

БИОИНФОРМАТИКА И КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ БИОЛОГИЧЕСКИХ СИСТЕМ

Ю.В. Черных¹, М.С. Маслов¹, Н.О. Майгур¹

¹ФГБОУ ВО «Воронежский государственный лесотехнический университет
имени Г.Ф. Морозова»

Аннотация. В работе рассматривается применение компьютерного моделирования в области биоинформатики, являющейся ответвлением математической биологии. Описываются виды биологических моделей, их преимущества и определенные ограничения в применении.

Ключевые слова: биоинформатика, компьютерное моделирование, математическая биология, исследование биологических систем

BIOINFORMATICS AND COMPUTER MODELING OF BIOLOGICAL SYSTEMS

Y.V. Chernykh¹, M.S. Maslov¹, N.O. Maygur¹

¹Voronezh State University of Forestry and Technologies named after G.F. Morozov

Abstract. This paper discusses the use of computer modeling in the field of bioinformatics, which is a branch of mathematical biology. The types of biological models, their advantages and certain limitations in application are described.

Key words: bioinformatics, computer modeling, mathematical biology, study of biological systems

Компьютерное моделирование является одним из главных источников и причин развития многих технологий. С его появлением в науке стало возможным рассмотреть даже невозможные для реализации в нашей жизни процессы (ядерный взрыв, распад частиц, процессы, требующие времени). Благодаря этой прекрасной возможности компьютерное моделирование нашло себе место в различных предметных областях таких как – экономика, астрофизика, математика, даже

в социология. Исключением не стала и наука о живой природе, т.к. в ней сложнее всего реализовать интересующие исследователей процессы. В данной статье будет рассмотрена связь биологии с, казалось бы, несовместимыми информатикой и моделированием.

Истоком компьютерного моделирования биологических систем стала математическая биология, а предпосылкой стали работы, принадлежащие Леонардо Фибоначчи. Итальянец описал задачу о размножении кроликов, тем самым построив математическую модель и начав путь развития математической биологии как таковой. Другой итальянский математик, Вито Вольтерра, разработал модель сосуществования двух популяций (хищник-жертва). Как итог данных трудов, две стороны – математики и биологи нашли в предметных областях друг друга интерес и потребность к систематизации науки, что подтверждает приход в биологию знаменитых отечественных ученых – Ляпунова, Полетаева. Благодаря ученику первого – Вадиму Александровичу Ратнеру, были описаны явления молекулярного уровня в сложных биологических системах (данные явления описаны в книге «Генетика, молекулярная кибернетика личности и проблемы»). Такие системы требуют производительности ввиду объема данных, поэтому дальнейшей целью была её наработка. С этого момента от математической биологии пошла отдельная ветвь, гордо носящая название биоинформатики – науки, применяющей алгоритмы для систематизации и анализа генетических данных, для определения молекулярных основ биологического процесса. Генетические данные, в первую очередь – белки и нуклеиновые кислоты. Подобное нововведение в науке позволило выявлять закономерности, связанные с генами и белками, зашифрованными в них, а также строить модели генов в клетке и даже разрабатывать лекарства, основываясь на данной информации.

На сегодняшний момент, главная цель биоинформатики – описание задач молекулярной биологии. Причиной этого служит удивительный объём имеющейся информации, необходимый для изучения функционирования молекул, что необходимо не только для прогнозирования генов в новом организме, но и для создания лекарственных препаратов.

Для полноценного понимания роли информатики в такой науке, как биоинформатика, следует рассмотреть все её направления.

Биоинформатика последовательностей – осуществляет анализ нуклеотидных и белковых последовательностей. Найти нужную последовательность из почти 160 миллионов символов невозможно без применения компьютера. Основными направлениями биоинформатики последовательностей являются:

- поиск особых участков, кодирующих белок, РНК;
- получение аминокислотных последовательностей из гена;
- сравнительный анализ двух последовательностей с целью найти отличия, прогнозировать вторичную структуру РНК.

Структурная биоинформатика – анализирует пространственные структуры (способ укладки последовательности аминокислот белка), выявленных экспериментом.

Данное направление особо важно для разработки лекарств, молекулы которых связываются и образуют комплексы с молекулами белков. Исследование этого механизма имеет ключевую роль, а все результаты – масштабные структуры комплексов для компьютерного анализа. Перечислим основные задачи:

- предсказание участков белковой молекулы (компьютерный анализ не может дать достоверную информацию, только примерное наблюдение, которое необходимо проверить экспериментально);
- классификация структур белков по пространственной структуре;
- предсказание воздействия молекул химических веществ (потенциального лекарства).

Компьютерная геномика. Прочтение генома – первый этап исследования функционирования клетки организмов, а без этого этапа не был бы исследован метаболизм бактерий (нельзя было бы подобрать лекарство), не создавались новые ферменты, которые участвуют в биотехнологическом производстве. В случае экспериментального исследования за длительное время и тщательную работу будет изучена лишь единица гена, тогда как при использовании компьютерного анализа (геномики) за неделю будут исследованы несколько тысяч генов с определенной точностью. Это облегчает проведение экспериментов, соответственно, и прогресс будет достигнут проще.

И в случае экспериментального исследования, лишь единица гена требует много ресурсов. Основные задачи компьютерная геномики:

- сравнительный анализ геномов;
- исследование регуляции функционирования генов;
- исследование транспортеров – генов, переносящих перенос питательных веществ;
- полногеномный анализ, исследование эволюции (перебор геномом вариантов последовательностей, что приводит к более эффективной эволюции вида).

Благодаря развитию биоинформатики для людей стали доступны многие возможности, например, открытие триклозана. Триклозан – универсальный распространённый антибактериальный препарат, используемого нами по сей день в мыле, санитайзерах и других антибактериальных средствах.

Конечно, ни одно направление в области биоинформатики не может обойтись без характерной информатики компонента – моделирования. Компьютерная модель – упрощённое математическое описание существующего объекта, выполненное на компьютере. В области биологии существуют ряд классических моделей, без которых невозможно оценить поведение популяции. Модель популяции (популяционная) – позволяет изучать динамику популяций живых организмов. Цель моделирования в таком случае – описание изменений численности, включая факторы рождаемости, смертности, миграции, отношений с другими видами.

Наиболее точными популяционными моделями являются:

Модель неограниченного роста (Т.Р. Мальтуса) – классическая математическая модель, предложенная Томасом Робертом Мальтусом. Ученый, заинтересовался идеей, что с ростом численности популяции падает рост ресурсов для питания, на основе человеческой популяции подытожил, что невозможно избежать нехватки ресурсов. Формула, описанная ученым, следующая:

$$x(i + 1) = (1 + a) \cdot x(i), \quad (1)$$

где a – коэффициент естественного прироста (отношение численности прироста за период изучения к начальной численности), $x(i + 1)$ – предсказываемая численность, $x(i)$ – нынешняя численность.

Модель ограниченного роста (П. Ф. Ферхюльста). Это более достоверная модель Мальтуса, учитывающая конкуренцию за ресурсы питания в популяции. Бельгийский математик, Пьер Франсуа Ферхюльст воздвигнул идею о том, что любая популяция достигает своего максимума, зависящего от факторов среды. Формула модели не сильно отличается от предыдущей:

$$x(i + 1) = x(i) + (a - b \cdot x(i)) \cdot x(i), \quad (2)$$

где b – коэффициент смертности от внутривидовой конкуренции.

Модель с критической численностью. Математическая модель ограниченного роста, учитывающая наименьшую критическую численность (существуют такие популяции, численность которых не опустится ниже критического уровня).

$$x(i + 1) = (a - b \cdot x(i)) \cdot (x(i) - L), \quad (3)$$

где L – критическая численность популяции.

Модель распространения болезней (эпидемиологическая) – позволяет оценить эффективность стратегий контроля и предотвращения распространения болезней. Модели должны учитывать следующие факторы: передача болезни, иммунитет, вакцинация и подобное. Для решения задач в данном направлении используется модель SIR. Население делится на три группы: здоровые (susceptible), зараженные (infectious), с иммунитетом, либо умершие (recovered). Модель предусматривает, как будет меняться количество людей в каждой группе в течение времени. Описывается системой дифференциальных уравнений и имеет следующий вид:

$$\begin{aligned} dS/dt &= -\beta IS, \\ dI/dt &= \beta IS - \gamma I, \\ dR/dt &= \gamma I, \end{aligned} \tag{4}$$

где β – скорость передачи болезни при контакте, γ – уровень излечения от болезни

Стоит учесть, что данная модель не рассматривает рождаемость или естественную смертность, поскольку считается, что период эпидемии длится меньше жизни человека. Однако, если исследователя это интересует, модель SIR можно модифицировать, так как это самая простая эпидемиологическая модель. Примерами разновидностей являются: SEIR, SIRS, SIS и другие, учитывающие факторы повторного заболевания и врожденного иммунитета.

Модель эволюции и генетики (эволюционная) – самая обширная модель, применяющая признаки теории Дарвина для построения систем, зачастую, для улучшения вычислительного интеллекта и описания социально-экономических систем.

Пример эволюционной модели – генетический алгоритм популяции, схема работы которого:

1. начальная популяция;
2. скрещивание/ мутация;
3. селекция;
4. формирование следующего поколения;
5. соответствие результатов (если нет – возврат к шагу 2);
6. итоговая популяция.

Модель эволюции помогает в понимании того, какие факторы могут привести к изменению генетических данных популяции и какие механизмы совершают работу в эволюционных изменениях.

Молекулярные модели – модели химических соединений, дающие представления о пространственном расположении атомов в соединениях в молекуле. Данные модели применяются в случае, когда невозможно проанализировать расположение атомов или оценить влияния атомов друг на друга. Простое соединение атомов же может быть исследовано без построения трехмерного или двухмерного изображения

В создании двухмерной модели своё участие принимают графы – абстрактные структуры из узлов и рёбер. Под молекулярным графом понимают атомы(узлы), объединенные между собой связями(ребра). Такой этап дает понятие о порядке связи между атомами в молекуле, но представления о пространстве не даёт. Это делают трехмерные модели, которые, в свою очередь, также помогают в исследовании стерических и электронных свойств молекулы.

Как можно заметить, моделей, используемых в биологии, достаточно количество. И это не просто так, ведь, у компьютерного моделирования есть особые преимущества перед традиционными методами изучения, в частности:

- экономия времени и ресурсов – моделирование позволяет проводить виртуальные эксперименты в короткие сроки;
- обучение – практическое изучение чего-либо материала на основе компьютерной модели осваивается студентами, школьниками и интернами намного лучше;
- прогнозирование – позволяет принять меры для предотвращения негативных последствий в зависимости от условий, в которую помещена виртуальная модель.

Но, как и у любого вида исследования и экспериментов, у компьютерного моделирования есть собственные ограничения и недостатки:

- недостаточная точность, т.к. в реальном мире гораздо больше факторов, влияющих на объект, которые в модели невозможно учесть, либо модели сознательно могут быть пропущены;
- недостаточный объём данных – надёжная модель требует достаточное количество данных, но бывает сложно собрать данные о всём существующем в изучаемой системе, что затрудняет анализ долгосрочных изменений системы;
- недостаточная универсальность модели – т.к. модель, построенная для определенного вида, не подойдет другому виду в силу ряда условий, например, среда, климат и другие важные факторы, что приводит к ограничению области применения той или иной модели;

- упрощение сложных частей системы – условием любой модели служит процесс декомпозиции чего-то сложного, описывать каждую особь в моделировании сложно и ненужно, поэтому виртуальная система воспринимает всех особей одинаково.

Таким образом, биоинформатика – ключ к пониманию биологических процессов и наука фокусируется на создании и исследовании процессов вычислительными методами, используя компьютерное моделирование для упрощения анализа и экспериментов.

В свою очередь, компьютерное моделирование – воздух в области биоинформатики. Оно позволяет использовать имитацию существующих процессов и явлений для анализа и исследования особенностей, прогнозирования недалекого будущего. Проведение виртуального тестирования систем помогает в экономии времени и ресурсов. Хранение огромных объемов информации упрощает работу, что также экономит важнейшие человеку ресурсы.

Список литературы

1. Биоинформатика - наука третьего тысячелетия (итоги Международной конференции по биоинформатике в Новосибирске, 14-20 июля 2002 Г.) // Информационный вестник ВОГиС. – 2003. – Т. 7, № 21-22. – С. 2. – EDN HRSXOP.

2. Гибадуллин, А. А. Молекулярная, эволюционная, генетическая, фрактальная биоинформатика и хемоинформатика / А. А. Гибадуллин // Академическая публицистика. – 2023. – № 9-1. – С. 34-36. – EDN DRXSER.

3. Ястребов, Д. С. Роль биоинформатики в дешифровке генетической информации / Д. С. Ястребов // Теория и практика современной науки. – 2023. – № 1(91). – С. 197-199. – EDN LKKRLE.

4. Савин, Е. И. Анализ показателей общественного здоровья и здравоохранения с точки зрения биоинформатики / Е. И. Савин, Т. В. Честнова, О. В. Гавриленко // Научное и образовательное пространство: перспективы развития : Сборник материалов III Международной научно-практической конференции: в 2-х томах, Чебоксары, 13 ноября 2016 года. Т. 1. – Чебоксары : Общество с ограниченной ответственностью "Центр научного сотрудничества "Интерактив плюс", 2016. – С. 36-38. – EDN XBIJFD.

5. Порозов, Ю. Б. Биоинформатика / Ю. Б. Порозов. – Санкт-Петербург : Университет ИТМО, 2012. – 54 с. – EDN ZUXZSD.

6. Колчанов, Н. А. Компьютерная системная биология и биоинформатика: задачи и методы / Н. А. Колчанов // Марчуковские научные чтения - 2019 : Тезисы Международной конференции, Новосибирск, 01–05 июля 2019 года. – Новосибирск: Новосибирский национальный исследовательский государственный университет, 2019. – С. 8-9. – EDN EDTFZM.

7. Гурьянова, С. В. Современные подходы системной биологии и биоинформатики в методологии диагностических исследований аллергических заболеваний / С. В. Гурьянова // Аллергология и иммунология. – 2017. – Т. 18, № 1. – С. 48. – EDN ZDNJKT.

References

1. Bioinformatics - the science of the third millennium (results of the International Conference on Bioinformatics in Novosibirsk, July 14-20, 2002) // Information Bulletin of VOGiS. – 2003. – Т. 7, No. 21-22. – P. 2. – EDN HRSXOP.

2. Gibadullin, A. A. Molecular, evolutionary, genetic, fractal bioinformatics and chemoinformatics / A. A. Gibadullin // Academic journalism. – 2023. – No. 9-1. – pp. 34-36. – EDN DRXSER.

3. Yastrebov, D. S. The role of bioinformatics in deciphering genetic information / D. S. Yastrebov // Theory and practice of modern science. – 2023. – No. 1(91). – pp. 197-199. – EDN LKKRLE.

4. Savin, E. I. Analysis of public health and healthcare indicators from the point of view of bioinformatics / E. I. Savin, T. V. Chestnova, O. V. Gavrilenko // Scientific and educational space: development prospects: Collection of materials of the III International Scientific- practical conference: in 2 volumes, Cheboksary, November 13, 2016. Volume 1. - Cheboksary: Limited Liability Company "Center for Scientific Cooperation "Interactive Plus", 2016. - P. 36-38. - EDN XBIJFD.

5. Porozov, Yu. B. Bioinformatics / Yu. B. Porozov. – St. Petersburg: ITMO University, 2012. – 54 p. – EDN ZUXZSD.

6. Kolchanov, N. A. Computer systems biology and bioinformatics: tasks and methods / N. A. Kolchanov // Marchukov Scientific Readings - 2019: Abstracts of the International Conference, Novosibirsk, July 01–05, 2019. – Novosibirsk: Novosibirsk National Research State University, 2019. – pp. 8-9. – EDN EDTFZM.

7. Guryanova, S. V. Modern approaches of systems biology and bioinformatics in the methodology of diagnostic studies of allergic diseases / S. V. Guryanova // Allergology and Immunology. – 2017. – Т. 18, No. 1. – P. 48. – EDN ZDNJKT.