

Задача кластеризации генотипов в системе поддержки принятия решений при управлении персонифицированным питанием

В. Н. Иванова¹, В. И. Карпов², Ю. И. Сидоренко³
Московский государственный университет технологий
и управления им. К. Г. Разумовского
(Первый казачий университет)
¹rektorat@mgutm.ru, ²vikarp@mail.ru,
³y.sidorenko@mgutm.ru

Н. А. Жученко
Первый Московский государственный медицинский
университет имени И.М. Сеченова
zhychenko64@mail.ru

Аннотация. В статье рассмотрена задача кластеризации генотипов, применяемая в системе поддержки принятия решений при управлении персонифицированным питанием. Приведены результаты решения задачи для конкретных объектов, прошедших лабораторные исследования состояний генов.

Ключевые слова: генотип; персонифицированное питание; кластеризация многомерных объектов; мягкие вычисления; метрика характеристик

I. ВВЕДЕНИЕ

Современные научные исследования в области нутригеномики показали, что подбор оптимального рациона питания с учетом генетических особенностей на 200-300% повышает эффективность таких программ по сравнению с традиционными подходами [1–5]. В настоящем докладе рассматривается применение мягких вычислений (метода кластеризации многомерных объектов) для определения типовых кластеров управления персонифицированным питанием клиентов на основе их генотипов.

II. ПОСТАНОВКА ЗАДАЧИ

Дано:

C_0 – исходное множество многомерных объектов,

$$C_0 = \{S_n\}, n = 1, \dots, N$$

$Mr(M)$ – метрика характеристик, $Mr(i)$ – весовой коэффициент i -ой характеристики, $i = 1, \dots, M$;

$X(n, i)$ – i -я характеристика состояния i -го объекта n , $n = 1, \dots, N$, $i = 1, \dots, M$.

В табл. 1 представлены результаты анализа состояния генов для одного человека. В столбце 4 отмечено состояние гена, названного в столбце 2. Для цифровой обработки в столбце 4 ставятся баллы 0,1,2. При

Работа выполнена в соответствии с госзаданием МОН «Исследование об организации питания и создании условий для сохранения здоровья обучающихся в общеобразовательных организациях», шифр проекта № 27.9928.2017/НМ.

выставлении баллов 0 1 2, предполагается, что 1 и 2 это средний и высокий риск предрасположенности к болезненному состоянию, 0 означает отсутствие риска к болезненному состоянию. Для оценки предрасположенности субъектов экспертизы к наиболее значимым болезненным состояниям экспертами были определены четырнадцать генов, по состоянию которых и была проведена кластеризация.

К числу контролируемых метаболических систем были отнесены: система биотрансформации ксенобиотиков; метаболизм витаминов; метаболизм липидов и углеводов. Указанные метаболические системы кодируются панелями генов, включающими следующие гены.

Ген *CYP1A1* (rs1048943). Ген *CYP1A1* кодирует аминокислотную последовательность фермента цитохрома P450 1A1 (I фаза биотрансформации), локализован на 15 хромосоме. Он относится к суперсемейству ферментов цитохрома P450. Представители этой большой группы являются монооксигеназами и катализируют реакции метаболизма ксенобиотиков, а также участвуют в синтезе холестерина, стероидов и других липидов.

Ген *CYP1A2* (rs762551). Ген *CYP1A2* находится в 15 хромосоме, локусе 15q22-qter, кодирует фермент полипептид 2-го типа цитохрома P-450, который локализован в эндоплазматическом ретикулуме клеток печени и активируется под воздействием полициклических ароматических углеводородов (ПАУ), таких как кофеин. Ферментная активность *CYP1A2* у разных людей отличается, поэтому отличается их способность метаболизировать кофеин и ПАУ.

Гены *GSTM1*, *GSTT1* и *NAT2* представляют (II фазу биотрансформации). Ген *GSTM1* расположенный в области короткого плеча первой хромосомы (1p13.3), экспрессируется в печени и клетках крови.. Глутатионопосредованная детоксикация играет ключевую роль в обеспечении резистентности клеток к перекисному окислению липидов, свободным радикалам, алкилированию белков и в предотвращении повреждений ДНК.

Ген NAT2 (rs1799929, rs1799930) кодирует аминокислотную последовательность фермента N-ацетилтрансферазы-2 (II фаза биотрансформации), локализован на 8-ой хромосоме. Этот фермент катализирует реакции переноса ацетильной группы на ксенобиотики (ариламины, гидразины) и канцерогены. N-ацетилтрансфераза-2 у человека активно экспрессируется в печени, тканях тонкого и толстого кишечника.

Метаболизм витаминов, отражает состояние антиоксидантной системы организма. Анализ полиморфизмов в генах *BCMO1*, *NBPF3*, *FUT2*, *MTHFR*, *APOA5*, *GC* оценивает метаболизм витаминов А, В₆, В₁₂, В₉, Е и D соответственно.

Ген BCMO1 (rs7501331), бета-каротин монооксидазы локализован на 16-й хромосоме, отражает способность превращать бета-каротин в ретиноловые эфиры. Полиморфизм A379V изменяет каталитическую активность бета-каротина и ассоциируется со снижением синтеза витамина А.

Ген NBPF3(Alpl) (rs4654748), неспецифической щелочной фосфатазы, локализован в первой хромосоме (1p36.12). Полиморфизм гена влияет на уровень витамина В₆ в крови, поскольку *Alpl* является основным ферментом, ответственным за клиренс витамина В₆.

Ген FUT2 (rs602662), альфа-1,2-фукозилтрансферазы, расположен на 19-ой хромосоме. У носителей вариантов в гене *FUT2* снижается вероятность развития мальабсорбции (нарушение всасывания) витамина В₁₂ в кишечнике и, как следствие, выше содержание витамина В₁₂ в крови.

Ген MTHFR (rs1801133), локализован на первой хромосоме (1p36.22), кодирует белок метилентетрагидрофолатредуктазу – внутриклеточный фермент, участвующий в превращении гомоцистеина в метионин.

Межгенный маркер (rs12272004), расположен возле гена *APOA5*.

Ген GC (rs2282679), расположенный на хромосоме 4p12, кодирует белок связывающий витамин D, ассоциирован с изменением уровня концентрации витамина D в крови. Предрасположенность к избыточной массе тела, оценивалась по метаболизму липидов, углеводов, а также склонности к повышенному аппетиту.

Ген ADRB2 (rs1042713, rs1042714), локализован на пятой хромосоме (5q32), кодирует бета-2-адренергический рецептор, встроенный в цитоплазматическую мембрану клетки, имеющий высокую степень родства к адреналину и обеспечивающий повышение или понижение активности иннервируемой ткани или органа. Бета-2-рецепторы преимущественно располагаются на поверхности клеток гладких мышц бронхов и артерий скелетных мышц, печени и жировой ткани. Выявлена ассоциация полиморфизмов гена *ADRB2* с метаболическим синдромом, избыточным весом и ожирением.

Ген PPARD (rs2016520), относится к семейству *PPAR* – ген рецепторов активации пролиферации пероксисом кодирует белок *PPARδ*, который специфически связывается с промоторами генов жирового и углеводного

обменов и регулирует их транскрипцию. **Ген PPARD** локализован в локусе 6p21.1- p21.2, активно экспрессируется в жировой ткани и в медленных мышечных волокнах скелетных мышц. Полиморфизм гена *PPARD*, способствует большему катаболизму жиров и в определенной степени снижает риск развития ожирения.

Ген SR(HTR2A) (rs6313) – постсинаптический рецептор серотонина 2А типа, локализован на 13-ой хромосоме в положении 13q14-q21 и состоит из трех экзонов и двух интронов. Однонуклеотидная замена тимина на цитозин в положении 102 (Rs6313, T102C) располагается в первом экзоне гена *HTR2A*, данный полиморфизм связан с уровнем экспрессии гена и является маркером повышенного аппетита.

Состояние вышеописанных генов было оцифровано по вышеописанной методике.

Метрика $Mp(M)$ нормирована (1).

$$\sum_{i=1}^N Mp(i) = 1 \quad (1)$$

Исходное множество C_0 нужно разделить на множества кластеров C_k (смотри формулы 2,3):

$$C_0 = \{ C_k \} \quad k=1,..,K \quad (2)$$

$$C_k = \{ S_z \}, \quad z=1,..,N_k \quad (3)$$

Любая пара кластеров не имеет общих элементов, т. е. любой объект может только в одном кластере (4):

$$\forall C_k \in C_0, \forall C_l \in C_0 : C_k \cap C_l = \emptyset \quad (4)$$

Требуется определить такие C_k , которые максимизируют критерий U (см. формулу 5):

$$U(K_o) = \max_{K=\overline{N,2}} \{ U_1(K) - U_2(K) \} \quad (5)$$

где $U(K_o)$ – оптимальное значение критерия качества кластеризации;

$U_1(K)$ – компактность классов при K кластерах;

$U_2(K)$ – мера близости классов при K кластерах.

Мера сходства между двумя объектами определяется на основании потенциальной функции $f(S_i, S_j)$:

$$f(S_i, S_j) = \frac{1}{1 + \rho^2(S_i, S_j)},$$

$$\rho(S_i, S_j) = \sqrt{\sum_{m=1}^M (Mp(m) * (X_{im} - X_{jm}))^2}$$

$$U_1(K) = \frac{1}{K} \sum_{k=1}^K \frac{2}{N_k(N_k - 1)} \sum_{S_i \in C_k} \sum_{S_j \in C_k} f(S_i, S_j), \quad i \neq j$$

где K – количество классов на текущем шаге классификации; C_k – k -ый класс объектов; N_k – количество объектов в классе C_k ; $f(S_i, S_j)$ – потенциальная

функция двух объектов S_i и S_j ; $\rho(S_i, S_j)$ – расстояние между объектами S_i и S_j в пространстве характеристик X , с учетом метрики

$$U_2(K)=\frac{2}{K(K-1)}\sum_{C_k\in C_p}\sum_{C_l\in C_p}F(C_k,C_l),\quad k\neq l$$

$$F(C_k,C_l)=\frac{1}{N_kN_l}\sum_{S_i\in C_k}\sum_{S_j\in C_l}f(S_i,S_j)$$

Таким образом, оптимальное разбиение на кластеры предполагает максимизацию критерия (5). Содержательно такая постановка означает, что в каждом кластере собраны родственные объекты, а между объектами разных кластеров имеются значительные различия. По своему классу, эта задача относится к задачам мягких вычислений, решаемых методами целочисленного математического программирования. Для ее решения использовался комплекс программ оценки качества многомерных объектов [6].

III. РЕЗУЛЬТАТЫ РЕШЕНИЯ ЗАДАЧИ КЛАСТЕРИЗАЦИИ

На рис. 1 представлены результаты решения задачи кластеризации по анкетам для мужчин. После обработки анкет женщин также было выделено три кластера. Интерпретация и характеристика этих кластеров, характерных для мужчин и женщин, представлена в табл. 2.

На рис. 2 представлена структурно-функциональная модель подсистемы принятия решений в процессе управления персонализированным питанием. На блок 1 поступает поток сведений о состоянии генов выбранной группы потребителей. На выходе этого блока – информация о состоянии генов участников группы, которая используется в блоке 2 для решения задачи кластеризации. Далее для каждого кластера в блоке 3 разрабатывается оптимальный рацион питания.



Рис. 1. Результаты кластеризации мужчин по состоянию их генов

ТАБЛИЦА I ВХОДНАЯ ИНФОРМАЦИЯ ДЛЯ ОДНОГО ОБЪЕКТА

Номер гена	ГЕНЫ	Полиморфизм/ мутация	wt/wt	Генотип
1	CYP1A1	wt, 2A, 2B, *4: 4887C>A, 4889A>G, 6235T>C)	0	wt/wt
2	CYP1A2	intron 1 -164 C>A (*1A>*1F)	2	*1A/*1A
3	GSTT1	wt, del	0	+
4	GSTM1	wt, del	0	+
5	NAT2	*4(wt) (T341C), 481C>T *5, 590G>A *6, 857G>A *7 S1, S2, S3 (*5, *6, *7) – медленные ацетиляторы	2	S1/S2
6	MTHFR	677 C>T	0	C/C
7	ADRB2	48A>G, 81C>G	2	G/G, C/G
8	SR(HTR2 A)	102 T>C	1	T/C
9	PPARD	294 T>C	0	T/T
10	NBPF3(AL PL)	rs4654748	2	C/C
11	FUT2	rs602662	0	A/A
12	BCMO1	rs7501331	0	C/C
13	GC(D)	rs2282679	0	
14	APOA5	INTERGENIC Rs12272004	2	C/C

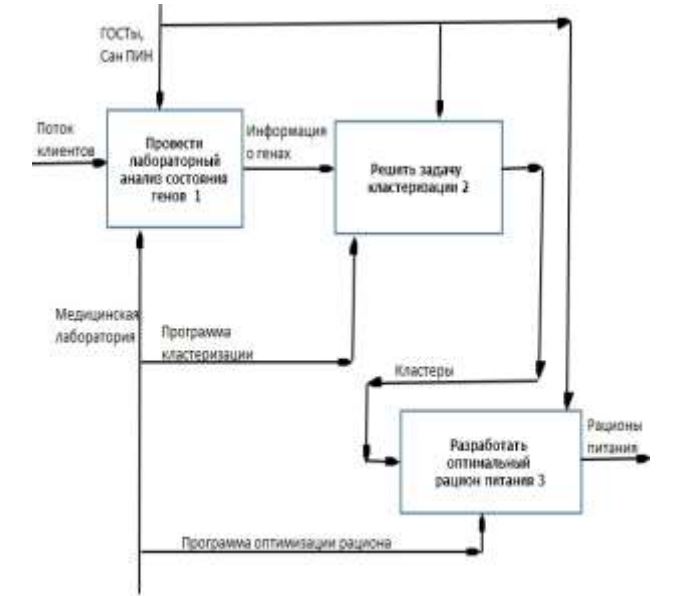


Рис. 2. Структурно-функциональная модель подсистемы принятия решений в процессе управления персонализированным питанием

IV. ЗАКЛЮЧЕНИЕ

В настоящем докладе рассмотрена постановка задачи кластеризации информации о состоянии генов потребителей для проектирования их рациона персонализированного питания. Предложена программа для решения этой задачи. Представлен результат решения

задачи кластеризации для реальных потребителей, по результатам анализа их генетического статуса, основанного на оценке 14 генов. Разработана методика кластеризации потребителей в рамках заданной социальной группы по признаку угроз их болезненного состояния, определенных по данным генома. Методика позволяет выделить кластеры с различным уровнем устойчивости к наследственным заболеваниям болезненным состояниям.

ТАБЛИЦА II ИНТЕРПРЕТАЦИЯ КЛАСТЕРОВ, ПОЛУЧЕННЫХ ПОСЛЕ РЕШЕНИЯ ЗАДАЧИ КЛАСТЕРИЗАЦИИ

Заболевание/состояния	Кластер №1	Кластер №2	Кластер №3
Окислительный стресс	Окислительный стресс (низкая или средняя предрасположенность к окислительному стрессу)	Окислительный стресс (средняя или высокая предрасположенность к окислительному стрессу)	Окислительный стресс высокий
Антиоксидантный	Антиоксидантный статус высокий	Антиоксидантный статус средний	Антиоксидантный статус низкий
Ожирение	Нет (не выявлена предрасположенность к ожирению)	Нет или есть (не выявлена предрасположенность или средняя предрасположенность к ожирению)	Есть (выявлена средняя предрасположенность к ожирению)
Питание/диета, с обогащением	Сбалансированное питание	Питание с низким содержанием углеводов, с обогащением витаминами	Питание с низким содержанием углеводов, с высоким уровнем обогащения витаминами

СПИСОК ЛИТЕРАТУРЫ

- [1] Сидоренко М.Ю. Персонализированное питание. М.: ДеЛи плюс, 2016. 192 с.
- [2] Сидоренко М. Ю., Стройкова А. А. Методология проектирования персонализированных рационов питания с учетом механизма психологической мотивации потребителей, Товаровед. продовольственных товаров, 2012, № 10. с. 15-19.
- [3] Nielsen DE, Shih S, El-Sohemy A.// Perceptions of genetic testing for personalized nutrition: a randomized trial of DNA-based dietary advice. J Nutrigenet Nutrigenomics. 2014;7(2):94-104. doi: 10.1159/000365508. Epub 2014 Aug 22.
- [4] Kim YJ, Huh I, Kim JY, Park S, Ryu SH, Kim KB, Kim S, Park T, Kwon O. // Integration of Traditional and Metabolomics Biomarkers Identifies Prognostic Metabolites for Predicting Responsiveness to Nutritional Intervention against Oxidative Stress and Inflammation. - Nutrients. 2017 Mar 4;9(3). pii: E233. doi: 10.3390/nu9030233.
- [5] Takenaka A, Nakamura S, Mitsunaga F, Inoue-Murayama M, Udon T, Suryobroto B.// Human-specific SNP in obesity genes, adrenergic receptor beta2 (ADRB2), Beta3 (ADRB3), and PPAR γ 2 (PPARG), during primate evolution. - PLoS One. 2012;7(8):e43461. doi: 10.1371/journal.pone.0043461. Epub 2012 Aug 24.
- [6] Карпов В.И., Печеная Л.Т., Пати М.О. Моделирование механизма гармонизации промышленной и торговой политики предприятий кондитерской промышленности, ж. Мир агробизнеса, 2012 г. №1, с. 24-26.