

От генома до оптимального сбалансированного рациона

Шлихт А.* , Краморенко Н.

Дальневосточный федеральный университет, Владивосток, Россия

* *schliht@mail.ru*

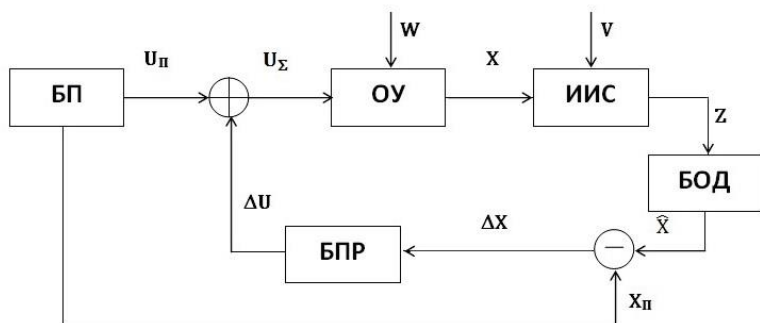
Ключевые слова: геномика; феномика; фудомика; оптимальный синтез рационов; базы данных; базы знаний; экспертные системы

Мотивация и цель: Целью исследования является создание интегрированной интеллектуальной биоинформационной системы, позволяющей проследить цепочку омиксных биохимических процессов, включая: геномику, эпигеномику, транскриптомику, экзомуку, протеомику, метаболомику, реактомику, геномонтологию, геномные мутации и далее перейти к диагностике генетических заболеваний, анализу и синтезу оптимальных сбалансированных рационов. Существующие хранилища омиксных данных на базе международных порталов (NCBI, KEGG, Ensembl, UniProt, GO и др.) не обеспечивают в должной мере автоматический анализ и интерпретацию омиксных данных. Решающее значение здесь приобретают технологии искусственного интеллекта (ИИ) на основе баз данных (БД) и баз знаний (БЗ). Омик-центрированные биологические объекты и процессы стали вызовом для математики, информатики, теории управления. Учет геномного статуса индивида является важной особенностью современной теории и практики наук о жизни. Мутации в генах дают мутации в соответствующих белках. Если этот белок фермент, то происходит нарушение биохимических реакций, в которых участвует фермент, далее нарушение работы метаболических путей и физиологических процессов, в частности, происходит избыточное накопление метаболитов, которые являются причиной заболевания. Часто сбалансированное питание способно нейтрализовать последствия геномных патологий, путем создания соответствующих рационов. Здесь работают другие омйки: нутригенетика, нутригеномика, нутрициология, фудомика, феномика, диетология и физиология питания в рамках модели: Генотип – Среда обитания – Фенотип. Ключевым фактором в заболеваниях, связанных с питанием, является химический состав и сбалансированность рациона.

Методы и алгоритмы: Для достижения этих целей были созданы высокоструктурированные реляционные омиксные модели БД и БЗ [1, 2], алгоритмы и программы, которые позволили формализовать биохимические и физиологические данные и знания и погрузить в системно-кибернетическую структурированную модель (СКСМ), а также в расширенную модель математического программирования, позволяющие учитывать большое разнообразие норм и ограничений в рационе [3]. Созданные модели позволяют перейти от больших количественных данных BigData, характерным для омик и химического состава продуктов, блюд и рационов, к качественным малым данным SmallData – диагнозам, рационам, удобно используемым врачами, диетологами, биологами [4, 5]. Для этого в СКСМ осуществляются анализы рационов в прямой цепи управления и оптимальный синтез дополнительных рационов в цепи

обратной связи, на основе кибернетических принципов наблюдаемости, управляемости, достижимости.

Результаты: В состав предлагаемой СКСМ [3] на базе модели пространства состояний входят подсистемы: норм питания (БП); анализа рациона (ИИС); оценки рациона (БОД); оптимальный регулятор в цепи обратной связи (БПР) (рис. 1). С векторами: X , $\hat{X}$ – соответственно векторы состояния, оценки состояния, Z – измерения, W – возмущения, V – ошибок измерения, ΔX – отклонение от норм, ΔU – дополнительное управление, $U_{\text{п}}$ – плановые продукты, $X_{\text{п}}$ – плановые нутриенты.



БП – блок планирования формирующий нормативные требования к управляемым объектам и процессам ($U_{\text{п}}, X_{\text{п}}$)

ОУ – объект или процесс управления с векторами управления, состояния и возмущения (U_{Σ}, X, W)

ИИС – информационно-измерительная система с векторами измерения и ошибок измерения (Z, V)

БОД – блок оценивания и диагностирования с вектором оценки состояния ($\hat{X}$)

БС – блок сравнения оценки текущего состояния и плана ($\Delta X = \hat{X} - X_{\text{п}}$)

БПР – регулятор в цепи обратной связи – блок принятия дополнительных решений (ΔU)

Рис. 1. Системно-кибернетическая структурированная модель

СКСМ строится по технологии глубоких экспертных систем на основе БД и БЗ с дополнительной оптимизацией на базе методов математического программирования (реализация задачи о диете). Оптимальный регулятор в цепи обратной связи определяет дефицитные нутриенты и дает рекомендации по качественному и количественному составу дополнительных продуктов.

Анализ традиционных рационов.

Применительно к задаче анализа и оптимального синтеза рациона в данном случае не учитываются возмущения (колебания химического состава продуктов), а также ошибки измерений, возникающие в процессе взвешивания и анализа химического состава продуктов. Это существенно упрощает задачу и позволяет перейти от стохастических моделей к детерминированным. Анализ химического состава традиционного рациона производится на базе системы алгебраических уравнений:

$$X = A \cdot U,$$

где U – вектор потребленных продуктов в рационе; X – вектор нутриентов;

$A = (a_{ij})$ – матрица $m \times n$, где a_{ij} – элементы матрицы, характеризующие количество i -го нутриента в j -ом продукте, m – количество анализируемых нутриентов, n – количество продуктов.

Оптимальный синтез корректирующих рационов.

Далее проводится сравнение результатов анализа с требуемыми нормами и определяется дефицит необходимых нутриентов $\Delta X = X - X_{\text{п}}$, т.е. вместо $\hat{X}$ в СКСМ используем просто X .

После этого решаем задачу оптимального синтеза добавления дефицитных нутриентов на основе методов линейного программирования путем минимизации линейной функции: $K = C^T \cdot \Delta U$,
где $\Delta U \geq 0$, $A \cdot \Delta U \geq \Delta X$, $C^T = (c_1, c_2, \dots, c_n)$ – вектор-строка $1 \times n$, характеризующая стоимость соответствующих продуктов.

Для решения задач анализа и оптимального сбалансированного синтеза рациона созданы БД и БЗ: норм питания, сырья, рецептур блюд, химического состава сырья и конкретных блюд, норм и ограничений в питании, суточных рационов, введены формулы питания (по А.А. Покровскому). Разработанные БД и БЗ интегрированы с алгоритмами и программами реализации симплекс метода. Для автоматического построения многомерных моделей (размерность десятки и сотни) разработаны запросы, которые вошли в состав дополнительных ограничений задачи линейного программирования и обеспечивают автоматический синтез многомерных моделей. В качестве критерия оптимизации могут быть использованы различные критерии: стоимости, минимизации калорий, веса, ограничений не только на нутриенты и их соотношения, но и на потребляемые первичные продукты.

На основе моделей баз данных и знаний, экспертных систем осуществляется: интеллектуальный анализ и интерпретация омиксных данных; формируются оптимальные сбалансированные рационы. При этом прослеживается вся цепочка причинно-следственных связей и событий в омиксных биохимических и физиологических процессах, а также синтезе оптимального сбалансированного рациона. Таким образом, мы имеем систему объяснений, характерную для экспертных систем, в отличие от интеллектуальных систем на базе обучаемых нейронных сетей, которые работают, как «черный ящик».

Выводы: Созданная биоинформационная система позволяет эффективно интегрировать многочисленные дифференциальные фундаментальные знания и внедрить в практику интеллектуальные методы анализа и интерпретации омиксных данных, автоматизировать процессы анализа и оптимального сбалансированного синтеза рационов, реализовать любую существующую диету. Для этого в автоматическом режиме создаются модели данных из созданных БД и многочисленных ограничений на сырье, первичные продукты, диеты. В случае, если не удастся синтезировать рацион на базе традиционных продуктов, т.е. рацион не управляем, либо не достигим (превышение массы, избыточные калории, нутриенты, ограничения в продуктах и т.д.), то можно добавлять в модель различные пищевые добавки, витамины. Система сама определит, что и в каком количестве подходит для сбалансированности синтезируемого рациона. Достоинством разработанной системы является то, что она не ограничивает индивида строгими рамками рациона, который получается на базе оптимального синтеза, но позволяет питаться по традиционной схеме питания, но с дополнительными продуктами, получаемыми в цепи обратной связи для коррекции рациона. Обычно, это составляет значительно меньшую долю общих затрат на питание, но требует более строгого подбора сырья и готовых продуктов, а также дисциплины питания.

From the genome to an optimal balanced diet

Shlikht A.*, Kramorenko N.

Far Eastern Federal University, Vladivostok, Russia

* schliht@mail.ru

Key words: genomics; phenomics; foodomics; optimal synthesis of diets; databases; knowledge bases; expert systems

Motivation and Aim: The goal of the study is to create an integrated intelligent bioinformatics system that allows you to trace the chain of omics biochemical processes, including: genomics, epigenomics, transcriptomics, exomics, proteomics, metabolomics, reactomics, genomontology, genomic mutations and then proceed to the diagnosis of genetic diseases, analysis and synthesis of optimal balanced diets. Existing omics data repositories based on international portals (NCBI, KEGG, Ensemble, UniProt, GO, etc.) do not adequately provide automatic analysis and interpretation of omics data. Artificial intelligence (AI) technologies based on databases (DB) and knowledge bases (KB) are becoming crucial here. Omics-centered biological objects and processes have become a challenge for mathematics, informatics and control theory. Taking into account the genomic status of an individual is an important feature of the modern theory and practice of life sciences. Mutations in genes produce mutations in the corresponding proteins. If this protein is an enzyme, then there is a disruption of the biochemical reactions in which the enzyme participates, further disruption of the metabolic pathways and physiological processes, in particular, there is an excessive accumulation of metabolites that are the cause of the disease. Often, a balanced diet is able to neutralize the effects of genomic pathologies by creating appropriate diets. Other omics work here: nutrigenetics, nutrigenomics, nutritionology, foodomics, phenomics, dietetics and nutrition physiology within the framework of the model: Genotype – Habitat – Phenotype. The key factor in nutrition-related diseases is the chemical composition and balance of the diet.

Methods and Algorithms: To achieve these goals, highly structured relational omics models of DB and KB [1, 2] were created, along with algorithms and programs. They made it possible to formalize biochemical and physiological data and knowledge and immerse them in the system-cybernetic structured model (SCSM), as well as in an extended model of mathematical programming, which allow taking into account a wide variety of norms and restrictions in the diet [3]. The created models make it possible to move from large quantitative data BigData, characteristic of omics and the chemical composition of products, dishes and diets, to high-quality SmallData – diagnoses, diets, conveniently used by doctors, nutritionists, biologists [4, 5]. For this purpose, the SCSM analyzes rations in the direct control chain and the optimal synthesis of additional rations in the feedback chain, based on the cybernetic principles of observability, controllability, and reachability.

Results: The proposed SCSM [3] based on the state space model includes subsystems: nutrition standards (BS); diet analysis (IMS); diet assessment (ADB); optimal regulator in the feedback control (DSS) (Figure 1). With vectors: X , $\hat{X}$ – vectors of state, condition assessment, respectively, Z – measurements, W – perturbations, V – measurement errors, ΔX – deviation from norms, ΔU – additional control, U_p – planned products, X_p – planned nutrients.

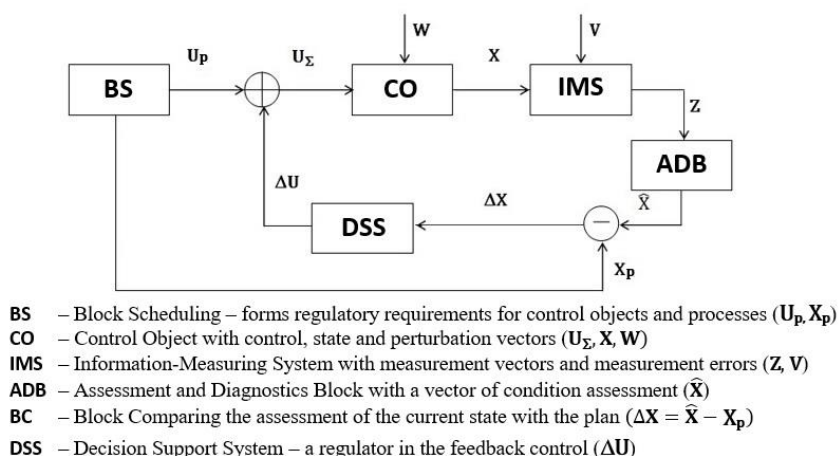


Fig. 1. The System-Cybernetic Structured Model

SCSM is based on the technology of deep expert systems based on DB and KB with additional optimization based on mathematical programming methods (implementation of the diet problem). The optimal regulator in the feedback control identifies deficient nutrients and provides recommendations on the qualitative and quantitative composition of additional products.

Analysis of traditional diets.

In relation to the problem of analyzing and optimally synthesizing a diet, in this case, disturbances (fluctuations in the chemical composition of foods), as well as measurement errors that occur during weighing and analyzing the chemical composition of food products, are not considered. This greatly simplifies the task and allows you to move from stochastic models to deterministic ones. The analysis of the chemical composition of traditional diet is conducted based on a system of algebraic equations:

$$X = A \cdot U,$$

where U – the vector of consumed products in the diet; X – the nutrient vector;

$A = (a_{ij})$ – the matrix $m \times n$, where a_{ij} – the elements of the matrix that characterize the amount of the i -th nutrient in the j -th product, m – the number of analyzed nutrients, n – the number of products.

Optimal synthesis of corrective rations.

Next, the results of the analysis are compared with the required standards and the deficiency of necessary nutrients is determined $\Delta X = X - X_p$, i.e., instead of $\hat{X}$ in SCSM we simply use X .

After that, we solve the problem of optimal synthesis of the addition of deficient nutrients based on linear programming methods by minimizing the linear function: $K = C^T \cdot \Delta U$, where $\Delta U \geq 0$, $A \cdot \Delta U \geq \Delta X$, $C^T = (c_1, c_2, \dots, c_n)$ – vector-line $1 \times n$ characterizing the values of the respective products.

To solve the problems of analysis and optimal balanced synthesis of the diet, DB and KB were created: nutrition norms, raw materials, recipes of dishes, chemical composition of raw materials and specific dishes, norms and restrictions in nutrition, daily rations, nutrition formulas were introduced (according to A.A. Pokrovsky). The

developed DB and KB are integrated with algorithms and programs for implementing the simplex method. For the automatic construction of multidimensional models (tens and hundreds of dimensions), queries have been developed that are part of additional constraints of the linear programming problem and provide automatic synthesis of multidimensional models. Various criteria can be used as an optimization criterion: cost, calorie minimization, weight minimization, restrictions not only on nutrients and their ratios, but also on consumed primary products.

Based on database models and knowledge bases, as well as expert systems, intelligent analysis and interpretation of omics data is performed, and optimal balanced diets are generated. At the same time, the entire chain of cause-and-effect relationships and events in the omics biochemical and physiological processes, as well as the synthesis of an optimal balanced diet, is traced. Thus, we have a system of explanations that is typical for expert systems, in contrast to intelligent systems based on trained neural networks, which operate like a "black box".

Conclusion: The created bioinformatics system makes it possible to effectively integrate numerous differential fundamental knowledge and put into practice intelligent methods of analysis and interpretation of omics data, automate the processes of analysis and optimal balanced synthesis of diets, implement any existing diet. To do this, data models are automatically created from the databases created and numerous restrictions on raw materials, primary products and diets. If it is not possible to synthesize a diet based on traditional products, i. e. the diet is not manageable or achievable (excess weight, excess calories, nutrients, food restrictions, etc.), then various dietary supplements and vitamins can be added to the model. The system itself will determine what and in what amount is suitable for the balance of the synthesized diet. The advantage of the developed system is that it does not restrict the individual to strict dietary limits, which are obtained through optimal synthesis, but rather allows them to eat according to a traditional nutritional plan, with additional products derived from the feedback control to correct the diet. This usually accounts for a much smaller proportion of the total cost of nutrition, but requires more rigorous selection of raw materials and finished products, as well as nutritional discipline.

Список литературы/References

1. Shlikht A.G., Kramorenko N.V. Artificial intelligence in the problems of analysis and interpretation of omics human data. *BGRS\SB-2018*. 2018;78. doi 10.18699/BGRSSB-2018-054
2. Shlikht A.G., Kramorenko N.V. Deep bioinformatics expert system of analysis, modeling and interpretation of omics BigData of the human genome. *MM-HPC-BBB-2018*. 2018;66. doi 10.18699/MM-HPC-BBB-2018-57
3. Шлихт А.Г., Краморенко Н.В. Геном-центрированная интеллектуальная системно-кибернетическая структурированная модель в задачах оптимального синтеза рационов человека. *Актуальные вопросы биологической физики и химии*. 2020;5(2):301-304 [Shlikht A.G., Kramorenko N.V. Genome-centered intelligent system-cybernetic structured model for optimal synthesis of human diet. *Aktualnyye voprosy biologicheskoy fiziki i khimii*. = *Russian Journal of Biological Physics and Chemistry*. 2020;5(2):301-304 (in Russian)]
4. Tamkovich S.N., Tutanov O.S., Serdukov D.S. et al. Protein Content of Circulating Nucleoprotein Complexes. *Adv Exp Med Biol*. 2016;924:133-136. doi 10.1007/978-3-319-42044-8_26
5. Naryzhny S., Volnitskiy A., Kopylov A., Zorina E., Kamyshinsky R., Bairamukov V., Garaeva L., Shlikht A., Shtam T. Proteome of Glioblastoma-Derived Exosomes as a Source of Biomarkers. *Biomedicines*. 2020;8(7):216. doi 10.3390/biomedicines8070216