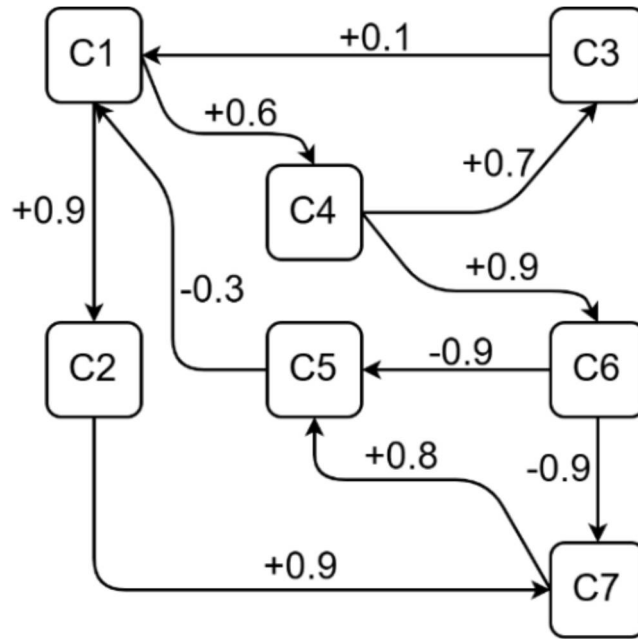


Рисунок 1. Пример нечеткой когнитивной карты

Для моделирования ОП рассмотрим активацию состояния одного из компонентов системы в момент времени  $t + 1$ :

$$A_j(t+1) = f \left( A_j(t) + \sum_{i \neq j} A_i(t) w_{ij} \right) \quad (1)$$

Данная функция позволяет определить влияние изменения состояния одного компонента на другие компоненты системы. При этом функция активации состояния компонента системы учитывает предшествующие состояния, что особенно важно для ОП, поскольку следующее состояние компонента зависит не только от внешних значений, но и предшествующих состояний.

Для контроля входных данных и обеспечения стабильности модели ОП на основе НКК необходимо ограничить сумму значений переменных, умноженных на соответствующие веса связей между ними. Для преобразования целесообразно использовать логистическую функцию:

$$f(x) = \frac{1}{1 + e^{-Cx}} \quad (2)$$

Предложенное описание состояния ОП с использованием нечетких когнитивных карт позволяет анализировать сложные взаимосвязи и влияния компонентов друг на друга.

В реальном мире сложные нелинейные системы имеют многопространственные группы компонентов с прямыми или косвенными связями, которые создают мультиреляционные данные. Более того, каждое такое пространство может содержать множество компонентов с различными отношениями уровня детализации. В частности, в экосистеме ОП существуют три пространства: академическая успеваемость, цифровой след и социально-психологический профиль. Они являются компонентами с высокой гранулярностью; а баллы за тесты, время сессии и активность в системе – компонентами с низкой гранулярностью.

Для прогнозирования изменений состояний компонентов с множественными зависимостями и высокой гранулярностью в сложной системе ОП применяется

мультиреляционная НКК с двумя уровнями и  $n$ -массивами  $U_n^2 = \langle C_n^2, W_n^2, A_n^2, f \rangle$ , где  $C_n^2$  – набор состояний компонентов с высокой гранулярностью. Компонент с высокой гранулярностью можно рассматривать как суммарное выражение его компонентов низкой гранулярности с множеством экземпляров на низком уровне измерения.

**Алгоритм обработки и анализа образовательных данных.** В основе разрабатываемого алгоритма обработки и анализа данных образовательного процесса лежит алгоритм построения НКК. Алгоритм построения НКК должен адекватно определять влияние одного компонента ОП на другой (например, влияние регулярности работы в системе дистанционного обучения на итоговый балл) и рассчитывать веса связей между состояниями компонентами. В настоящее время наиболее популярными для построения когнитивных карт являются алгоритмы на основе принципа Хебба и эволюционные алгоритмы.

Принцип Хебба гласит, что узлы, которые активируются вместе, соединяются вместе. Для НКК данный принцип определяет правила обновления весов связей между состояниями образовательных компонентов на основе их совместной активности (корреляции данных в цифровом следе). Как правило, для пересчета весов и точного моделирования сложных зависимостей используются нелинейные функции (нелинейное обучение Хебба).

#### *Нелинейное обучение Хебба*

Нелинейное обучение Хебба относится к методам обучения нейронной сети без учителя. Для нелинейного выходного блока определим случайный пресинаптический входной вектор  $x$

(входные метрики обучающегося), весовой вектор  $w$  и выходной вектор  $z = f(w^T x)$ .

Нелинейная функция активации  $f$  является сигмовидной функцией. Функция критерия  $J$  максимизируется по принципу Хебба и может быть записана в виде:

$$J = E \left[ z^2 \right]. \quad (3)$$

Дополнительное ограничение  $\|w\|=1$  позволяет стабилизировать правила обучения модели НКК. Стохастическое аппроксимирующее решение используется для следующего нелинейного правила обучения Хебба в виде:

$$\Delta w_{ji} = \eta_k z \frac{dz}{dy} (x_j - w_{ji} y_i) \quad (4)$$

Таким образом, представленные правила нелинейного обучения позволяют определять набор весовых параметров при условии наибольшей дисперсии выходных данных нелинейного модуля.

#### *Непрерывный генетический алгоритм.*

Непрерывный генетический алгоритм (НГА) применяется для разработки матрицы связей между состояниями компонентов ОП на основе ресурса данных (логов из LMS). В терминах генетических алгоритмов НГА определяет каждую хромосому как вектор с плавающей точкой. В случае НКК с  $N$  узлами каждая хромосома состоит из  $N(N-1)$  генов, которые представляют собой числа с плавающей точкой диапазона  $[-1,1]$ , определенные следующим образом:

$$E = [w_{12}, w_{13}, \dots, w_{1N}, w_{21}, w_{23}, \dots, w_{2N}, \dots, w_{NN-1}]^T, \quad (5)$$

где  $w_{ij}$ - значение веса ребра НКК от  $i$ -го до  $j$ -го состояния компонента.

Функция приспособленности определяется для каждой хромосомы путем вычисления разницы между откликом системы, сгенерированным с использованием весов НКК, и соответствующим откликом системы, полученным непосредственно из базы реальных данных обучающихся. Мера ошибки рассчитывается по формуле:

$$Error\_Lp = \alpha \sum_{t=1}^{M-1} \sum_{n=1}^N \left| A_n(t) - \hat{A}_n(t) \right|^p \quad (6)$$

где  $\alpha$  применяется для нормализации частоты ошибок;  $N$ - количество состояний компонентов в НКК;  $M$ - количество итераций (учебных периодов). Мера ошибки используется как ядро функции приспособленности в виде:

$$F = \frac{1}{a \cdot Error\_Lp + 1} \quad (7)$$

Обучение по НГА завершается при достижении порогового значения функции приспособленности  $max\_fitness$ . Если ни одно из условий остановки НГА не выполняется с увеличением итераций, то используются эволюционные операторы (скрещивание или мутация) и стратегия выбора для создания нового поколения решений. Точность модели определяется как нормализованная средняя ошибка между соответствующими значениями состояний компонентов на каждой итерации:

$$error = \frac{1}{M \cdot T} \sum_{t=0}^T \sum_{j=1}^M \left| A_j(t) - \hat{A}_j \right| \quad (8)$$

Следовательно, весовой вектор в низкоуровневой НКК рассчитывается дополненным многовариантным нелинейным обучением Хебба. Для получения точной классификации и выявления причинно-следственных связей в образовательном процессе применялся непрерывный генетический алгоритм.

**Результаты апробации математической модели на основе реальных данных.** На основе предложенной модели мультиреляционной НКК проводилось моделирование и прогнозирование состояния образовательной экосистемы. Очевидно, что рассматриваемая система имеет сложную структуру с мультиреляционным набором данных. Состояние экосистемы выражается как множественная взаимосвязь массивов компонентов с низкой гранулярностью. Выделим три набора исходных обезличенных данных о состоянии обучающихся.

Набор  $BK_0$ : Базовый профиль. Каждый вектор включает характеристики направления подготовки, формы обучения и результаты входного контроля (ЕГЭ).

Набор  $BK_1$ : Цифровой след. Каждый вектор включает метрики активности в LMS (частота входов, время выполнения тестов, пропуски), помимо тех характеристик, что определены в  $BK_0$ .

Набор  $BK_2$ : Социально-психологический профиль. Включает характеристики социальной активности (участие в форумах, проектной деятельности), помимо тех, что определены в  $BK_1$ .

Эксперимент реализован для прогнозирования изменения состояния ОП с применением мультиреляционной НКК и разработанного алгоритма ее построения при различных пороговых значениях функции приспособленности. В качестве обучающей выборки для

гибридной сети использовано 7000 комплектов обезличенных данных обучающихся, а в качестве тестовой выборки – 3000 комплектов. Время построения НКК показано при пороговых значениях функции приспособленности, принимающей значение от 0,6 до 0,9 (рисунок 2).

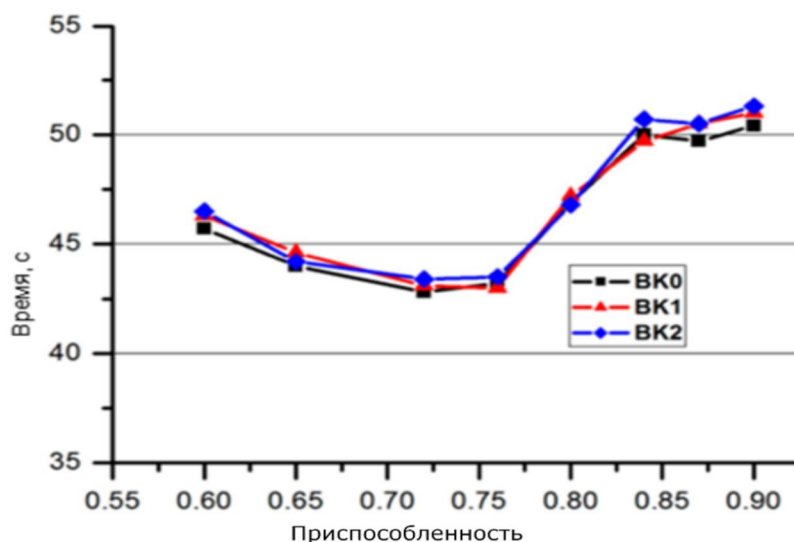


Рисунок 2. Зависимость времени построения НКК от значения функции приспособленности

Как видно из графика, время построения НКК для различных наборов исходных данных практически не отличается, при этом наименьшее затрачиваемое вычислительное время отмечается для приспособленности в интервале [0,65;0,75][0,65;0,75].

Рассмотрим результаты прогноза состояний экосистемы при  $max\_fitness=0,7$  для трех видов исходных данных (таблица 1).

Таблица 1. Эффективность прогноза состояния образовательной экосистемы

Набор исходных данных	Время выполнения (с)	Точность
$BK_0$	0,78	92,3%
$BK_1$	0,80	97,9%
$BK_2$	0,80	97,7%

Время прогноза изменения состояний экосистемы с набором данных, содержащем большее количество связей, больше. При этом наилучший показатель точности прогноза зафиксирован для  $BK_1$ .

Точность прогноза подтверждается и рассчитывается как степень соответствия фактических данных результатам, предсказанным прогнозной моделью по формуле:

$$Точность = \left( 1 - \frac{MAE}{\bar{y}} \right) \cdot 100\% \quad (9)$$

Точность прогноза 97,9% означает, что средняя ошибка прогноза составляет 2,1% от среднего значения фактических данных. Такая точность свидетельствует о высокой степени соответствия модели реальным условиям образовательного процесса. Для каждого обучающегося рассчитывается прогнозное значение состояния (итоговый балл или риск отчисления).

Затем сравниваются прогнозные и фактические значения. Рассчитывается средняя абсолютная ошибка (MAE):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (10)$$

где  $y_i$  - фактическое значение состояния,  $\hat{y}_i$  - прогнозное значение состояния,  $n$  – количество обучающихся.

**Оценка эффективности полученных результатов.** Предложенная система оценки состояния образовательного процесса показывает значительное повышение эффективности и точности оценки по сравнению с традиционными методами академического мониторинга.

В отличие от традиционных методов, требующих длительных замеров успеваемости вручную (по итогам сессий), предиктивная система обеспечивает автоматизированный сбор, обработку и анализ данных в реальном времени или с минимальной задержкой.

Благодаря интеграции с базами данных LMS сокращается время выявления обучающихся в «зоне риска» с нескольких недель до нескольких часов, что повышает оперативность принятия педагогических решений. Традиционные методы выявляют локальные нарушения (академические задолженности) постфактум.

Разработанное решение выявляет скрытые паттерны (снижение мотивации, прокрастинацию) с точностью до 97,9%, что позволяет своевременно принимать меры по корректровке индивидуальных образовательных траекторий.

**Заключение.** В результате выполнения работы разработана математическая модель прогнозирования состояния образовательного процесса, позволяющая учитывать нелинейные зависимости между академическими, цифровыми и социально-психологическими факторами.

Получены результаты апробации разработанной модели на основе реальных обезличенных данных Big Data.

Разработан алгоритм обработки и анализа данных, интегрирующий многовариантное нелинейное обучение Хебба и непрерывный генетический алгоритм.

Осуществлена оценка эффективности: разработанный подход обеспечивает существенное повышение эффективности академического мониторинга за счет ускорения процессов, увеличения точности предиктивной аналитики и снижения затрат на администрирование образовательного процесса.

#### Список литературы

- [1] Аверкин А.Н. Генетический алгоритм обучения нечетких когнитивных карт / А.Н. Аверкин, А.А. Паринов // Научные труды ВЭО России. – 2010. - № 143. – С. 69-74.
- [2] Петухова А.В. Методы прогнозирования развития сложных систем с применением теории нечетких когнитивных карт / А.В. Петухова, А.В. Коваленко // Computational Mathematics and Information Technologies. – 2022. – № 6(2). – С. 81-95.
- [3] Shevnina Ju.S. On the Issue of Modeling Complex Nonlinear Systems / Ju.S. Shevnina, L.G. Gagarina, A.V. Chirkow // AIP Conference Proceedings. – 2021. – Vol. 2467. – P. 565.

#### Авторский вклад

**Шевнина Юлия Сергеевна** – постановка задачи исследования, разработка рассматриваемых подходов и алгоритмов прогнозирования состояния образовательного процесса.

**Прокофьев Алексей Валентинович** – разработка методики оценки эффективности полученных результатов, проведение эксперимента и анализ результатов.

**Ашарафьянова Ольга Андреевна.** - сбор и предобработка данных, реализация алгоритмов обучения Хебба и генетического алгоритма, проведение экспериментов и анализ результатов.

## **PREDICTING THE STATE OF THE EDUCATIONAL PROCESS USING FUZZY COGNITIVE MAPS AND NONLINEAR LEARNING ALGORITHMS**

***Yu. S. Shevnina***

*National Research University  
of Electronic Technology  
(MIET), Professor, Institute of  
Systems and Software  
Engineering and Information  
Technology, Doctor of  
Engineering, Associate  
Professor, yusm@rambler.ru*

***A. V. Prokofiev***

*National Research University of  
Electronic Technology, graduate  
student at the Institute of Systems and  
Software Engineering and Information  
Technology,  
alexei.prokofiev2015@yandex.ru*

***O. A. Asharafyanova***

*National Research  
University of Electronic  
Technology (MIET), third-  
year student at the Institute  
of Systems and Software  
Engineering and  
Information Technology,  
ashaolga@mail.ru*

**Abstract.** This article examines an approach to predicting the state of the educational process using multirelational fuzzy cognitive maps and Hebbian nonlinear learning algorithms. It describes a methodology for generating datasets to determine the current state of the educational process and its dynamics based on students' digital footprints, which include academic, systemic, and socio-psychological indicators. The results of testing the approach are presented, confirming the high accuracy of predicting the risk of academic failure.

**Keywords:** educational process, fuzzy cognitive maps, Hebbian algorithm, forecasting, digital footprint, Advanced Analytics, Big Data.