

**ПРИМЕНЕНИЕ ДВУХКОМПОНЕНТНЫХ ИНДИВИДУУМ-
ОРИЕНТИРОВАННЫХ МОДЕЛЕЙ ДЛЯ АНАЛИЗА РАСПРОСТРАНЕНИЯ
СОЦИАЛЬНО ЗНАЧИМЫХ ЗАБОЛЕВАНИЙ****В.Н. Леоненко, Н.В. Перцев (Омск)****Введение**

Математическое моделирование является популярным и доказавшим на практике свою эффективность способом изучения распространения социально значимых заболеваний. Существует множество подходов к построению математических моделей (использование систем нелинейных дифференциальных и интегро-дифференциальных уравнений, систем уравнений в частных производных, цепей Маркова, случайных процессов, агентных моделей и пр.), каждый из которых имеет свои достоинства и недостатки. Для ряда задач выгодным является применение нескольких подходов в рамках одной двухкомпонентной модели, что позволяет более гибко описывать интересующие исследователя процессы в моделируемой системе и при необходимости уточнять описания отдельных компонент, не перестраивая модель. В России исследования, посвященные использованию двухкомпонентных моделей сложных систем, ведутся, в частности, сотрудниками ВЦ РАН (г. Москва) [1,2] и НИИ НКТ СПбГУ ИТМО [3]. Большое количество работ, посвященных двухкомпонентным моделям распространения заболеваний, публикуется за рубежом, например, [4,5].

В настоящей работе приводится подход к построению двухкомпонентных индивидуум-ориентированных моделей распространения заболеваний на основе совмещения популяционного и дискретно-событийного описаний. Представлен общий вид двухкомпонентной модели, а также методы распараллеливания алгоритмов моделирования для проведения численных экспериментов на многоядерных персональных компьютерах и высокопроизводительных вычислительных системах. Указанный подход применялся авторами для моделирования распространения туберкулеза [6] и ВИЧ [7], а также сравнения эффективности методов выявления индивидуумов, предрасположенных к колоректальному раку [8].

Популяционная модель

Основой для построения двухкомпонентной модели является популяционная модель распространения заболевания, задаваемая системой стохастических разностных уравнений с целочисленными неотрицательными переменными и дискретным временем. Система уравнений описывает динамику групп индивидуумов A_i , $i = 1, 2, \dots, n$, составляющих население некоторого региона. Единица времени выбирается исходя из длительности характерных процессов, происходящих с индивидуумом в рамках модели (от одних суток до одного года). Обозначим через $x_i(t)$ численность индивидуумов группы A_i в момент времени $t = 1, 2, \dots, T$, $i = 1, 2, \dots, n$. Система уравнений модели в общем виде выглядит следующим образом:

$$x_1(t) = \hat{x}_1(t) - \sum_{k=2,\dots,n} u_{1k}(t) + \sum_{l=2,\dots,n} u_{l1}(t) + f_1(t), \quad (1)$$

$$x_2(t) = \hat{x}_2(t) - \sum_{k=1,\dots,n,k \neq 2} u_{2k}(t) + \sum_{l=1,\dots,n,l \neq 2} u_{l2}(t) + f_2(t), \quad (2)$$

$$\dots\dots\dots$$

$$x_n(t) = \hat{x}_n(t) - \sum_{k=1,\dots,n-1} u_{nk}(t) + \sum_{l=2,\dots,n} u_{ln}(t) + f_n(t), \quad (3)$$

$$x_i(0) = x_i^{(0)}, \quad i = 1, 2, \dots, n; \quad t = 1, 2, \dots, T. \quad (4)$$

В системе уравнений (1)–(4) слагаемые $f_i(t) \geq 0, i = 1, 2, \dots, n$ отражают приток населения в соответствующие группы за промежуток времени $(t-1, t]$ и представляют собой некоторые случайные процессы с заданными вероятностными характеристиками. Случайные величины $\hat{x}_j(t)$ означают количество индивидуумов группы A_j , доживших от момента времени $t-1$ до момента времени t . При фиксированном $x_j(t-1) \geq 0$ величина $\hat{x}_j(t)$ имеет биномиальное распределение с параметрами $x_j(t-1), \rho_j$, где $0 < \rho_j < 1$ – вероятность дожития индивидуума группы A_j от момента времени $t-1$ до момента времени $t, j = 1, \dots, n$. Случайные величины $u_{jk}(t)$ описывают количество индивидуумов, переходящих из группы A_j в группу A_k в течение промежутка времени $(t-1, t]$. Для заданного j и фиксированного $\hat{x}_j(t)$ величины $u_{j1}(t), \dots, u_{jn}(t)$ являются координатами случайного вектора с мультиномиальным законом распределения. Параметрами этого закона являются вероятности $0 < \gamma_{jk} < 1$ переходов индивидуумов из группы A_j в группу A_k и количество доживших индивидуумов $\hat{x}_j(t)$. Вектор начальных численностей $(x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)})$ имеет неотрицательные компоненты и описывается некоторым заданным законом распределения.

Дискретно-событийная модель

Предполагаем далее, что из совокупности S выделяется подмножество $F \subseteq S$, состоящее из групп индивидуумов, подвергаемых воздействию заболевания. Каждый индивидуум α , принадлежащий группе $A_i \in F$, в заданный момент времени t характеризуется конкретным значением неотрицательного параметра $\Delta_\alpha(t)$. Это значение отражает степень воздействия заболевания на индивидуума и, в зависимости от задачи, выражается в конкретных величинах (размер злокачественного образования, концентрация вирусов в крови) или условных безразмерных единицах. Для индивидуума $\alpha \in A_i, A_i \in F$, перешедшего из группы $A_j \in S \setminus F$ за промежуток времени $(t_0 - 1, t_0]$, начальное значение параметра $\Delta_\alpha(t_0)$ задается константой или генерируется согласно некоторому закону распределения. Аналогично получают значения $\Delta_\alpha(t_0)$ для индивидуумов, поступивших в группы $A_i \in F$ вследствие миграции извне (случайные величины $f_i(t)$), а также значения $\Delta_\alpha(0)$ для индивидуумов, находившихся в группах $A_i \in F$ в начальный момент времени. Изменение значения $\Delta_\alpha(t)$ описывается некоторым детерминированным или стохастическим законом.

Поведение каждого индивидуума $\alpha \in A_i, A_i \in F$ моделируется независимо от других индивидуумов с помощью дискретно-событийного подхода [9]. В фиксированный момент времени индивидуум α находится в определенном состоянии, соответствующем группе $A_i \in F$. В течение каждого временного интервала $(t-1, t]$ состояние индивидуума может

изменяться в соответствии с заданным набором вероятностей. Эти вероятности могут зависеть от значения параметра $\Delta_\alpha(t)$. Дискретно-событийное моделирование изменения состояния индивидуума α прекращается в момент времени, когда он переходит в группу $A_j \in C \setminus F$ или покидает популяцию в силу гибели или миграции. Численности индивидуумов в каждом состоянии (группе) $A_i \in F$, $i = 1, \dots, n$ могут влиять на параметры законов распределения случайных величин $\hat{x}_j(t)$ и $u_{jk}(t)$ популяционной модели (1)–(4), описывающей динамику численностей индивидуумов из групп $A_i \in C \setminus F$.

Двухкомпонентная модель

Используя совместно представленные выше модели, приходим к двухкомпонентной индивидуум-ориентированной модели, в которой одна часть индивидуумов популяции неразличима внутри своих групп и описывается популяционной подмоделью, а другая часть моделируется с учетом значений индивидуальных параметров в рамках дискретно-событийной подмодели. Для иллюстрации взаимодействия компонентов модели приведем следующий простой пример. Опираясь на популяционную модель, вычислим количество $M(t_0)$ восприимчивых к инфекции индивидуумов, инфицированных за промежуток времени $(t_0 - 1, t_0]$. Каждый индивидуум из числа $M(t_0)$ пополняет группу инфицированных индивидуумов. Далее вычисления в популяционной модели приостанавливаются, а все вычисления проводятся для каждого отдельно взятого индивидуума из числа $M(t_0)$. Информация по таким индивидуумам заносится в память на моменты времени $t_0, t_0 + 1, t_0 + 2, \dots, T$. После завершения работы со всеми индивидуумами из числа $M(t_0)$ возобновляется работа популяционной модели и проводятся вычисления для промежутка времени $(t_0, t_0 + 1]$. Затем фиксируется количество $M(t_0 + 1)$ индивидуумов. Для них выполняются аналогичные действия и т.д.

Описанный подход позволяет варьировать детальность описания рассматриваемой популяции с учетом доступных вычислительных мощностей и реальных данных для калибровки модели.

Вычислительный алгоритм

Целью численных экспериментов с моделями является оценка различных характеристик групп индивидуумов (например, математических ожиданий, дисперсий, корреляций, совместных законов распределения численностей групп) с помощью стандартных методов математической статистики.

Численные эксперименты проводятся на основе методов Монте-Карло. Для генерации необходимых случайных величин применяются вычислительные методы и датчик псевдослучайных чисел, описанные в [10, 11]. Для нахождения выборочных значений характеристик групп индивидуумов используется набор реализаций моделируемого случайного процесса с разными стартовыми значениями датчика псевдослучайных чисел.

Отметим, что учет индивидуумов с динамически изменяемыми значениями параметров может приводить к существенным вычислительным затратам при изучении популяций, в которых количество индивидуумов достаточно велико (например, от 10^5 до 10^6 индивидуумов, что соответствует населению крупного региона или города). В таких случаях для ускорения работы моделирующей программы применяются параллельные и распределенные вычисления.

Распараллеливание по реализациям

Реализации моделируемого случайного процесса независимы друг от друга и не нуждаются в обмене данными между собой. В связи с этим простым и эффективным способом ускорения вычислений является применение параллельного алгоритма с распределением реализаций по доступным ядрам (для персонального компьютера) или вычислительным узлам (для вычислительного кластера). Результирующий алгоритм обладает высокой масштабируемостью и не зависит от деталей формализации конкретной модели в рамках подхода. Данный подход был реализован авторами с применением технологий OpenMP и MPI для проведения вычислительных экспериментов на суперкомпьютерах «Ломоносов» и BlueGene/P Суперкомпьютерного центра МГУ (г. Москва), а также НКС-30Т Сибирского суперкомпьютерного центра (г. Новосибирск).

Использование графических процессоров

Быстрое развитие технологий, связанных с вычислениями на графических процессорах (GPU), дало возможность использования GPU для проведения вычислительных экспериментов с двухкомпонентной моделью. В силу особенностей строения графических процессоров алгоритм для расчетов на GPU должен состоять из большого числа «легковесных» потоков, что делает невозможным применение в программах параллелизма по реализациям. На данный момент авторами реализовано два варианта алгоритмов с использованием GPU на основе технологии №VIDIA CUDA.

1. Алгоритмы для моделей неинфекционных заболеваний без учета демографических характеристик популяции. Рассмотрим класс моделей, для которых верны два предположения: (1) больные индивидуумы не влияют друг на друга и на восприимчивых индивидуумов (моделируется неинфекционное заболевание); (2) популяция в начальный момент времени состоит из индивидуумов одного возраста и предполагается отсутствие притока в популяцию в течение модельного времени. Для указанного класса моделей возможно применение алгоритма с распараллеливанием по индивидуумам, где возникновение и развитие заболевания у каждого происходит в отдельном потоке. Структура такого алгоритма оптимальна для применения вычислений на графических процессорах и согласно результатам вычислительных экспериментов дает весомый прирост производительности (вплоть до двух порядков) по сравнению с последовательной версией алгоритма.

2. Использование GPU для генерации псевдослучайных чисел. Для моделей, которые не удовлетворяют приведенным в п.1. требованиям, возможно применение алгоритмов с генерацией псевдослучайных чисел на графическом процессоре. Поскольку генерация псевдослучайных чисел является весьма ресурсоемкой операцией и занимает значительную часть времени, расходуемого на численный эксперимент, получаемая экономия времени существенно превышает временные затраты на пересылку псевдослучайных чисел с графического на центральный процессор. Численные эксперименты показали, что использование библиотеки CUDA CURAND для генерации псевдослучайных чисел на GPU дает прирост производительности моделирующей программы примерно на 50% по сравнению с использованием датчиков псевдослучайных чисел, выполняемых на центральном процессоре.

Заключение

Приведенный в работе подход к построению двухкомпонентных индивидуум-ориентированных моделей может быть применен для моделирования распространения различных социально значимых заболеваний с необходимой степенью детализации. Описанная выше двухкомпонентная структура модели может быть расширена с помощью до-

бавления дополнительных подмоделей, например, подмодели распространения инфекции в гетерогенной среде. Дальнейшее развитие двухкомпонентной модели предполагает реализацию комбинированного подхода к распараллеливанию алгоритма с использованием технологий OpenMP, MPI и CUDA. Такой подход позволит использовать вычислительные мощности гибридных кластеров и проводить вычислительные эксперименты с наиболее ресурсоемкими и высокодетализированными моделями распространения заболеваний.

Работа поддержана СО РАН (Интеграционный проект № 47, 2012–2014 гг.)

Литература

1. **Белотелов Н. В., Бродский Ю. И., Павловский Ю. Н.** Разработка инструментальной системы распределенного моделирования [Электронный ресурс]. URL: <http://simul.ccas.ru/articles/kirov.pdf>
2. **Бродский Ю. И., Павловский А. А.** Разработка инструментальной системы распределенного имитационного моделирования // Информационные технологии и вычислительные системы. – 2009. – № 4. – С. 3–15.
3. **Бухановский А. В., Ковальчук С. В., Марьин С. В.** Интеллектуальные высокопроизводительные программные комплексы моделирования сложных систем: концепция, архитектура и примеры реализации // Известия высших учебных заведений. Приборостроение. – 2009. – Т. 52. – № 10. – С. 5–24.
4. **Ford D. A., Kaufman J. H., Eiron I.** An extensible spatial and temporal epidemiological modelling system // International Journal of Health Geographics. – 2006. – Vol. 5. – No. 4. doi:10.1186/1476-072X-5-4
5. **Tian C., Ding W., Cao R., Jiang S.** Extensive Epidemic Spreading Model Based on Multi-agent System Framework // Lecture Notes in Computer Science. Vol. 4490. Springer, Heidelberg. – 2007. – P. 129–133.
6. **Pertsev N. V., Leonenko V. N.** Stochastic individual-based model of spread of tuberculosis // Russ. J. Numer. Anal. Math. Modelling. – 2009. – V.24. – No. 4. – P. 341–360.
7. **Pertsev N. V., Leonenko V. N.** Discrete stochastic model of HIV infection spread within a heterogeneous population // Russ. J. Numer. Anal. Math. Modelling. – 2012. – V.27. – No. 5. – P. 459–477.
8. **Леоненко В. Н., Перцев Н. В.** Анализ эффективности программ выявления индивидуумов, предрасположенных к колоректальному раку, на основе имитационного моделирования // Управление большими системами. – 2011. – № 35. – С. 207–236.
9. **Замятина Е. Б.** Современные теории имитационного моделирования. Специальный курс. – Пермь: ПГУ, 2007. – 119 с.
10. **Михайлов Г.А., Войтишек А.В.** Численное статистическое моделирование. Методы Монте-Карло. М.: Издательский центр «Академия», 2006.
11. **Марченко М.А.** Комплекс программ MONC для распределенных вычислений методом Монте-Карло // Сибирский журнал вычислительной математики. – 2004. – Т. 7. – № 1. – С. 43–55.