

Настоящая разработка подготовлена для конференции во ВНИГИМе, проведение которой планируется на декабрь 2023 г.

Препринт

УДК 631.6.02

ПРОЦЕДУРА ПРЕОБРАЗОВАНИЯ БАЗ ДАННЫХ В БАЗЫ ЗНАНИЙ

Шабанов В.В. д.т.н., проф., научный руководитель «Проблемной лаборатории по разработке теоретических основ управления водным, солевым и тепловым режимами мелиорируемых земель». 515vvsh@gmail.com. Российский государственный аграрный университет (Московская сельскохозяйственная академия) им. К.А. Тимирязева, Москва, Россия

Аннотация. В статье рассматриваются преобразование баз данных в базы знаний. Обсуждаются процедуры преобразования. Приводится пример, на основе преобразования базы данных влагозапасов в почве, в базу знаний о продуктивности сельскохозяйственной культуры. Отмечается возможность выполнения таких преобразований с использованием аппарата нейросетей.

Ключевые слова. База данных, база знаний, преобразование баз данных в базы знаний, распознавание образов, классификация природно-антропогенных объектов, экспертные системы, процесс принятия решений, мелиорация.

PROCEDURE FOR CONVERTING DATABASES INTO KNOWLEDGE BASES TO CREATE DECISION-MAKING TOOLS Shabanov V.V. Doctor of Technical Sciences, Professor, Scientific Director of the Problem Laboratory for the Development of Theoretical Foundations for the Management of Water, Salt and Thermal Regimes of Reclaimed Lands. 515vvsh@gmail.com. Russian State Agrarian University (Moscow Agricultural Academy) named after K.A. Timiryazev, Moscow, Russia

Abstract. This article discusses how to convert databases into knowledge bases. Conversion procedures are discussed. An example is given based on the transformation of a database on soil moisture reserves into a knowledge base on crop productivity. The possibility of performing such transformations using the apparatus of neural networks is noted. Keywords. Database, knowledge base, transformation of data-

bases into knowledge bases, pattern recognition, classification of natural and anthropogenic objects, expert systems, decision-making process, land reclamation.

Введение. База данных и база знаний (отличия и единство). Считается, что база данных (БД) — это структурированное хранилище данных, в котором информация организована таким образом, чтобы она была легко доступна для решения различных задач.

Для задач обоснования необходимости и эффективности мелиорации, Базы Данных представляются, в виде матриц значений факторов внешней среды, которыми предстоит управлять, в частности, продуктивными влагозапасами в почве под различными сельскохозяйственными культурами [1, 2, 3]. Это позволяет хранить большие объёмы информации, обрабатывать её и извлекать нужную.

Процесс формирования баз данных может быть отнесён к процессам «цифровизации».

Базы Данных без предварительной обработки не могут качественно (хорошо - плохо; много - мало) характеризовать объект и результат мелиоративного управления, для этого нужны Базы Знаний.

Создание баз знаний можно отнести к процессу интеллектуализация обработки информации (ИОИ).

ИОИ — процесс использования интеллектуальных методов и алгоритмов для анализа, интерпретации и обработки данных. Он может включать в себя использование машинного обучения, искусственного интеллекта, статистических методов и других технологий для улучшения качества и эффективности обработки информации. Целью интеллектуализации обработки информации, является повышение точности прогнозирования, оптимизация процессов принятия решений и улучшение результатов работы в различных областях, таких как бизнес, наука, медицина и т.д.

Базы знаний – определение. База знаний в матричном виде, даёт оценочную информацию (больше, меньше, хорошо, плохо).

Базу знаний связывает с базой данных преобразующий слой, по структуре, аналогичный таковому в нейросетях. Именно в нём происходит преобразование каждой ячейки матрицы исходных данных в некоторую новую информационную субстанцию.

База знаний может быть представлена как некоторый графический образ, поэтому для таких объектов может быть использована процедура "распознавания образов", с применением, и здесь, таких логических понятий как много-мало, хорошо - плохо.

Методология. Ниже приводится один из возможных вариантов преобразования элементов базы данных влагозапасов БД(φ) в базу знаний относительной продуктивности БЗ(S) [2].

Для i -го момента времени жизни биологического объекта и для j -го фактора при $\varphi_{\min} = 0$ модель требований живого организма можно записать в виде:

$$S_{ij} = \left(\frac{\varphi}{\varphi_{opt}} \right)^{\gamma \varphi_{opt}} \left(\frac{\varphi_{max} - \varphi}{\varphi_{max} - \varphi_{opt}} \right)^{\gamma (\varphi_{max} - \varphi_{opt})} \quad (1)$$

где, S- относительная продуктивность; φ - текущее значение фактора; φ_{max} - максимальное значение фактора; φ_{opt} - оптимальное значение фактора; γ - коэффициент саморегулирования растения [2].

База данных влагозапасов образует матрицу:

$$\begin{matrix} \varphi_{11} & \varphi \dots & \varphi_{1i} \\ \dots & \dots & \dots \\ \varphi_{1k} & \varphi \dots & \varphi_{ik} \end{matrix} \quad (2)$$

где, значения k – элементы ландшафтной катены от водораздела k=1, до водоприёмника, а значения i – декады вегетации.

Преобразуя каждый элемент «матрицы данных» (2) по формуле (1), получаем, матрицу знаний (3) о продуктивности биологического объекта в пространстве и во времени [3].

$$\begin{matrix} S_{11} & S \dots & S_{1i} \\ \dots & \dots & \dots \\ S_{1k} & S \dots & S_{ik} \end{matrix} \quad (3)$$

Проведя процедуру классификации (градации) продуктивности в пространстве «матрицы знаний», можно выделить определённые классы относительной продуктивности: (1 - 0.8;) – оптимальные условия; (0.8 - 0.6) – хорошие условия; (0,6 – 0,4) – средние условия; (0.4 - 0.2 - ниже средних); (0.4 - 0.2 – плохие условия) Это позволяет «принять решение» о необходимых действиях по оптимизации эффективности управления.

Описанное выше, относится к «однофакторному» управления жизнедеятельностью биологического объекта. Вместе с тем продуктивность биологических объектов зависит от множества факторов. Основные из них: водный φ_w ; тепловой φ_t ; пищевой φ_f ; радиационный φ_r .

В этой ситуации, построение непрерывных правил управления, представляется весьма сложным. Поэтому были определены, в соответствии с «законом минимума фактора», правила основанные на аппарате алгебры логики.

Эти правила можно сформулировать так: в каждый момент времени роста и развития биологического объекта, критическим, определяющим продуктивность, является фактор, находящийся в минимуме.

$$\min \{ \varphi_w; \varphi_t; \varphi_f; \varphi_r \} \quad (4)$$

В этой ситуации матрица данных (БД) становится трёхмерной. В «вертикальном» направлении располагаются основные слои (факторы жизни биологического объекта).

$$\begin{array}{ccc} \varphi_{11j} & \varphi \dots & \varphi_{1ij} \\ \dots & \dots & \dots \\ \varphi_{1kj} & \varphi \dots & \varphi_{ikj} \end{array} \quad (5)$$

При расчёте, выбор слоя осуществляется в соответствии с правилом:

$$\min \{ \varphi_w \wedge \varphi_t \wedge \varphi_f \wedge \varphi_r \} \quad (6)$$

где, $\varphi_w \wedge \varphi_t \wedge \varphi_f \wedge \varphi_r$ – факторы внешней среды; соответственно: водный, тепловой, пищевой, радиационный; \wedge - оператор логического умножения (или).

На «территории» катены могут располагаться несколько биологических объектов с разными требованиями к условиям внешней среды. Например, сельскохозяйственная культура и почвенное биологическое сообщество. В общем случае требования (мелиоративные режимы) этих двух объектов могут быть раз-

личны, т.е. функция (1) остаётся неизменной, но меняются её параметры. В таком случае получаются две базы знаний из одной исходной базы данных.

Преобразуя полученные базы знаний, по определённым алгоритмам, можно получать новые базы. Так, например, переводя относительные продуктивности в абсолютные величины и зная долю содержания углерода в биомассе¹ данного биологического объекта, можно получить базу данных распределения депонированного углерода, в пространстве и во времени.

Таким образом, при определённых условиях, базы, производные от баз знаний, могут становиться основанием для новой цифровизации – новыми базами данных.

При развитии этого направления целесообразно использовать аппарат нейросетей. Это представляется возможным, так как структура преобразования БД в БЗ, аналогична нейросетевой. Существует входной слой, в нашем случае это вектор – столбец БД – влагозапасы почвы на каждом элементе катены в i -ую декаду. Существует «внутренний слой» - функция, переводящая влагозапасы в продуктивность, уравнение (1). Выходной слой – относительная продуктивность на каждом из 6-7 элементов катены (типов водного питания по Брудастову), в любую из 10 декад вегетации.

Можно используя входной слой (матрица 6×10) и выходной (матрица 6×10), попытаться определить структуру скрытого слоя (количество «воспринимающих элементов» и связи между ними).

Входные матрицы влагозапасов под определённой культурой, средние (за 30 лет наблюдений).

При необходимости увеличения объёма исходных данных, можно использовать, установленную в Проблемной лаборатории, закономерность - связь влагозапасов по катене: величины влагозапасов в почве вдоль ландшафтной катены, связаны между собой линейно.

Это позволяет использовать линейную интерполяцию (между столбцами) и увеличить количество элементов входной матрицы в n -раз.

¹0.4-0.6 от биомассы

В результате решения этой задачи можно будет проанализировать изменение структуры скрытого слоя при увеличении объёма исходных данных. Это позволит найти необходимый объём исходных данных.

Графическая интерпретация образа выходной матрицы (рисунка) и классификация её к той или иной «степени необходимости мелиорации» даёт возможность, на этой основе, сделать инструменты принятия решения по применению мелиорации к каждой части катены, с оценкой эффективности принятого решения (экономической, экологической, климатической).

К сожалению, не хватает законодательных и нормативных материалов по оценке комплексной эффективности – стоимость плодородия, стоимость поглощаемого углерода, стоимость «поставленного» при фотосинтезе, с этих земель, кислорода, стоимость «поставленной» дистиллированной воды возникающей в процессе транспирации растений.

Таким образом, в связи с усложнением задачи, возможно, будут эффективны и нейросетевые подходы. Вместе с тем, обычно количество функций «преобразующих» входные переменные (внутренний слой) существенно ограничены, и это может явиться причиной ограниченного применения нейросетевого подхода, в современном виде. В связи с этим, возможно, что появится необходимость введение новых преобразующих функций.

Заключение – представляется, что рассмотренные задачи актуальны для оценки эффективности взаимодействия человека и природы, и поэтому могут быть востребованы в практике.

Примечание: текст, выделенный курсивом, сформирован на основании «бесед» с системой **GigaChat**, создателям которой, автор весьма признателен.

Литература

1. Голованов А.И. Природообустройство: учебник / А. И. Голованов, Ф. М. Зимин, Д. В. Козлов, И. В. Корнеев и др.; под редакцией Голованова А.И. — 2-е изд., испр. и доп. — Санкт-Петербург: Лань, 2015. — 560 с. — ISBN 978-5-8114-1807-7. — Текст: электронный //Лань: электронно-библиотечная

- Разработка - ПРОЦЕДУРА ПРЕОБРАЗОВАНИЯ БАЗ ДАННЫХ В БАЗЫ ЗНАНИЙ (2023) система. — URL: <https://e.lanbook.com/book/64328> (дата обращения: 19.11.2023). — Режим доступа: для авториз. пользователей.
2. Шабанов В.В. Водообеспеченность яровой пшеницы и её расчёт. Л., Гидрометеиздат. 1982.
 3. Шабанов В. В. Солошенко А. Д. Взаимосвязь типов увлажнения ландшафтной катены // Проблемы управления водными и земельными ресурсами: материалы междунар. науч. форума. –М.: Издательство РГАУ-МСХА, 2015. – С. 104–110.
 4. Шабанов В.В. Использование базы данных "Декоративные растения" в системе инженерной поддержки ландшафтного дизайна // Материалы международной научно-практической конференции "Роль мелиорации в обеспечении продовольственной и экологической безопасности России". — 2009. — Ч. 1. — 366-369.

References

1. Golovanov A.I., Zimin F.M., Kozlov D.V., Korneev I.V. et al.; edited by A.I. Golovanov — 2nd ed., ispr. St. Petersburg: Lan, 2015. — 560 p. — ISBN 978-5-8114-1807-7. — Text: electronic //Lan: electronic library system. — URL: <https://e.lanbook.com/book/64328> (accessed: 19.11.2023). — Access mode: for authorizations. Users.
2. Shabanov V.V. Bioclimatic justification of hydrothermal reclamation. Leningrad, Gidrometeoizdat. 1973.
3. Shabanov V. V., Soloshenkov A. D. Vzaimsvyaz tipov uvhuzheniya landscape noy katena [Interconnection of types of moistening of landscape catena]. Scientific. Forum. Moscow: RSAU-MSHA Publ., 2015. P. 104–110.
4. Shabanov V.V. Ispol'zovanie databasea "ornamental'nye rastenii" v sisteme inzhenernoy podderzhka landscape design [Use of the database "decorative plants" in the system of engineering support for landscape design] // Proceedings of the International Scientific and Practical Conference "The Role of Land Reclamation in Ensuring Food and Environmental Safety of Russia". ...