

СИСТЕМА ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ НА ОСНОВЕ МУЛЬТИМЕДИЙНОЙ ИСТОРИИ БОЛЕЗНИ: КОЛИЧЕСТВЕННОЕ СРАВНЕНИЕ И МНОГОМЕРНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ (НА ПУТИ К УНИФИЦИРОВАННОМУ ЯЗЫКУ ВОЗ В УКАЗАННОЙ ПРЕДМЕТНОЙ ОБЛАСТИ)

Шакин В. В., Шкловский-Корди Н. Е., Пташко Г. О.

(Россия, Москва)

Система поддержки принятия решений, или СППР, предназначена для сравнительного анализа историй болезни между собой, а также для сопоставления историй болезни с кластерами, т.е. обобщенными описаниями однотипных историй болезни, на основе моделирования и обобщения верифицированной клинической информации о конкретных пациентах. Используя специальные процедуры нормализации по времени и по соответствующим подпространствам параметров, а также специальные мультимедийные интерполяционные методы, СППР делает данные отдельно взятых мультимедийных историй болезни попарно сравнимыми. Это создает основу для единого унифицированного языка, отвечающего запросам и нуждам ВОЗ в части количественного описания сравнения и обобщенного описания образов конкретных пациентов.

Введение

Как известно, лечащие врачи, в процессе своей работы, отлично справляются с болезнью пациента, используя для этого все свои силы и знания. Любой врач знает или, по крайней мере, представляет себе ход болезни в пределах одного пациента. Врач может предсказать ход болезни с той или иной ошибкой, может сказать «хорошо» или «плохо» болеет пациент, то есть легкий ли это или тяжелый случай болезни. Но все это врач может сделать только в пределах одного пациента. Он не может сказать хуже или лучше болеет один пациент, чем другой. Так

же не может сказать хуже или лучше лечат одного пациента, чем другого. Для того чтобы врач мог оперировать такими понятиями, как качество лечения пациента в сравнении с другими или принимать решения, подкрепленные поддержкой компьютера, который умеет анализировать истории болезни и делать соответствующие выводы, нужно в первую очередь обобщить и формализовать историю болезни пациента.

История болезни в общем случае представляется, как записи о состоянии пациента, например в виде результатов клинических анализов, и записи о воздействиях, оказываемых на пациента, например назначение каких-либо лекарств. Все записи представлены во времени. В данной статье предлагается формализация истории болезни, как записи о состоянии пациента во времени. Рассматривается вопрос сравнения историй болезни, имеющих разные длительности по времени.

История болезни, как элемент функционального пространства

Для каждого пациента история болезни (далее ИБ) начинается с того момента, как он поступил в больницу, и заканчивается его выпиской. Таким образом, можем выделить два момента времени t_b и t_e — начальное время (поступление) и конечное время (выписка). Имеем отрезок времени $[t_b, t_e]$, в течение которого мы рассматриваем записи о состоянии пациента. Для того чтобы обобщить ИБ нужно, как минимум, абстрагироваться от начального и конечного времени конкретного случая и поставить каждой ИБ одно и тоже начальное и конечное время. Отообразим отрезок $[t_b, t_e]$ в отрезок $[0, T]$, где $T = t_e - t_b$. Ясно, что для каждого пациента время T будет разным.

Рассмотрим конечное множество ИБ и обозначим его как $H = \{H_i\}$, $i = \overline{1, n}$. Каждой ИБ H_i соответствует свое время T_i . В пределах множества H будем считать, что параметры, отражающие состояние пациента, фиксированы и их m штук.

Обозначим их $p_1^i, p_2^i, \dots, p_m^i$, $i = \overline{1, n}$, где p_j^i соответствует j – му параметру i – ой ИБ H_i . Каждый параметр p_j^i можно представить, как сеточную функцию, заданную на сетке на отрезке $[0, T_i]$. Обозначим эти сетки T_j^i и получим функции p_j^i следующего вида $p_j^i(t)$, где $t \in T_j^i = \{t_{j,1}^i, t_{j,2}^i, \dots, t_{j,k}^i\}$.

Используя такое задание сеточных функций p_j^i , интерполируем их кусочно-линейно на соответствующих отрезках времени $[0, T_i]$. После интерполяции получим соответствующие интерполянты P_j^i . Теперь каждой ИБ $H_i \in H$ будем ставить в соответствие вектор функцию следующего вида

$$\begin{aligned} f_i : [0, T_i] &\rightarrow \mathbf{R}^m \times [0, T_i], \\ f_i(t) &= (P_1^i(t), P_2^i(t), \dots, P_m^i(t), t) \end{aligned} \quad (1)$$

Отметим, что в вектор, стоящий в правой части выражения (1) входит время t , как компонент, от которого нам в дальнейшем предстоит избавиться. Данная функция описывает ход болезни пациента на основе m параметров во времени $[0, T_i]$.

Для того чтобы сравнивать полученные функции между собой, тем самым, проводя сравнительный анализ историй болезни, мы должны нормализовать их по времени и по параметрам.

Рассмотрим нормализацию параметров. Каждый рассматриваемый параметр P_j^i имеет минимальное и максимальное значение в пределах множества H . Обозначим их соответственно P_{\min}^j и P_{\max}^j : $P_{\min}^j = \min_i P_j^i$, $P_{\max}^j = \max_i P_j^i$.

Так же каждый параметр P_j^i имеет оптимальное значение, которое в нашем случае назначается врачом-экспертом. Обозначим его P_{opt}^j . Так, например, для температуры оптимальное

значение — это $P_{opt} = 36,6^\circ$. Масштабируем теперь область значений каждой из функций $P_j^i(t)$ следующим образом

$$\bar{P}_j^i(t) = \frac{P_j^i(t) - P_{opt}^j}{\max(P_{opt}^j - P_{\min}^j, P_{\max}^j - P_{opt}^j)}.$$

Получим нормализованные параметры \bar{P}_j^i , соответствующие параметрам $P_j^i(t)$. При таком масштабировании лучшее значение параметра переходит в ноль, а худшее переходит либо в единицу, либо в минус единицу, в зависимости от того, в какую сторону было отклонение от нормы.

Нормализацию по времени проведем следующим образом. Масштабируем время T_i для каждой $H_i \in H$. Обозначим T_{\max} максимальное время ИБ из H и T_{\min} минимальное время ИБ, т.е. $T_{\max} = \max_i \{T_i\}$, $T_{\min} = \min_i \{T_i\}$.

Тогда каждому T_i поставим в соответствие время, полученное следующим образом:

$$\bar{T}_i = \frac{T_i - T_{\min}}{T_{\max} - T_{\min}}.$$

После данных преобразований мы получим, что каждой функции f_i ставится в соответствие новая функция \bar{f}_i

$$\bar{f}_i : [0, \bar{T}_i] \rightarrow [-1, 1]^m \times [0, \bar{T}_i], \text{ где } \bar{T}_i \in [0, 1].$$

Таким образом, мы провели унификацию всех параметров процессов ИБ $H_i \in H$, включая время. Теперь наша задача состоит в том, чтобы избавиться от времени, как от параметра.

Для каждой функции $\bar{f}_i(t)$, $t \in [0, \bar{T}_i]$ будем строить новую функцию $s_i(t)$, которая принимает значение длины кривой, описываемой функцией $\bar{f}_i(t)$ в момент времени t . Данная

функция будет непрерывной и монотонно возрастающей, что позволяет нам построить новую функцию $t_i(s)$, являющуюся обратной к соответствующей $s_i(t)$. Теперь для каждой $t_i(s)$ сожмем область определения и получим новую функцию $\bar{t}_i(s)$, где $s \in [0, 1]$:

$$\bar{t}_i(s) = t_i(s \cdot S_{max}^i), \text{ где } S_{max}^i = s_i(\bar{T}_i).$$

Таким образом, мы получим функцию времени от относительной длины соответствующей кривой, описываемой функцией $\bar{f}_i(t)$.

Наконец, каждой функции $\bar{f}_i(t)$ будем ставить в соответствие новую функцию $\bar{\bar{f}}_i$ следующего вида

$$\begin{aligned} \bar{\bar{f}}_i : [0, 1] &\rightarrow [-1, 1]^m, \text{ где} \\ \bar{\bar{f}}_i(s) &= (\bar{P}_1^i(\bar{t}_i(s)), \bar{P}_2^i(\bar{t}_i(s)), \dots, \bar{P}_m^i(\bar{t}_i(s))), \quad s \in [0, 1] \end{aligned} \quad (2)$$

То есть теперь будем рассматривать ИБ СН $H_i \in H$, как элемент функционального пространства функций вида (2). В данном пространстве можно ввести скалярное произведение следующего вида

$$\langle f, g \rangle = \int_0^1 (\bar{P}_1^f(\bar{t}_f(s)) \cdot \bar{P}_1^g(\bar{t}_g(s)) + \dots + \bar{P}_m^f(\bar{t}_f(s)) \cdot \bar{P}_m^g(\bar{t}_g(s))) ds$$

а норму и метрику, как

$$\|f\| = \sqrt{\langle f, f \rangle}, \quad \rho(f, g) = \|f - g\|.$$

Итак, мы имеем обобщенное представление истории болезни пациента, как элемента функционального пространства, где мы можем проводить сравнительный анализ историй болезни между собой.

Пример сравнительного анализа истории болезни нескольких пациентов

В качестве сравниваемых историй болезни были взяты двенадцать мультимедийных историй болезни (ММИБ [1]) пациентов ГНЦ РАМН. Среди рассматриваемых пациентов (обозначим их p_1, p_2, \dots, p_{12}) болезни распределились следующим образом

| | |
|-------------------------|----------------------|
| Пациент 1 (p_1) | В-лимфопролиферация |
| Пациент 2 (p_2) | ЛГМ |
| Пациент 3 (p_3) | Апластическая анемия |
| Пациент 4 (p_4) | Лимфосаркома |
| Пациент 5 (p_5) | ЛГМ |
| Пациент 6 (p_6) | Пурпура |
| Пациент 7 (p_7) | Острый лейкоз |
| Пациент 8 (p_8) | Острый лейкоз |
| Пациент 9 (p_9) | ХМЛ |
| Пациент 10 (p_{10}) | Острый лейкоз |
| Пациент 11 (p_{11}) | Острый лейкоз |
| Пациент 12 (p_{12}) | Острый лейкоз |

Используемые параметры

Время — t

В нашем случае значения тех или иных параметров «снимаются» с пациента каждый день. Соответственно наименьший шаг по времени – это один день. Для простоты показательных расчетов каждый день, начиная с первого, отображается в целое положительное число. Первый день отображается в число 0, второй в 1 и т.д. После данного преобразования получаем отрезок времени $[0, T]$.

Пример отображения времени

| | | | | | |
|--------------------|------------|------------|------------|-----|------------|
| Время реальное | 04.04.1997 | 05.04.1997 | 06.04.1997 | ... | 21.01.1998 |
| Время для расчетов | 0 | 1 | 2 | ... | 26 |

Температура – $Temp(t)$, $t \in [0, T]$

Температура пациента представлена в обычном виде, в градусах Цельсия. В нашем случае, как говорилось выше, мы имеем сетку с шагом по времени равным 1. Температура пациента оказалась самым часто измеряемым параметром, поэтому для нее имеем самую полную таблицу задания.

Пример табличного задания температуры

| | | | | | |
|-------------|------|------|------|-----|------|
| Время | 0 | 1 | 2 | ... | 26 |
| Температура | 36,6 | 36,7 | 36,6 | ... | 36,8 |

Тромбоциты — $Tr(t)$, $t \in [0, T]$

Тромбоциты — это клетки крови, отвечающие за свертывание крови. Значения параметра «тромбоциты» представляют собой величину порядка 10^9 , которая в истории болезни записывается в виде $Tr \cdot 10^9$ клеток/л. Поскольку абсолютное значение параметра на данном этапе нас в принципе не интересует, то для расчетов мы будем брать величину Tr . Следует отметить тот факт, что значение параметра отмечается в истории болезни в том случае, если оно отошло от нормы. Но если в истории болезни в момент времени t_b или t_e значение данного параметра отсутствует, мы берем значение второго или соответственно предпоследнего замера данного параметра.

Пример табличного задания параметра «тромбоциты»

| | | | | | | |
|------------|-------|-------|-------|-------|-------|-------|
| Время | 0 | 2 | 18 | 21 | 24 | 26 |
| Тромбоциты | 295,0 | 295,0 | 186,0 | 227,0 | 170,0 | 170,0 |

Лейкоциты — $Le(t)$, $t \in [0, T]$

Лейкоциты — это клетки крови, выполняющие защитную функцию организма. Значение параметра «лейкоциты» представляют собой величину порядка 10^9 , которая в истории болезни записывается в виде $Le \cdot 10^9$ клеток/л. По аналогии с параметром «тромбоциты» для расчетов мы будем брать величину Le , а в качестве отсутствующих значений на концах временно-

го отрезка $[t_b, t_e]$ мы так же берем значение второго или соответственно предпоследнего замера данного параметра.

Пример табличного задания параметра «лейкоциты»

| | | | | | | |
|-----------|------|------|-----|-----|-----|-----|
| Время | 0 | 2 | 18 | 21 | 24 | 26 |
| Лейкоциты | 12,8 | 12,8 | 1,3 | 1,9 | 2,4 | 2,4 |

Нормализация

Как говорилось выше, для нормализации параметров, мы должны указать максимальное, минимальное и нормальное значение каждого параметра. В нашем случае максимальное и минимальное значение по каждому параметру по всем рассматриваемым пациентам. Нормальное значение P_{opt} по каждому параметру бралось согласно следующей таблице

| | |
|-------------|-----------------------|
| Температура | 36,6° |
| Тромбоциты | $250 \cdot 10^9$ кл/л |
| Лейкоциты | $6 \cdot 10^9$ кл/л |

Расчет

Для расчетов расстояний была использована система Matlab. Расстояния между историями болезни были получены в виде квадратной, симметричной матрицы с нулями на главной диагонали. Далее, эта матрица была использована в качестве входных данных в процедуре многомерного шкалирования в системы SPSS. Эта процедура была проведена для отображения историй болезни в точки на плоскости, в соответствии с полученными расстояниями между ними. Таким образом, мы получили наглядное представление о распределении между собой рассматриваемых историй болезни (Рис. 1).

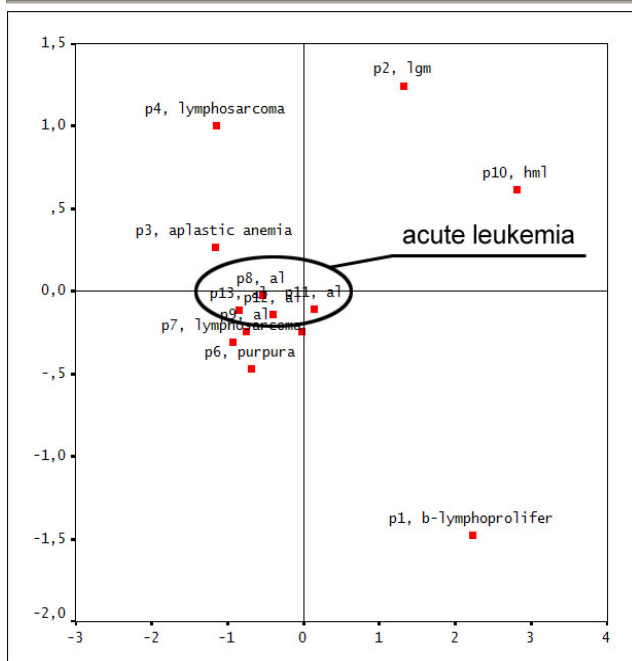


Рис. 1

На рисунке видно, как сгруппировались истории болезни. Конечно же, столь малое количество параметров, на основе которых проведено сравнение не может дать объективной картины. Так же не учтены всяческие воздействия на пациента, к числу которых можно отнести, как прописываемые лекарства, так и события, которые случаются совершенно случайно, например, какая-либо травма, случившаяся с пациентом не по вине врачей (падение, например). Но даже на такой малой выборке мы можем видеть, что большинство историй болезни пациентов, болеющих острым лейкозом, сгруппировались рядом. Так же примечателен тот факт, что на тех же трех рассматриваемых параметрах некоторые истории болезни совсем отделились от создавшихся групп.

Результаты, обсуждения и перспективы

Данный подход позволяет задать на множестве ММИБ «меру сходства», что достигается путем попарного сравнения ММИБ, нормализованных по времени и по пространству параметров. Более того, становится возможным применить методы многомерного статистического анализа для кластеризации и моделирования (обобщения) образов ММИБ в виде кластеров, что позволяет улучшить наблюдение за пациентами и их лечение. При этом удается соблюсти правила и принципы «Доказательной медицины». Всё это создает основу для единого унифицированного языка, отвечающего запросам и нуждам ВОЗ в части количественного описания сравнения и обобщенного описания образов конкретных пациентов. Эти «обобщенные образы» и служат моделями кластеров, состоящих из похожих образов ММИБ. Чтобы ставить диагнозы и предлагать оптимальные лечебные предписания, предсказательные модели комбинируются с целенаправленными доказательными параметрами. Таким образом, СППР сохраняет все преимущества ММИБ при работе с конкретными историями болезни [1]. С другой стороны, СППР позволяет количественно сравнивать конкретные истории болезни. Более того, СППР — это средство для создания обобщенных образов ММИБ.

Работа поддержана грантом РФФИ 04-01-00401-а.

Список литературы:

1. Nikita Shklovskiy-Kordi. Multimedia Case History as a Tool for Management of Cancer Patients and Clinical Trials. UICC Abstracts. — Dublin, 2004.
2. Vsevolod V. Shakin, et al. Life Support Systems in Medicine and Biology. ORM-2004 — Operations Research International Conference, Proceedings. — Moscow, Dorodnicyn Computing Center of the Russian Academy of Sciences, 2004.
3. Шакин В.В. Вычислительная электрокардиография. — М.: «Наука», 1981.

**DECISION SUPPORT SYSTEM USING MULTIMEDIA
CASE HISTORY: QUANTITATIVE COMPARISON AND
MULTIVARIATE STATISTICAL ANALYSIS
(TOWARDS UNIFIED WHO LANGUAGE IN THE
SUBJECT AREA)**

Shakin V. V., Shklovskiy-Kordi N. E., Ptashko G. O.

(Russia, Moscow)

Decision Support System, or DSS, for case-to-case and case-to-cluster comparative and multivariate statistical analysis based on modeling and generalization of verified clinical information on individual patients. The DSS makes individual MMCH data pair-wise comparable using normalization procedures in individual time-and-subspace time domain and interpolation techniques for multimedia data. All that serves as background for a Unified Language to meet needs of WHO in quantitative description, comparison and generalization of individual patients' patterns.