ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ
ҰЛТТЫҚ ҒЫЛЫМ АКАДЕМИЯСЫ

әл-Фараби атындағы Қазақ ұлттық университетінің

ХАБАРЛАРЫ

ИЗВЕСТИЯ

НАЦИОНАЛЬНОЙ АКАДЕМИИ
НАУК РЕСПУБЛИКИ КАЗАХСТАН
Казахский национальный
университет имени аль-Фараби

N E W S

OF THE ACADEMY OF SCIENCES
OF THE REPUBLIC OF
KAZAKHSTAN
al-Farabi Kazakh National University

SERIES
PHYSICO-MATHEMATICAL

2 (342)

APRIL – JUNE 2022

PUBLISHED SINCE JANUARY 1963

PUBLISHED 4 TIMES A YEAR

ALMATY, NAS RK

«ҚР ҰҒА Хабарлары. Физика-математикалық сериясы».
ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

Меншіктеуші: «Қазақстан Республикасының Ұлттық ғылым академиясы» РҚБ (Алматы қ.). Қазақстан Республикасының Ақпарат және қоғамдық даму министрлігінің Ақпарат комитетінде 14.02.2018 ж. берілген **No 16906-Ж** мерзімдік басылым тіркеуіне қойылу туралы куәлік.

Тақырыптық бағыты: *физика және ақпараттық коммуникациялық технологиялар сериясы.*

Қазіргі уақытта: *«ақпараттық технологиялар»бағыты бойынша ҚР БҒМ БҒСБК ұсынған журналдар тізіміне енді.*

Мерзімділігі: *жылына 4 рет.*

Тиражы: *300 дана.*

Редакцияның мекен-жайы: *050010, Алматы қ., Шевченко көш., 28, 219 бөл., тел.: 272-13-19*
http://www.physico-mathematical.kz/index.php/en/

Типографияның мекен-жайы: «Аруна» ЖК, Алматы қ., Мұратбаев көш., 75.

«Известия НАН РК. Серия физика-математическая».
ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

Собственник: *Республиканское общественное объединение «Национальная академия наук Республики Казахстан» (г. Алматы).*
Свидетельство о постановке на учет периодического печатного издания в Комитете информации Министерства информации и общественного развития Республики Казахстан **No 16906-Ж** выданное 14.02.2018 г.
Тематическая направленность: *серия физика и информационные коммуникационные технологии.*
В настоящее время: *вошел в список журналов, рекомендованных ККСОН МОН РК по направлению «информационные коммуникационные технологии».*
Периодичность: *4 раз в год.*
Тираж: *300 экземпляров.*
Адрес редакции: *050010, г. Алматы, ул. Шевченко, 28, оф. 219, тел.: 272-13-19*
http://www.physico-mathematical.kz/index.php/en/

Адрес типографии: ИП «Аруна», г. Алматы, ул. Муратбаева, 75.

News of the National Academy of Sciences of the Republic of Kazakhstan.
Physical-mathematical series.
ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

Owner: RPA «National Academy of Sciences of the Republic of Kazakhstan» (Almaty). The certificate of registration of a periodical printed publication in the Committee of information of the Ministry of Information and Social Development of the Republic of Kazakhstan **No. 16906-Ж,** issued 14.02.2018
Thematic scope: *series physics and information technology.*
Currently: *included in the list of journals recommended by the CCSES MES RK in the direction of «information and communication technologies».*
Periodicity: *4 times a year.*
Circulation: *300 copies.*
Editorial address: *28, Shevchenko str., of. 219, Almaty, 050010, tel. 272-13-19*
http://www.physico-mathematical.kz/index.php/en/

Address of printing house: ST «Aruna», 75, Muratbayev str, Almaty.

NEWS OF THE NATIONAL ACADEMY OF SCIENCES
OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
ISSN 1991-346X
Volume 2, Number 342 (2022),19–38
https://doi.org/10.32014/2022.2518-1726.127

УДК 519.68.02
МРНТИ 53.31.23

Е.С. Голенко[1]*, А.А. Исмаилова[1], А.С. Жумаханова[2]

[1]Казахский агротехнический университет им. С. Сейфуллина,
Нур-Султан, Казахстан;
[2]Евразийский национальный университет им. Л.Н. Гумилёва,
Нур-Султан, Казахстан.
E-mail: *golenko.katerina@gmail.com*

ПРЕДСКАЗАНИЕ ФУНКЦИЙ БЕЛКОВ ПРИ ПОМОЩИ БАЗЫ ДАННЫХ «GENE ONTOLOGY» И МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ

Аннотация. Прогнозирование функций белков является важной частью аннотации генома. В последнее время методы прогнозирования быстро развиваются благодаря появлению высокопроизводительных технологий секвенирования. Среди доступных баз данных для определения терминов функций белка важным ресурсом, описывающим функциональные свойства белков, является Gene Ontology (GO). Gene Ontology предлагает классификацию функций, которая базируется на некотором списке или словаре четко сформулированных терминов, каждый из которых принадлежит одной из категорий – молекулярным функциям, биологическим процессам и клеточным компонентам. Из этой базы данных можно по названию белка или его идентификационному номеру найти присвоенные ему термины Gene Ontology или аннотации, которые были сделаны на основе расчётных или экспериментальных данных. Каждому термину в Gene Ontology присваивается ряд атрибутов: уникальный цифровой идентификатор, название, словарь, к которому термин принадлежит, и определение. Термины могут иметь синонимы, которые делятся на точно соответствующие значению термина, более широкие, более узкие и имеющие

некоторое отношение к термину. Исследователи используют различные подходы для эффективного прогнозирования терминов GO. Между тем, глубинное обучение, быстро развивающаяся дисциплина в подходе, основанном на данных, демонстрирует впечатляющий потенциал в отношении присвоения терминов GO аминокислотным последовательностям. В данной статье авторами рассмотрены доступные сегодня вычислительные методы аннотации GO для белков, начиная от традиционного и заканчивая методом глубинного обучения. Также вынесены на обсуждение основные проблемы в этой области и подчеркнуты будущие направления предсказания функций белка с помощью GO.

Ключевые слова: Gene Ontology, предсказание функций белка, машинное обучение, глубинное обучение, аннотация белков.

Е.С. Голенко[1*], А.А. Исмаилова[1], А.С. Жумаханова[2]

[1]С. Сейфуллин атындағы Қазақ агротехникалық университеті,
Нұр-Сұлтан, Қазақстан;
[2]Л.Н. Гумилёв атындағы Еуразия ұлттық университеті,
Нұр-Сұлтан, Қазақстан.
E-mail: *golenko.katerina@gmail.com*

«GENE ONTOLOGY» БАЗАСЫН ЖӘНЕ МАШИНАЛЫҚ ОҚЫТУ ҮЛГІЛЕРІН ПАЙДАЛАНА ОТЫРЫП АҚУЫЗ ФУНКЦИЯЛАРЫН БОЛЖАУ

Аннотация. Ақуыз функциясын болжау геномдық аннотацияның маңызды бөлігі болып табылады. Соңғы уақытта жоғары өнімді секвенирлеу технологияларының пайда болуына байланысты болжау әдістері қарқынды дамып келеді. Ақуыз функциясының терминдерін анықтауға арналған дерекқорлардың арасында Gene Ontology (GO) белоктардың функционалдық қасиеттерін сипаттайтын маңызды ресурс болып табылады. Гендік онтология әрқайсысы категориялардың біріне жататын - молекулалық функцияларға, биологиялық процестерге және жасушалық компоненттерге жататын нақты анықталған терминдердің тізбесіне немесе сөздік қорына негізделген функциялардың жіктелуін ұсынады. Бұл дерекқордан ақуыздың атын немесе оның сәйкестендіру нөмірін оған тағайындалған Гендік онтология терминдерін немесе есептелген немесе эксперименттік деректер негізінде жасалған аннотацияларды табу үшін пайдалануға болады. Гендік онтологиядағы

әрбір терминге бірнеше атрибуттар тағайындалады: бірегей сандық идентификатор, атау, термин жататын сөздік және анықтама. Терминдердің синонимдері болуы мүмкін, олар терминнің мағынасына сәйкес келетін, кеңірек, тар және терминге қандай да бір қатысы бар болып бөлінеді. Зерттеушілер GO терминдерін тиімді болжау үшін әртүрлі тәсілдерді пайдаланады. Сонымен қатар, терең оқыту, деректерге негізделген тәсілдегі жылдам дамып келе жатқан пән, аминқышқылдарының тізбегіне GO терминдерін тағайындаудың әсерлі әлеуетін көрсетеді. Бұл мақалада авторлар дәстүрліден терең оқыту әдісіне дейінгі протеиндерге арналған GO аннотациясының қазіргі уақытта қол жетімді есептеу әдістерін қарастырады. Ол сондай-ақ осы саладағы негізгі проблемаларды талқылайды және GO көмегімен ақуыз функцияларын болжаудың болашақ бағыттарын көрсетеді.

Түйін сөздер: Gene Ontology, ақуыз функциясын болжау, машиналық оқыту, терең оқыту, ақуыз аннотациясы.

Y.S. Golenko[1*], A.A. Ismailova[1], A.S. Zhumakhanova[2]

[1]S. Seifullin Kazakh Agrotechnical University, Nur-Sultan, Kazakhstan;
[2]L.N. Gumilyov Eurasian National University, Nur-Sultan, Kazakhstan.
E-mail: *golenko.katerina@gmail.com*

PREDICTING PROTEIN FUNCTIONS USING THE «GENE ONTOLOGY» DATABASE AND MACHINE LEARNING MODELS

Abstract. Protein function prediction is an important part of genome annotation. Recently, predictive methods have advanced rapidly due to the advent of high throughput sequencing technologies. Among the available databases for defining protein function terms, Gene Ontology (GO) is an important resource describing the functional properties of proteins. Gene Ontology proposes a classification of functions that is based on a list or vocabulary of well-defined terms, each of which belongs to one of the categories - molecular functions, biological processes, and cellular components. From this database, you can use the name of the protein or its identification number to find the Gene Ontology terms assigned to it or annotations that were made on the basis of calculated or experimental data. Each term in Gene Ontology is assigned a number of attributes: a unique numeric identifier, a name, the dictionary to which the term belongs, and a definition. Terms can have synonyms, which are divided into exactly

corresponding to the meaning of the term, broader, narrower, and having some relation to the term. Researchers use a variety of approaches to efficiently predict GO terms. Meanwhile, deep learning, a rapidly evolving discipline in a data-driven approach, is showing impressive potential for assigning GO terms to amino acid sequences. In this article, the authors consider the currently available computational methods for GO annotation for proteins, ranging from the traditional one to the deep learning method. It also discusses the main problems in this area and highlights future directions for predicting protein functions using GO.

Key words: Gene Ontology, protein function prediction, machine learning, deep learning, protein annotation.

Введение. Белки представляют собой органические макромолекулы, которые являются фундаментальными детерминантами структуры и функций живых организмов. Они играют роль во многих процессах, включая биохимические реакции, передачу сигналов, транспорт питательных веществ и т.д. Поэтому понимание свойств белков важно не только с биологической и эволюционной точек зрения, но и с точки зрения использования их потенциала в биомедицинских и фармацевтических сферах и других областях.

Как правило, идентификация функций белка осуществляется с помощью ручной или компьютерной аннотации. Первый подход является золотым стандартом для функциональных аннотаций, потому что он реализуется экспертами и дает высококачественные результаты. Тем не менее, этот подход является дорогостоящим и трудоемким, и поэтому его трудно масштабировать. Кроме того, из-за развития высокопроизводительных технологий секвенирования, таких как секвенирование нового поколения (Next-Generation Sequencing или NGS), количество последовательностей, подлежащих аннотированию, резко возросло. Таким образом, методы вычислительной аннотации были разработаны как необходимость для автоматической обработки большого объема вновь сгенерированных последовательностей, а также для повышения точности аннотированных данных.

Из-за изменчивости словаря, используемого для определения функций белка, были предложены различные базы данных для обеспечения стандартизированной схемы, такие как Enzyme Commission (EC), Функциональный каталог (FunCat) и Киотская энциклопедия генов и геномов (KEGG). В настоящее время Gene Ontology (GO) (Consortium GO, 2015) является наиболее полным ресурсом, поскольку он обладает всеми нужными свойствами системы функциональной классификации.

Консорциум GO создал базу данных для контролируемого словаря, описывающего функциональные свойства геномных продуктов (например, генов, белков и РНК). Каждая онтология (словарь) принадлежит к одной из трех категорий: молекулярная функция (МФ, MF, Molecular Function), биологический процесс (БП, BP, Biological Process) и клеточный компонент (CC, Cellular Component). С точки зрения структуры, GO следует иерархической организации в виде направленного ациклического графа (DAG), в котором каждый термин является узлом, а каждое ребро, соединенное с двумя узлами, представляет отношение родитель-потомок. Его можно использовать для вывода многих типов информации, таких как «является» («is-a») или «является частью» («part-of»). Кроме того, «является частью» подразумевает, что дочерний узел обязательно является частью родителя. Это позволяет гибко аннотировать белки в отношении различных уровней функции — от общих терминов до конкретных – в зависимости от доступных данных.

Автоматизированное прогнозирование функций (Automated Function Prediction - AFP) на основе системы GO представляет собой сложную задачу биоинформатики. Во многих исследованиях обсуждалась функциональная аннотация белка с разных точек зрения. Предыдущие обзоры были посвящены AFP (Rost, 2003) с точки зрения используемого типа данных (Shehu, 2016), недостатков и соответствующих решений, сети взаимодействия белков, типов классифицированных функций и назначения GO на основе информации о последовательности. В (Bonetta, 2020) демонстрируется предсказание функции белка в рабочем процессе машинного обучения. В (Zhao, 2022) рассматривают прогнозирование функции генов с точки зрения моделирования GO. Все эти исследования представили независимые взгляды на проблему, однако не представлено подробного обзора глубинного обучения, которое является новым подходом к прогнозированию функций белков с помощью терминов GO. Таким образом, в статье предлагается возможность прогнозирования функций белка с использованием как обычного, так и глубинного обучения, что также указывает на то, что можно ожидать более высокой прогностической эффективности при сравнении нескольких методов друг с другом.

Основываясь на предыдущих исследованиях и вышеупомянутых обзорах, наша статья разделена на две основные части. Первая часть охватывает традиционный подход и включает решения, не использующие глубинное обучение, а вторая часть описывает методы, кото-

рые основываются на глубинном обучении для решения проблемы функциональной аннотации белка.

Ниже кратко рассмотрен традиционный подход, так как по этому вопросу уже доступно большое количество работ и обзоров литературы. В этой части суммируются три основные подкатегории традиционного подхода и упоминаются известные и самые последние соответствующие исследования.

Основная часть. Материалы и методы. Стандартный подход для прогнозирования GO терминов белкам.

В целом, было предложено несколько методов для присвоения терминов GO белкам с использованием различных типов данных и методов. Ниже рассмотрены репрезентативные решения, внимание сосредоточено на трех категориях: методы на основе подобия, вероятностные методы и методы машинного обучения.

Методы, основанные на сходстве. Первоначально функциональные аннотации присваивались неохарактеризованным белкам на основе простого принципа: последовательность белка искалась в базах данных экспериментально выбранных белков и, если были извлечены какие-либо похожие белки («похожие» специально определенным образом), термины GO, связанные с извлеченными последовательностями, были присвоены запрашиваемому белку. Самое первое решение методов, основанных на сходстве, было основано на гомологии. Эти инструменты основывались на инструментах поиска локальных выравниваний, таких как Базовый инструмент поиска локальных выравниваний (BLAST). В этих методах неизвестная последовательность ищется в базе данных, в которой хранятся хорошо аннотированные белки. Затем идентифицируется извлеченная последовательность с наивысшей оценкой выравнивания в соответствии с заданным порогом, и ее аннотация передается запрашивающему белку. OntoBlast, GOFigure, GOblet и GOtcha являются типичными системами аннотаций, использующими сходство последовательностей, определяемое поисковым механизмом BLAST. Более подробная информация об этих инструментах представлена в Таблице 1.

Таблица 1 Предикторы аннотаций GO, использующие традиционный подход

Метод	**Назавание программы**	**Описание**	**Год исследования**
Основанный на сходстве	OntoBlast	Веб-сервер, который является частью инструмента «Ontologies TO GenomeMatrix». Он предсказывает функцию белка по взвешенному списку терминов GO, связанных с последовательностями попаданий BLAST в девяти базах данных генома.	2003
	GOFigure	Производит вывод в виде кликабельного графа в четыре этапа, включая поиск гомологичной последовательности, построение графа с минимальным покрытием и назначение онтологий после их оценки.	2003
	Goblet	Программный пакет, позволяющий пользователю определять чувствительность к E-значению, и базы данных, используемые для поиска BLAST. Все онтологии, идентифицированные после сопоставления белков, строятся в сводной структуре DAG, где количество последовательностей, имеющих общий термин GO, является кумулятивным, чтобы представить значимость термина.	2004
	Gotcha	Переносит ассоциации GO, полученные из BLAST, в генные продукты с помощью новой схемы ранжирования. E-значения производных терминов распространяются на их родительские термины в DAG, обеспечивая нормализованные оценки достоверности для отдельной онтологии GO.	2009
	PFP	Функциональная система аннотаций, основанная на совпадениях PSI-BLAST, использующая совпадающие последовательности до необычно высокого порога и извлекающая очень важные термины GO из UniProt для оценки функций для неизвестных белков.	2015
	INGA	Функциональный сервер аннотаций выводит прогнозы GO из множества источников данных.	2016
	GoFDR	Метод на основе множественного выравнивания с использованием FDR и PSSM для ранжирования терминов GO для аминокислотных последовательностей.	2003

Вероятностный	-	Вероятностная структура, основанная на PPI, прогнозирует назначение GO для 10% неаннотированных белков в дрожжах.	2007
	-	Использует разнородные данные для улучшения вероятностной модели функционального прогнозирования Saccharomyces cerevisiae.	2010
	BMRF	Вариант подхода на основе MRF к PPI, который позволяет делать новые прогнозы для неаннотированных белков.	2006
Машинное обучение	GOPET	Инструмент прогнозирования и оценки терминов GO, предоставляющий термины GO для молекулярных функций и биологических процессов для любого организма.	2010
	PoGO	Модель ансамбля со сходством последовательностей InterPro, биохимическими свойствами и третичной структурой белка для прогнозирования GO белков грибов.	2018
	FFPred3	Инструмент на основе SVM, обеспечивающий независимое от гомологии назначение терминов GO для эукариотических белков.	2018
	PANNZER2	Взвешенный классификатор k-ближайших соседей, обеспечивающий функциональную аннотацию для неохарактеризованных белков с терминами GO и описаниями в произвольном тексте.	2018
	Deep Text2GO	Сочетание текстового метода и метода, основанного на последовательностях, для улучшения крупномасштабного предсказания функции белка.	2018
	NetGo	Эффективный веб-сервер на основе LTR, объединяющий как последовательность, так и массивную сетевую информацию о белках для аннотирования генных продуктов.	2019

PFP (Piovesan, 2015) это еще один метод, который использует функциональную информацию, связанную с удаленными гомологами, путем применения позиционно-специфического итерированного BLAST (PSIBLAST). Инструмент является улучшением с точки зрения охвата и точности, что определяется путем анализа эталонного набора данных по сравнению с GOtcha, PSI-BLAST и InterProScan. Другим вариантом решения, основанным на выравнивании последовательностей, является INGA (Piovesan, 2015), в котором данные сети белок-белкового взаимодействия (Protein-Protein Interaction - PPI) объединяются с назначением домена и сходством последовательностей

из BLAST для достижения консенсуса в прогнозировании функций GO с использованием анализа обогащения. Кроме того, GoFDR (Gong, 2016) берет соответствующие термины GO из запросов множественного выравнивания последовательностей (MSA) с помощью поиска BLAST или PSI-BLAST. Вероятность присвоения термина последовательности запроса определяется функционально различающими остатками (Functionally Discriminating Residues - FDR), оценочной матрицей для конкретных позиций (PSSM) для FDR и таблицей оценки с максимальной вероятностью, подготовленной с использованием обучающих последовательностей.

Хотя методы, основанные на поиске локального выравнивания, все еще актуальны, просты и в некоторой степени работают хорошо, они также имеют некоторые недостатки, включая ошибки аннотаций базы данных или чрезмерную передачу функций, относительность порога и низкую чувствительность или специфичность. В дополнение к методам, основанным на выравнивании, доступно несколько других предикторов, которые аннотируют функции переноса на основе сходства на уровне структуры белка, семейства белков или филогеномики (Shehu, 2016).

Вероятностные методы. Для определения функции белков был разработан ряд вероятностных моделей. В (Nariai, 2007) использовали граф функциональных связей, построенный на основе сети PPI дрожжей Saccharomyces cerevisiae. Предположение заключалось в том, что вероятность совместного использования функций между (узлами) белками, находящимися близко друг к другу на графе, выше, чем вероятность для узлов, которые не находятся в непосредственной близости. В этом методе вероятности терминов GO, присвоенных белковой последовательности, выводятся из биномиальной модели, включающей алгоритм случайного поля Маркова (MRF). Позже, в (Nariai, 2007) объединили несколько источников данных (PPI, данные об экспрессии генов, информацию о белковых мотивах, данные о мутантном фенотипе и данные о локализации белка), используя байесовскую структуру, чтобы повысить эффективность прогнозирования по сравнению с моделью, которая использует только PPI.

Работая с тем же модельным организмом, S. cerevisiae, который использовался в описанных выше методах, в (Kourmpetis, 2010) было предложено применить байесовский подход к модели MRF. Это предложение считалось улучшением для оценки параметров модели и предоставления прогнозов на основе сетевых данных. Кроме того, Pinoli, Chicco и Masseroli (Pinoli, 2015) сравнили различные

схемы взвешивания в сочетании с тремя алгоритмами, а именно: разложение усеченного сингулярного значения (truncated Singular Value Decomposition - tSVD), семантически улучшенное tSVD (SIM) и вероятностный латентно-семантический анализ с нормализацией (probabilistic Latent Semantic Analysis with normalization - pLSAnorm). Первые два метода основаны на линейной алгебре, а последний представляет собой модифицированную вероятностную модель, основанную на байесовском выводе. Эти методы были успешно использованы для создания новых аннотаций GO для трех модельных организмов: Bos taurus, Danio rerio и Drosophila melanogaster.

Методы машинного обучения. Инструменты на основе машинного обучения были разработаны для выявления скрытых взаимосвязей между различными характеристиками белка (последовательностью, структурой или другими соответствующими эволюционными доказательствами) и функциональными метками на основе обучающего набора (группы полностью охарактеризованных макромолекул) и использования этой информации для создания аннотаций для новых белков. Были предложены различные методы, основанные на машинном обучении, включающие различные линии доказательств в качестве функций для обучения классификаторов и прогнозирования терминов GO.

Использование классификаторов множественной машины опорных векторов (Support Vector Machine - SVM) является преобладающим выбором в нескольких исследованиях. Например, GOPET передает несколько функций, связанных с терминами GO (меры сходства последовательностей, основанные на поиске BLAST, частоте, качестве аннотаций гомологов и уровне аннотаций в иерархии GO) для 99 классификаторов SVM. Конкретный термин будет «правильным» или «неправильным» ярлыком для неизвестной последовательности, с вычисленными показателями достоверности голосования. В качестве другого примера, FFPred первоначально был установлен для неаннотированного протеома человека, но продемонстрировал обобщение на другие протеомы. Последняя версия FFPred – FFPred3, которая по-прежнему основана на SVM, расширена для исследования корреляций между характеристиками признаков, извлеченными из последовательностей и структур в рамках трех подонтологий (MF, BP и CC). Другой инструмент, Prediction of Gene Ontology terms (PoGO), был разработан на основе автоматической аннотации функционального класса белка (Automatic Annotation of Protein Functional Class - AAPFC). Вместо того, чтобы использовать только

термины InterPro в качестве характеристик, PoGO объединяет еще три источника (сходство последовательностей, биохимические свойства и третичная структура белка). Впоследствии SVM и линейный классификатор используются в качестве классификаторов базового уровня перед этапом метаобучения.

Используя алгоритм k-ближайшего соседа (k-Nearest Neighbor - k-NN), PANNZER2 предоставляет быстрые функциональные системы аннотаций, основанные на гомологии последовательностей и других предикторах аннотаций. Между тем, MS-k NN объединяет разнородные данные, чтобы предложить конкурентоспособную модель для предсказания функции белка. DeepText2GO – это согласованный подход, объединяющий глубокое семантическое представление текста из цитат MEDLINE (NCB, 2018) в виде текстовой информации и информации о последовательности, полученной с помощью BLAST и InterProScan. Эти функции передаются в модели k-NN и логистической регрессии для создания функций белков без знаний в большом масштабе. NetGO является расширением GoLabeler, в котором используется модель обучения для ранжирования (Learning-To-Rank - LTR) для интеграции доказательств на основе последовательностей. NetGO повышает производительность крупномасштабного AFP за счет доступа к огромной белок-белок сети из более чем 2000 видов в базе данных STRING.

Подходы машинного обучения считаются будущим направлением для AFP. Некоторые новые белки либо не имеют идентифицируемых гомологичных последовательностей, либо их обнаруживаемым гомологам не были присвоены какие-либо GO-метки. Следовательно, определение функции белка с нуля с использованием машинного обучения, то есть прямого вывода аннотации из аминокислотной последовательности без доступа к каким-либо дополнительным ссылкам или базам данных, является актуальной задачей.

Аннотации терминов белков с использованием подхода глубинного обучения

Потенциал глубинного обучения был продемонстрирован в нескольких областях применения, включая биоинформатику (Li, 2019). Основной характеристикой, которая отличает этот подход от других методов, является процесс обучения. Они автоматически извлекают функции высокого уровня из необработанных данных и обеспечивают комплексные прогнозы. В отличие от обычных моделей машинного обучения, точная классификация достигается с помощью созданных вручную функций. В настоящее время количество генерируемых

геномных данных, а также количество сложных алгоритмов и вычислительных ресурсов быстро растут. Эти ресурсы поддерживают глубинное обучение для решения проблемы функциональных аннотаций. Ниже рассмотрены предлагаемые методы, основанные на двух критериях: модели и используемых входных данных. Сводная таблица рассмотренных методов представлена в Таблице 2.

Таблица 2 Методы, основанные на глубинном обучении, для назначения терминов GO белкам

Используемая функция	**Название программы**	**Описание**	**Исполь-зуемая модель DL**
Основанный на последо-вательности		Метод глубокого AE для прогнозирования термина GO	AE
		Выводит отсутствующие аннотации GO для белков Homo sapiens, S. cerevisiae, Mus musculus и Drosophila с использованием машин Больцмана с глубоким ограничением	DRBM
	ProLanGO	Модель на основе LSTM, использующая идею NMT для функциональной аннотации аминокислотных последовательностей.	LSTM
	SECLEF	Генерирует функцию GO и семейство белков UniProt из данных последовательности. Основные модели были разработаны Szalkai & Grolmusz (2018a). Веб-интерфейс доступен по адресу https://pitgroup.org/seclaf/.	CNN
	DEEPred	Прогнозирование функции белка на основе GO с набором связанных MTDNN. Исходный код и данные доступны по адресу https://github.com/cansyl/DEEPred.	DNN
	DeepSeq	Предсказывает функцию белка H. sapiens только на основе данных о последовательности, используя архитектуру глубокого обучения на основе CNN. Исходный код и данные доступны на https://github.com/recluze/deepseq.	CNN
	DeepGOPlus	Расширенный метод DeepGO, основанный на структуре нескольких слоев CNN. Доступно на http://deepgoplus.bio2vec.net/.	CNN

Интеграция на основе данных/ на основе структуры		Глубинная модель CNN, классифицирующая функциональность на основе третичной структуры белков человека.	CNN
	DeepGO	Фреймворк глубинного обучения, использующий один сверточный слой. Аннотации Protein GO выводятся на основе иерархически структурированного классификатора. Онлайн-инструмент доступен по адресу https://deepgo.cbrc.kaust.edu.sa/deepgo/.	CNN
	deepNF	Глубинное слияние сетей, фиксирует высокоуровневые функции нескольких сетевых данных для AFP. Исходный код и данные доступны на https://github.com/VGligorijevic/deepNF.	AE
		Реализация двух сложенных многослойных структур для прогнозирования терминов GO на основе свойств последовательности и структуры. Исходный код и данные доступны по адресу http://bioinf.cs.ucl.ac.uk/downloads/mtdnn.	DNN
	DeepFunc	Выводит аннотации GO путем объединения функций сети на основе InterPro и взаимодействия белков.	FCDN
	DeepGOA	Интегрирует исчерпывающую информацию о последовательности и PPI для AFP.	Bi-LSTM, CNN
	SDN2GO	Интегрирует архитектуры с тремя подмоделями и классификатором взвешивания для прогнозирования терминов GO. Исходный код и данные доступны на https://github.com/Charrick/SDN2GO.	CNN
	DeepAdd	Структура на основе CNN, предсказывающая функцию белка на основе последовательности и дополнительной информации (PPI или SSP).	CNN
	FFPred-GAN	FFPred, в котором используется обучающая выборка, дополненная архитектурой глубокого обучения.	GAN
	GONET	Глубинная модель для прогнозирования функции белка с использованием обучения представлению для встраивания белковых последовательностей и сетей.	CNN, RNN

Модельный подход. Обучение с учителем. Обучение с учителем является важным подходом для прогнозирования функции белков in silico. Модели под наблюдением учатся аннотировать функции, руководствуясь обучающим набором данных, который пред-

ставляет собой группу охарактеризованных белков с надежными, экспериментально подтвержденными аннотациями. После этапа обучения создается модель, которая фиксирует взаимосвязь между признаком и функцией и используется для прогнозирования терминов GO для новых аминокислотных последовательностей.

Глубокая нейронная сеть (Deep Neural Network - DNN) представляет собой архитектуру с прямой связью. Обычно она состоит из входного слоя, нескольких скрытых слоев и выходного слоя. Входные данные обрабатываются однонаправленно по слоям, от первого до конечного этапа. Что касается функциональной аннотации на основе GO, DNN использовались в качестве многозадачных DNN (MTDNN) (Fa, 2018) или серии MTDNN (Rifaioglu, 2019).

Сверточная нейронная сеть (Convolutional Neural Network - CNN) изначально была разработана для обработки двумерных изображений для распознавания рукописных цифр. Однако с тех пор CNN стала эффективной архитектурой не только для многомерных данных, но и для одномерных входных данных, таких как предложения или геномные последовательности. Модель CNN начинается со сверточных слоев, единицами которых являются карты объектов. Каждая единица получается путем вычисления операций свертки между локальными участками единицы в предыдущем слое и фильтром (набором весов). Слои нелинейного пула объединяются в сверточные слои, что позволяет интегрировать функции разного масштаба в эту архитектуру. CNN могут применяться в AFP либо отдельно (Du, 2020), либо в сочетании с другими архитектурами (Spalević, 2020).

Рекуррентная нейронная сеть (Recurrent Neural Network - RNN) – это архитектура глубокого обучения, разработанная специально для последовательных данных. RNN имеют магистраль DNN, блоки скрытого уровня связаны между собой. Таким образом, скрытая единица получает информацию от входного слоя на текущем временном шаге, а также от скрытых единиц на предыдущем этапе. Долгосрочная кратковременная память (Long Short-Term Memory - LSTM) – это вариант модели RNN, разработанный для фиксации долгосрочных зависимостей между входными последовательностями. Двунаправленный LSTM (Bi-LSTM) – это модель LSTM, которая обрабатывает данные в двух направлениях, вперед и назад. Эти нейронные сети используются для обработки белковых последовательностей на естественном языке или в сочетании с моделью CNN для предоставления аннотаций GO.

Другой тип нейронной сети, полносвязная глубокая сеть (Fully

Connected Deep Network - FCDN), представляет собой серию полностью связанных слоев. Эта модель использовалась для преобразования входных векторов из InterPro и прогнозирования терминов GO.

Обучение без учителя. В отличие от обучения с учителем, обучение без учителя независимо выявляет скрытые закономерности в распределении входных данных. Это имеет решающее значение для изучения немаркированных данных и предоставляет важную информацию для контролируемых архитектур. Как правило, модели обучения без учителя используются для кластеризации, уменьшения размеров и преобразования данных.

Autoencoder (AE) – это модель обучения без учителя, разработанная для присвоения терминов GO аминокислотным последовательностям. AE – это нейронная сеть для преобразования данных. После кодирования и декодирования целевой размер вывода совпадает с размером ввода. Используя обученную модель AE, прогнозы GO для белков можно вывести непосредственно из сгенерированной матрицы. В качестве альтернативы низкоразмерные представления признаков в скрытом слое извлекаются и передаются в классификатор SVM для окончательной классификации или в модель CNN для окончательной классификации.

Ограниченная машина Больцмана (Restricted Boltzmann Machine - RBM) имеет один скрытый слой для представления скрытых признаков и входной слой, кодирующий наблюдаемые данные. Для аннотации GO единицами входного слоя являются доступные термины GO. Zou, Wang и Yu (Zou, 2017) представили глубокий RBM (DRBM), в котором многослойные RBM обучаются и разворачиваются, что приводит к прогнозированию функции белка.

Наконец, порождающая состязательная сеть (Generative Adversarial Network - GAN) состоит из двух сетей: порождающей модели, которая создает синтетические реалистичные данные, и дискриминационной модели, которая оценивает, являются ли данные реальными или нет. В AFP модель GAN улучшает прогноз, создавая синтетические функции для контролируемых классификаторов или используя онтологические корреляции (Seyyedsalehi, 2021).

Подход на основе данных. Что касается входных данных, методы, основанные на глубинном обучении, в основном используют два подхода для присвоения терминов GO неизвестным генным продуктам. Одним из них является подход только к последовательности, который полезен для предсказания функций новых белков в отсутствие гомологичной информации или других ссылок. Второй основан на струк-

туре или иным образом использует большие данные из нескольких доступных ресурсов.

Модели на основе последовательностей (результаты и обсуждение). Одним из первых методов прогнозирования аннотаций GO, основанных на глубинном обучении, является метод, предложенный в (Chicco, 2014). Авторы сравнили два решения для аннотаций, tSVD и нейронную сеть AE. Форма вывода двух методов была одинаковой; однако последний показал лучшие результаты на шести разных наборах данных. Позже, вместо использования архитектуры AE, Zou, Wang и Yu (Zou, 2017) предложили DRBM для аннотирования продуктов генов четырех модельных видов: Homo sapiens, S.cerevisiae, Musmusculus и Drosophila.

ProLanGo – это первый инструмент, который применяет нейронный машинный перевод (Neural Machine Translation - NMT), разработанный Google в AFP. Аминокислотные последовательности и термины GO были преобразованы в языки «ProLan» и «GOLan» соответственно. Затем после трансляции с использованием модели с тремя слоями RNN была сгенерирована аннотация GO нового белка. DeepSeq — еще один метод присвоения терминов GO аминокислотным последовательностям, основанный на сверточных слоях. Однако авторы предсказали только пять наиболее частых онтологий MF для белков H. sapiens.

Используя архитектуру DNN, DEEPred (Rifaioglu, 2019) реализует стек многозадачных DNN с прямой связью. Каждая отдельная сеть построена для определенного уровня терминов GO в DAG, что позволяет выполнять иерархическую постобработку прогнозов. Белковые последовательности экспериментально представлены тремя типами дескрипторов (карта профиля подпоследовательности, состав псевдоаминокислот и функция объединенной триады), при этом карта профиля подпоследовательности лучше всего подходит для анализа.

DeepGOPlus был разработан для преодоления существующих ограничений DeepGO, таких как длина последовательности, недоступные функции PPI и количество меток GO. Функциональная аннотация предсказывается многослойной структурой CNN в сочетании со сходством последовательностей. TALE был разработан для создания прогнозов GO путем интеграции шаблонов последовательности на основе кодировщика преобразователя и совместного сходства термина последовательности. PFP-WGAN (Seyyedsalehi, 2021) – одна из последних идей, использующих GAN для определения функциональных возможностей белков. В то время как сеть генератора

обрабатывает необработанные последовательности, дискриминатор получает два входа; один состоит из аннотированных белков из базы данных SwissProt, а другой – из необработанных последовательностей с соответствующими аннотациями, синтезированными из генератора.

Интегрированные модели на основе данных/структуры. Функциональное назначение аминокислотных последовательностей трехмерных структур было предложено в (Tavanaei, 2016). В основном анализе использовалась модель CNN, которая была обучена и протестирована на пяти наборах данных белков человека, каждому из которых было присвоено два термина GO. Однако прогноз не распространялся на дерево DAG для унаследованных терминов GO. DeepGO – это функциональный сервер аннотаций, основанный на архитектуре CNN, который был разработан для изучения особенностей аминокислотных последовательностей и сети PPI. Метки GO назначаются входным белкам посредством иерархической классификации, структурированной в виде дерева DAG. DeepAdd был вдохновлен сервером DeepGO и предоставляет решение для AFP, используя структуру CNN для изучения векторных представлений из последовательностей и дополнительной информации. Кроме того, в DeepAdd добавляется профиль белковой последовательности (SSP), если информация о сети PPI недоступна. На основе аналогичной концепции объединения первичной структуры белка и PPI была построена новая модель GONET с использованием CNN, RNN и Уровня внимания (Attention layer) для последовательностей человека и мыши.

Работая с набором данных, идентичным тому, который был собран FFPred3, Fa et al. (Fa, 2018) протестировали MTDNN, в которой предсказание функции белка рассматривалось как проблема классификации с несколькими метками. Решение состояло из слоев, общих для всех задач (меток GO), которые сложены параллельно слоям, специфичным для задачи. DeepFunc – это новый предиктор, который превосходит DeepGO, FFPred3 и BLAST. Во-первых, информация о белковом домене, семействе и мотиве запрашивается у InterPro и кодируется перед прохождением через полностью связанные слои. Затем алгоритмом Deepwalk получаются топологические характеристики PPI. Наконец, этот метод объединяет два типа признаков (последовательный и сетевой ввод), чтобы соответствовать FCDN. Архитектура DeepGOA более сложная, чем у DeepFunc. В дополнение к информации, генерируемой InterPro и PPI, глобальные и локальные семантические признаки аминокислотных последовательностей извлекаются с помощью Bi-LSTM и сверточного слоя соответственно.

Используя те же типы функций, SDN2GO использует три подмодели для каждого источника информации, при этом все выходные данные интегрированы в окончательную взвешенную модель.

В другом исследовании deepNF был создан с помощью мультимодального глубокого AE для захвата скрытой информации в белках из различных типов сетей взаимодействия. DeepMNE-CNN имеет более высокую производительность, чем deepNF в человеческих данных, за счет использования слоев CNN вместо SVM для модели классификации. С другой стороны, FFPred-GAN (Wan, 2020) использует GAN для расширения возможностей традиционных моделей машинного обучения, особенно SVM. Реальные особенности — это биофизическая информация, извлеченная из необработанных аминокислотных последовательностей с помощью FFPred, а генератор использует скрытые переменные для увеличения синтетических образцов.

Выводы. Вычислительная GO-аннотация белков была активно решаемой и сложной задачей в биоинформатике примерно с 2000-х годов; это ответ на необходимость преодолеть разрыв между известными и неизвестными, недавно открытыми аминокислотными последовательностями.

С одной стороны, понимание функции белка необходимо для расшифровки биологической эволюции и для многочисленных задач, таких как разработка лекарств и лечение болезней. База данных GO облегчила всеобъемлющий словарь для функциональной аннотации, поскольку она представляет структурированные функции GO в трех доменах (MF, BP и CC), тем самым эффективно поддерживая назначение функций белков in silico. Здесь мы представили текущее состояние области и сравнили решения AFP на основе GO, которые классифицируются как традиционные подходы и подходы глубинного обучения.

Одной из основных трудностей в этой области является использование входных функций для достижения эффективной работы. Хотя гетерогенные входные данные полезны для предсказания GO, для большинства неаннотированных белков доступна только информация о последовательности. Что касается конкретных подходов, модели машинного обучения могут быть ограничены неоднородностью геномных данных при анализе различных источников информации, в то время как настройка параметров и гиперпараметров является сложным этапом подхода глубинного обучения. Что касается выходных данных, база данных GO обновляется, потому что аннотации GO все еще несбалансированы и не полны для всех видов. Его сложная структура идеально подходит для описания функциональных ролей

белков, но также делает задачу предсказания сложной задачей с несколькими метками.

Тем не менее, несмотря на существующие проблемы, поиск успешного аннотирования будет продолжаться во многих направлениях благодаря усилиям научного сообщества. Наряду с постоянным расширением баз данных -omics методы, основанные на данных, продемонстрировали превосходное применение в различных областях с многообещающей тенденцией в области функциональной аннотации. В зависимости от доступных источников они могут быть основаны только на интегрированных данных или информации о последовательности, учитывая гибридный подход. Были предложены и разработаны некоторые подходы для работы с несбалансированными данными; они включают работу над подгруппами, в которых классы более сбалансированы, увеличение данных с использованием GAN и рассмотрение показателей оценки, указанных для несбалансированных данных, таких как AUPR. Проблема с несколькими метками может быть решена за счет продвижения вычислительных ресурсов и четко определенных решений, например, путем объединения множества отдельных решений. Наконец, успешные решения для прогнозирования срока GO могут быть расширены за счет включения других функциональных ресурсов (EC, путей и т. д.), чтобы понять биологическую роль и потенциал белков в науке о жизни.

Information about authors:

Golenko Y.S. – Doctoral Student, S. Seifullin Kazakh Agrotechnical University, Nur-Sultan, Kazakhstan; *golenko.katerina@gmail.com*; https://orcid.org/0000-0002-4643-4571;

Ismailova A.A. – Ph.D., Associate Professor, S. Seifullin Kazakh Agrotechnical University, Nur-Sultan, Kazakhstan; *a.ismailova@mail.ru*; *https://orcid.org/0000-0002-8958-1846;*

Zhumakhanova A.S – Master's degree, L.N. Gumilyov Eurasian National University, Nur-Sultan, Kazakhstan; *zhumanar@gmail.com*; *https://orcid.org/0000-0003-4128-2107*;

REFERENCES

Bonetta R., Valentino G. (2020) Machine learning techniques for protein function prediction, Proteins: Structure, Function, and Bioinformatics, 88(3):397–413. DOI: 10.1002/prot.25832.

Chicco D., Sadowski P., Baldi P. (2014) Deep autoencoder neural networks for gene ontology annotation predictions, In: Proceedings of the 5th ACM conference on bioinformatics, computational biology, and health informatics, 533–540.

Consortium GO. (2015) Gene ontology consortium: going forward, Nucleic Acids Research, 43(D1):D1049–D1056. DOI: 10.1093/nar/gku1179.

Du Z., He Y., Li J., Uversky V.N. (2020) Deepadd: protein function prediction from k-merembedding and additional features, Computational Biology and Chemistry, 89:107379. DOI: 10.1016/j.compbiolchem.2020.107379.

Fa R., Cozzetto D., Wan C., Jones D.T. (2018) Predicting human protein function with multi-task deep neural networks, PLOS ONE 13(6):e0198216. DOI: 10.1371/journal.pone.0198216.

Gong Q., Ning W., Tian W. (2016) Gofdr: a sequence alignment-based method for predicting protein functions, Methods, 93:3–14. DOI: 10.1016/j.ymeth.2015.08.009.

Kourmpetis Y.A., Van Dijk A.D., Bink M.C., Van Ham R.C., ter Braak C.J. (2010) Bayesian markov random field analysis for protein function prediction based on network data, PLOS ONE, 5(2):e9293. DOI: 10.1371/journal.pone.0009293.

Li Y., Huang C., Ding L., Li Z., Pan Y., Gao X. (2019) Deep learning in bioinformatics: introduction, application, and perspective in the big data era, Methods, 166:4–21. DOI: 10.1016/j.ymeth.2019.04.008.

Nariai N., Kolaczyk E.D., Kasif S. (2007) Probabilistic protein function prediction from heterogeneous genome-wide data, PLOS ONE, 2(3):e337. DOI: 10.1371/journal.pone.0000337.

Pinoli P., Chicco D., Masseroli M. (2015) Computational algorithms to predict gene ontology annotations, BMC Bioinformatics, 16(6):1–15.

Piovesan D., Giollo M., Leonardi E., Ferrari C., Tosatto S.C. (2015) Inga: protein function prediction combining interaction networks, domain assignments and sequence similarity, Nucleic Acids Research, 43(W1):W134–W140. DOI: 10.1093/nar/gkv523.

Rifaioglu A.S., Doğan T., Martin M.J., Cetin-Atalay R., Atalay V. (2019) Deepred: automated protein function prediction with multi-task feed-forward deep neural networks, Scientific Reports, 9(1):1–16.

Rost B., Liu J., Nair R., Wrzeszczynski K.O., Ofran Y. (2003) Automatic prediction of protein function, Cellular and Molecular Life Sciences, 60(12):2637–2650. DOI: 10.1007/s00018-003-3114-8.

Seyyedsalehi S.F., Soleymani M., Rabiee H.R., Mofrad M.R. (2021) Pfp-wgan: protein function prediction by discovering gene ontology term correlations with generative adversarial networks, PLOS ONE, 16(2):e0244430. DOI:10.1371/journal.pone.0244430.

Shehu A., Barbará D., Molloy K. (2016) A survey of computational methods for protein function prediction, Cham, Switzerland: Springer International Publishing, 225–298.

Spalević S., Veličković P., Kovačević J, Nikolić M. (2020) Hierachial protein function prediction with tails-gnns, ArXiv preprint. arXiv:2007.12804.

Tavanaei A., Maida A.S., Kaniymattam A., Loganantharaj R. (2016) Towards recognition of protein function based on its structure using deep convolutional networks, In: 2016 IEEE international conference on bioinformatics and biomedicine (BIBM), Piscataway: IEEE, 145–149.

Wan C., Jones D.T. (2020) Protein function prediction is improved by creating synthetic feature samples with generative adversarial networks, Nature Machine Intelligence, 2(9):540–550. DOI: 10.1038/s42256-020-0222-1.

Zhao Y., Wang J., Chen J., Zhang X., Guo M., Yu G. (2020) A literature review of gene function prediction by modeling gene ontology, Frontiers in Genetics, 11:400. DOI: 10.3389/fgene.2020.00400.

Zou X., Wang G., Yu G. (2017) Protein function prediction using deep restricted Boltzmann machines, BioMed Research International, 2017:1729301.

МАЗМҰНЫ

СОДЕРЖАНИЕ

CONTENTS

Publication Ethics and Publication Malpracticein
the journals of the National Academy of Sciences of the Republic of Kazakhstan

For information on Ethics in publishing and Ethicalguidelines for journal publication see http://www.elsevier.com/publishingethics and http://www.elsevier.com/journal-authors/ethics.

Submission of an article to the National Academy of Sciences of the Republic of Kazakhstan implies that the described work has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis or as an electronic preprint, see http://www.elsevier.com/postingpolicy), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any otherlanguage, including electronically without the written consent of the copyright-holder. In particular, translations into English of papers already published in another language are not accepted.

No other forms of scientific misconduct are allowed, such as plagiarism, falsification, fraudulent data, incorrect interpretation of other works, incorrect citations, etc. The National Academy of Sciences of the Republic of Kazakhstan follows the Code of Conduct of the Committee on Publication Ethics (COPE), and follows the COPE Flowcharts for Resolving Cases of Suspected Misconduct (http://publicationethics.org/files/u2/New_Code.pdf). To verify originality, your article may be checked by the Cross Check originality detection service http://www.elsevier.com/editors/plagdetect.

The authors are obliged to participate in peer review process and be ready to provide corrections, clarifications, retractions and apologies when needed. All authors of a paper should have significantly contributed to the research.

The reviewers should provide objective judgments and should point out relevant published works which are not yet cited. Reviewed articles should be treated confidentially. The reviewers will be chosen in such a way that there is no conflict of interests with respect to the research, the authors and/or the research funders.

The editors have complete responsibility and authority to reject or accept a paper, and they will only accept a paper when reasonably certain. They will preserve anonymity of reviewers and promote publication of corrections, clarifications, retractions and apologies when needed. The acceptance of a paper automatically implies the copyright transfer to the National Academy of Sciences of the Republic of Kazakhstan.

The Editorial Board of the National Academy of Sciences of the Republic of Kazakhstan will monitor and safeguard publishing ethics.

Правила оформления статьи для публикации в журнале смотреть на сайтах:
www:nauka-nanrk.kz
http://physics-mathematics.kz/index.php/en/archive
ISSN2518-1726 (Online),
ISSN 1991-346X (Print)

Директор отдела издания научных журналов НАН РК *А. Ботанқызы*
Заместитель директор отдела издания научных журналов НАН РК *Р. Жәлиқызы*
Редакторы: *М.С. Ахметова, Д.С. Аленов*
Верстка на компьютере *Г.Д. Жадыранова*

Подписано в печать 29.06.2022.
Формат 60х881/8. Бумага офсетная. Печать – ризограф.
9,0 п.л. Тираж 300. Заказ 1.

Национальная академия наук РК
050010, Алматы, ул. Шевченко, 28, т. 272-13-19