

ПОСТРОЕНИЕ МОДЕЛИ МОБИЛИЗАЦИИ И АФЕРЕЗА КРОВЕТВОРНЫХ СТВОЛОВЫХ КЛЕТОК У ПАЦИЕНТОВ С ОНКОЛОГИЧЕСКИМИ ЗАБОЛЕВАНИЯМИ С ИСПОЛЬЗОВАНИЕМ ГИБРИДНОЙ ДСМ-СИСТЕМЫ

И.В. Высочин¹, И.Б. Яшков², Д.В. Виноградов²

¹НИИ Скорой помощи им. Н.В. Склифосовского, Москва

²Всероссийский институт научной и технической информации, Москва

E-mail: i.vysochin@yandex.ru

В статье представлена гибридная комбинаторно-логическая система, разработанная для предсказания результатов сбора гемопоэтических стволовых клеток, основанного на особенностях пациента и параметрах лейкоцитафереза. Стволовые клетки могут использоваться для восстановления гемопоэза и собираются у пациента после проведения химиотерапии, которая вызывает мобилизацию клеток в периферическую кровь. Работа основана на данных, описывающих приблизительно 320 пациентов и 585 процедур сбора, которые проводились с 1994 по 2010 гг. Процедура анализа включала следующие шаги: анализ и предварительная подготовка данных; создание ДСМ-модуля, используемого для построения групп объектов из начальных данных и построения правил для присвоения одной или нескольких групп новому пациенту; создание статистического модуля, используемого для построения регрессионной модели по каждой из групп (регрессионные модели использовали для предсказания результатов сбора); анализ выходных данных. Регрессионные модели использованы для предсказания результатов сбора клеток на основе анализа данных системы.

Работа над системой может быть продолжена в следующих направлениях: коррекция регрессионных моделей, основанных на новых данных; изменение модулей, созданных для ежедневного предсказания уровня стволовых клеток в периферической крови пациента; создание интегральной системы из созданных модулей, которыми может полностью управлять один пользователь.

Ключевые слова: гемопоэтические стволовые клетки, сбор стволовых клеток, регрессионный анализ, гибридный ДСМ-модуль.

Key words: hematopoietic stem cells, stem cells collection, regression models, JSM-module.

интенсивность предшествующей химиотерапии; интервал от последней химиотерапии и т.д.

Таким образом, актуальность работы обоснована потребностью в математической модели, связывающей характеристики пациента и параметры процедуры лейкоцитафереза и качество клеточного сепарата, иначе говоря, позволяющей спрогнозировать объем и оптимальный день для начала сбора клеток в зависимости от индивидуальных характеристик пациента и результатов анализа крови, полученного после введения препарата, способствующего мобилизации. Первый вариант такой модели был реализован в рамках дипломной работы [2] по созданию гибридной ДСМ-системы под руководством к.ф.-м.н. Д.В. Виноградова (ВИНИТИ).

Материалы и методы

В качестве экспериментального массива данных была взята база данных с информацией о пациентах с онкологическими заболеваниями и процедурах сбора стволовых клеток у них. В период с 1994 по 2010 г. проведено 585 сеансов лейкоцитафереза у 320 пациентов. В исследование были включены как здоровые доноры, так и пациенты с онкологическими заболеваниями. Процедуру лейкоцитафереза выполняли с целью получения гемопоэтических стволовых клеток, которые были использованы в последующем как аутотрансплантат. Все пациенты накануне сеанса лейкоцитафереза получали препараты гликозилированного и негликозилированного Г-КСФ в течение

Несмотря на то что области использования гемопоэтических стволовых клеток из периферической крови хорошо известны и постоянно расширяются, механизм их мобилизации изучен недостаточно. Ng-Cashin и соавт. рассмотрено несколько эмпирических правил, позволяющих определить оптимальный день для начала сбора клеток [1]. Под оптимизацией понимается минимизация количества процедур лейкоцитафереза при максимальном результирующем

объеме сбора гемопоэтических стволовых клеток. Кроме того, указывается на неопределенность относительно объема лейкоцитафереза – часть экспериментов показывает, что большие объемы позволяют проводить меньше процедур для сбора необходимого числа клеток, в то время как другие свидетельствуют о том, что количество необходимых процедур не зависит от объема. Также рассматривается ряд факторов, влияющих на интенсивность и скорость мобилизации:

нескольких дней в дозе от 5 до 10 мкг/кг в сут. Онкологическим больным (онкогематологическим и с солидными опухолями; всего около 40 нозологий) начинали стимуляцию кроветворения препаратами Г-КСФ при выходе из нейтропении, развившейся на фоне ранее проведенной химиотерапии. Отсроченное назначение препаратов Г-КСФ, основанное на кинетике восстановления гематопоза, позволяет более точно прогнозировать время начала первого сеанса лейкоцитоза, уменьшает суммарную курсовую дозу препарата и стоимость процедуры, а также снижает частоту возникновения и тяжесть побочных эффектов. Лейкоцитоз проводили на непрерывно-поточном и прерывисто-поточном клеточных сепараторах Cobe spectra и Baxter CS-3000plus.

Контролем качества полученного клеточного сепарата служили показатели: содержание ядросодержащих и позитивных по поверхностному мембранному маркеру CD 34. После сбора гемопоэтические стволовые клетки криоконсервировали с криофилактиком ДМСО в режиме медленного контролируемого замораживания 1–3°C/мин. Хранили клетки в жидком азоте. В последующем гемопоэтические стволовые клетки использовали для аутотрансплантации онкологическим больным.

Для математического анализа использовали исходные данные, которые разделили на два основных массива:

- 1) характеристики пациентов;
- 2) параметры сборов (некоторым пациентам провели несколько сеансов лейкоцитоза).

В качестве целевого объекта была выбрана пара пациент–сбор, причем из множества характеристик отобраны потенциально значимые:

1. Возраст пациента (€ N).
2. Пол (€ {‘m’, ‘f’}).
3. Вес (€ Q).
4. Диагноз (конечное множество значений мощностью ≈40). В базе было представлено 40 диагнозов, среди которых было выделено 9 групп.
5. Является ли заболевание рецидивом (€ {t, f}).
6. Имеются ли метастазы (€ {t, f}).
7. Количество курсов химиотерапии перед лейкоцитозом (€ N).

8. Доза препарата (колониестимулирующего фактора) (€ Q).

9. Промежуток между началом курса химиотерапии и первым днем введения Г-КСФ в днях (€ N).

10. Промежуток между окончанием курса химиотерапии и первым днем введения Г-КСФ в днях (€ N).

11. Продолжительность введения Г-КСФ в днях (€ N).

12. Число тромбоцитов в крови (€ Q).

13. Число лейкоцитов в крови (€ Q).

14. Гематокрит (€ Q).

15. Объем обработанной крови (€ Q).

16. Скорость потока крови в сепараторе (€ Q).

17. Количество ядросодержащих клеток в крови пациента (€ Q).

18. Содержание клеток CD34+ в периферической крови пациента (€ Q).

Целевыми параметрами, значения которых необходимо спрогнозировать, выбраны две величины:

1. Количество CD34+ клеток в лейкоконцентрате на 1 кг массы пациента (€ Q).

2. Количество CD34+ клеток в лейкоконцентрате на 1 кг массы пациента, отнесенное к количеству обработанного объема циркулирующей крови, – ОЦК (€ Q).

Список объектов, полученных на этом этапе, необходимо было разделить на группы с помощью алгоритма группировки, основанном на **ДСМ-методе автоматического порождения гипотез**.

Обоснованность выбора для группировки именно ДСМ-метода подтверждается следующими обстоятельствами:

– хотя значения большинства свойств объекта являются числовыми параметрами, отдельные характеристики представлены элементами дискретных множеств, что не дает возможности использовать для кластеризации алгоритмы с вычислением расстояния в евклидовом пространстве. Учитывать же такие параметры, как «Диагноз», в ходе группировки необходимо, т.к. они являются одними из наиболее существенных для результативности сбора факторов;

– алгоритмы кластеризации типа Expectation-Maximization, применяемые при кластеризации объектов с категориальными параметрами, требуют наличия для них функции плотности вероятности, однако невозможно адекватно вычислить

такую функцию для параметров, полученных путем приведения числовых величин к категориальному виду разбиением на равные интервалы;

– наконец, результатом работы алгоритма группировки должны быть группы объектов, для которых определен набор факторов, оказывающий наибольшее влияние на результативность процесса сбора. ДСМ-метод позволил выделить такие группы на этапе формирования гипотез.

Группы объектов – результат работы алгоритма ДСМ-кластеризации – используются в качестве входных данных для статистического модуля, реализующего алгоритм линейной регрессии.

В качестве математической модели для описания интересующей нас зависимости была выбрана **пошаговая линейная многопараметрическая регрессия**. Такой выбор оправдан гипотезой о существовании зависимости результатов сбора от характеристик пациента и параметров лейкоцитоза. Скорее всего, эта зависимость представляется в виде сложной функции, однако «мы можем стремиться к тому, чтобы отразить ее с помощью математического уравнения данного вида. Если уравнение физически бессмысленно, то оно тем не менее может оказаться весьма ценным для предсказания значений ряда переменных по известным значениям других переменных, быть может, при определенных ограничениях» [3]. Дрейпер имеет в виду, что в уравнении линейного вида зависимость может выглядеть как сумма разнородных параметров с коэффициентами, которая в общем случае не имеет физического смысла. Тем не менее такой вид уравнений может быть полезен для предсказания значений. Такими ограничениями в данном случае служит принадлежность пациента одной из групп – в зависимости от этого выбирается то или иное уравнение регрессии.

В качестве параметров для регрессии были выбраны те из потенциально важных характеристик, которые изначально представлялись в виде целого или действительного числа. В качестве откликов использовались те же параметры, что и в ходе ДСМ-группировки – «Количество CD34+ клеток в лейкоконцентрате на 1 кг

массы пациента» и «Количество CD34+ клеток в лейкоконцентрате на 1 кг массы пациента, отнесенное к количеству обработанных ОЦК». Вычисление коэффициентов уравнения регрессии производилось по точкам, принадлежащим одной группе.

Результаты

В результате работы алгоритма ДСМ-группировки выделены 19 групп объектов (пара пациент–сбор). Объект считается принадлежащим группе n, если его параметры удовлетворяют ограничениям, указанным в описании этой группы. Один объект может принадлежать нескольким группам.

Принадлежность объекта группе определяет уравнение регрессии, которое используется для прогнозирования результативности сбора.

Отметим, что распределение по группам проводилось исключительно по результатам работы алгоритма, без попытки объяснить причину, по которой именно такой набор значений факторов стал определяющим для данной группы. Тем не менее можно заметить, что определенные значения факторов (например, малое значение гематокритного числа, отсутствие метастазов или малый вес пациента) чаще появляются в условиях принадлежности группам, чем другие значения тех же факторов. Можно выдвинуть гипотезу о большей значимости факторов, находящихся в этих диапазонах, однако для ее проверки требуется более тщательный анализ.

Применять полученные данные о распределении объектов по группам можно для прогнозирования результативности сбора у нового объекта. Для этого при-

меняется уравнение регрессии, которое строится по значениям факторов тех объектов, которые принадлежат той же группе, что и новый. Для этого применяется статистический модуль, обрабатывающий данные из базы: задаются номер группы, ограничение на уравнение регрессии и строится уравнение.

Например, для группы номер 5 с ограничением не больше 5 факторов уравнение имеет следующий вид (см. рисунок):

$$[Количество\ CD34+\ кеток\ в\ лейкоконцентрате\ на\ 1\ кг\ массы\ пациента] = -1,5643 + 1,3904 * [Процент\ клеток\ CD34+\ в\ периферической\ крови\ пациента,\ \%] + 0,1325 * [Вес\ пациента,\ кг] + -0,0014 * [количество\ тромбоцитов\ в\ крови,\ \cdot 10^9] + -0,0457 * [Скорость\ потока\ крови\ в\ сепараторе,\ мл/мин] + 0,3644 * [Количество\ ядросодержащих\ клеток\ в\ крови\ пациента,\ \cdot 10^9].$$

Список групп

№	Признаки
1	Мужской пол, продолжительность введения Г-КСФ 8–32 дня, количество ядросодержащих клеток в крови пациента – 1,1–1,7 · 10 ⁹ /л, гематокрит < 22%
2	Интервал между окончанием курса химиотерапии и первым днем введения Г-КСФ 2–6 дней, количество ядросодержащих клеток в крови пациента < 10 ⁹ /л или количество тромбоцитов в крови 79–117 · 10 ⁹ /л
3	Доза препарата 9,5–10 мкг/кг, количество ядросодержащих клеток в крови пациента < 10 ⁹ /л или количество курсов химиотерапии 4–6
4	Женский пол, продолжительность введения Г-КСФ 8–32 дня, количество лейкоцитов в крови < 3,9 · 10 ⁹ /л, скорость потока крови в сепараторе < 25 мл/мин
5	Женский пол, отсутствие метастазов, интервал между окончанием курса химиотерапии и первым днем введения Г-КСФ 2–6 дней, продолжительность введения Г-КСФ 8–32 дня, число тромбоцитов в крови 41–78 · 10 ⁹ /л
6	Наличие метастазов, количество курсов химиотерапии < 2, количество CD34+ клеток в периферической крови пациента 0,61–0,94%
7	Возраст пациента 13–14 лет, отсутствие рецидива, количество курсов химиотерапии < 2, количество CD34+ клеток в периферической крови пациента 0,95–1,7% или продолжительность введения Г-КСФ < 3 дней
8	Доза препарата (Г-КСФ) 10,4–21,9 мкг, продолжительность введения Г-КСФ < 3 дней
9	Рецидив основного заболевания, наличие метастазов, количество ядросодержащих клеток в крови пациента 6–60 · 10 ⁹ /л
10	Возраст пациента < 3 лет, количество CD34+ клеток в периферической крови пациента 2,9–13%
11	Диагноз – нейробластома, количество ядросодержащих клеток в крови пациента 1,73–2,33 · 10 ⁹ /л
12	Количество лейкоцитов в крови < 3,9 · 10 ⁹ /л, гематокрит < 22%
13	Продолжительность введения Г-КСФ 8–32 дня, количество ядросодержащих клеток в крови пациента < 10 ⁹ /л, количество CD34+ клеток в периферической крови пациента 0,95–1,7%
14	Отсутствие рецидива, отсутствуют метастазы или количество курсов химиотерапии < 2, интервал между окончанием курса химиотерапии и первым днем введения Г-КСФ 11–14 дней
15	Вес пациента < 13 кг, женский пол, отсутствие рецидива, отсутствуют метастазы
16	Интервал между окончанием курса химиотерапии и первым днем введения Г-КСФ 2–6 дней, объем обработанной крови < 10 ³ мл/кг, количество CD34+ клеток в периферической крови пациента 2,9–13%
17	Вес пациента < 13 кг, продолжительность введения Г-КСФ 8–32 дня
18	Отсутствие рецидива, количество лейкоцитов в крови < 3,9 · 10 ⁹ /л, объем обработанной крови < 103 мл/кг, количество ядросодержащих клеток в крови пациента < 10 ⁹ /л
19	Отсутствие рецидива, отсутствуют метастазы, число лейкоцитов в крови < 3,9 · 10 ⁹ /л, гематокрит < 22



данные о пациентах и уточнять математическую модель.

Так как нами была создана база данных для хранения текущих данных и результатов их анализа, а также для пополнения данных о пациентах и сборах, возможна дальнейшая работа с уже имеющимися и новыми данными с применением более адекватных предметной области алгоритмов статистического анализа.

Задача прогнозирования оптимального дня для начала процесса сбора остается открытой, однако существующая структура базы данных позволяет вносить информацию об уровне лейкоцитов и дозе препарата в ежедневном разрезе. При наличии достаточного количества примеров эта информация может быть проанализирована с применением тех же методов, что использовались в текущей работе.

Также остается открытой задача проверки корректности прогноза – как предварительного, с использованием ДСМ-гипотез, так и основанного на регрессионной модели. Оценку корректности также можно будет дать в ходе пополнения базы.

Авторы выражают благодарность сотрудникам отделения трансплантации костного мозга НИИ детской онкологии и гематологии РОНЦ им. Н.Н. Блохина РАМН за предоставление базы данных по лейкаферезам и ведущему научному сотруднику отделения, д.м.н. И.С. Долгополову за консультативную помощь.

На приведенном графике, который формируется статистическим модулем, показан результат вычисления уравнения регрессии для группы объектов (точки на оси x – коды объектов, на оси y – значения отклика «Количество CD34+ клеток в лейкоконцентрате на 1 кг массы пациента, отнесенное к количеству обработанных ОЦК»). Среднюю величину ошибки можно со временем минимизировать, если пополнять базу данных пациентов и сборов.

Полученные уравнения регрессии рекомендуется применять для прогнозирования оптимального дня для начала сбора, что является одной из главных целей нашей дальнейшей работы.

Заключение

Создана система для хранения информации о пациентах и процедурах сбора и ее анализа в соответствии с разработанной математической моделью. В ходе работы достигнуты следующие результаты:

1. Сформировано 19 гипотез, позволяющих распределять пациентов по группам и прогнозировать результативность процесса сбора.
2. Для каждой из групп построены регрессионные модели, с помощью которых можно дополнить прогноз.
3. Создана база данных и приложение для работы с ней, что позволит пополнять

ЛИТЕРАТУРА

1. Ng-Cashin, Judith. Shea, Thomas. Mobilization of Autologous Peripheral Blood Hematopoietic Cells for Support of High-Dose Cancer Therapy / Blume, Karl G. Forman, Stephen J. Appelbaum, Frederick R // Hematopoietic Cell Transplantation: Third Edition. – Blackwell Publishing, Inc. 2004.

2. Яшков И.Б. Гибридная ДСМ-система для моделирования стимуляции и сбора кроветворных стволовых клеток у доноров и онкологических больных детей // М. – РГГУ, Институт лингвистики, кафедра математики, логики и интеллектуальных систем в гуманитарной сфере, 2010.

3. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. Издание второе, переработанное и дополненное // М. – Финансы и статистика, 1986.

SUMMARY

HYBRID JSM-SYSTEM FOR STEM CELLS MOBILIZATION AND COLLECTION MODELING

I.V. Vysochin¹, I.B. Yashkov², D.V. Vinogradov²

¹Scientific Research Institute for Emergency Care of N.V. Sklifosovsky

²VINITIRAS

This article describes a hybrid combinatorial-logical system designed to predict the result of the hematopoietic stem cells harvest based on the patient's characteristics and the parameters of the operation. Hematopoietic stem cells can be used to recover hematopoietic system after it was damaged and are collected from a patient after they were given chemotherapy that

causes cell's mobilization into their peripheral blood. The work is based on the data describing about 320 patients and 585 harvest procedures that took place from 1994 to 2010. The following steps are described: data analysis and preprocessing; creating the JSM-module used to build object clusters from the initial data and generate the rules to ascribe the patient to one of the clusters; creating the statistical module used to build the regression models for each of the clusters. The regression models are used to predict the result of the collection; analyzing the system output.

The work on the system can be continued in the following directions: correction of the regression models based on new data; modifying the modules created to make day-to-day predictions about stem cells level in the patient's peripheral blood; making an integral system from the created modules, that can be wholly operated by one user.