Московский государственный университет имени М. В. Ломоносова

Факультет Вычислительной Математики и Кибернетики

Кафедра Математических Методов Прогнозирования

Афанасьев Павел Андреевич

Методы повышения точности прогноза ключевых биомедицинских индикаторов, основанных на анализе биосигналов

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

Научный руководитель:

д.ф-м.н.

О. В. Сенько

Москва, 2018

Содержание

Аннотация

В данной работе описывается способ повышения обобщающей способности на основании расширения признакового пространства, полученного с использованием двумерных моделей оптимальных разбиений. Этот метод дает возможность повысить точность распознавания групп с повышенным и нормальным систолическим давлением по синхронным сигналам, полученных с ЭКГ и ФПГ.[1]

[1]Фотоплетизмограмма - метод регистрации кровяного потока.

1 Введение

Систолическое (верхнее) артериальное давление — это давление крови в артериях в момент сокращения сердца. Его повышение приводит к ускорению естественного старения внутренних органов, особенно сердца, мозга и почек и является очень важным показателем риска сердечных заболеваний и развития фибрилляции предсердий. Повышенное артериальное давление признано фактором риска развития многих опасных заболеваний, в том числе нарушения мозгового кровообращения. Высокие цифры давления создают избыточную нагрузку на стенки сосудов, что может со временем привести к их разрыву и кровоизлиянию.

Основным прибором для измерения уровня кровяного давления является осциллометрический тонометр с надувной манжетой, который обеспечивает измерение систолического давления в плечевой артерии, а также частоты пульса. В то же время у тонометров с манжетой есть недостатки: они слишком громоздки чтобы постоянно носить их с собой; процедура одевания и накачки манжеты неудобна; они не могут не производиться в непрерывном режиме.

Фотоплетизмограмма(ФПГ) — метод регистрации кровяного потока, позволяющая измерять объёмный пульс крови, вызванный периодическим изменением кровяного объёма при каждом ударе сердца, частоту сердцебиения, вариабельность сердечного ритма. Разработанный в России кардиомонитор CardioQvark дает возможнось измерить ЭКГ, ФПГ, пульс и другие показатели, на основании которых можно получить синхронные показания фотоплетизмографических и электромагнитных датчиков. Процедура измерения давления заключается в прикладывании в течение некоторого времени пальцев каждой руки к датчикам. В целом эта процедура представляется гораздо более простой и удобной для потребителя чем тонометр с манжетой.

Целью данной работы является повышение точности расчета уровня систолического давления, который производился на основании синхронных сигналов ЭКГ и ФПГ, записанных устройством CardioQvark в течение пяти минут. На основании данных сигналов рассчитывался набор из 147 показателей: 23 геометрических и 124 спектральных. Так как период измерений вмещает около трехсот кардиоциклов, признаки расчитывались для каждой волны в отдельности и затем усреднялись. Так,

для спектральных признаков использовались среднее значение и дисперсия, а для геометрических – медиана.

В данной работе решается задача прогнозирования. Данная задача является очень трудной и до сих пор не удается достигнуть сколь либо приемлимой точности оценок. Однако прибор CardioQvark целесообразно использовать в качестве средства предварительной оценки возможности повышения давления. Такая задача по своей сути является задачей распознавания, поскольку фактически сводится к задаче отнесения каждого из измерений к группе с нормальным и повышенным давлением.

1.1 Описание признаков

Геометрические признаки: MB0 — медиана периода от R-пика до точки максимального роста пульсовой волны, MB1 — медиана периода от R-пика до точки начала пульсовой волны, MB2 — медиана периода от от R-пика до точки максимального значения пульсовой волны, MSEP – медиана периода от R-пика до пика прямой систолической волны, MSRP — медиана периода от R-пика до пика отраженной систолической волны, MDP – медиана периода от R-пика до диастолического пика, MRR — медиана периода между R-пиками, MdSEP – медиана периода между пиками прямых систолических волн, MdSRP — медиана периода между пиками отраженных систолических волн, MdDP — медиана периода между диастолическими пиками, MASEP — медиана амплитуды пиков прямых систолических волн, MASRP — медиана амплитуды пиков отраженных систолических волн, MADP — медиана амплитуды диастолических пиков, MAB2 — медиана амплитуды максимумов пульсовых волн, MSNR — медиана соотношения сигнал/шум, MS5S2 — медиана соотношения площадей сегментов под кривой пульсовой волны, MPI — медиана перфузионного индекса. Кроме того, использовались: BR — частота дыхания, quality — процент кардиокомплексов, удовлетворяющих условию качества, spqrst, spq, sqrs, sst — площади под графиком, соответствующие сегментам кардиоцикла.

В качестве границ спектральных признаков были R-пики ЭКГ.

- Первая группа спектральных признаков содержала 30 средних и 30 дисперсий разностей натуральных логарифмов спектральных амплитуд ЭКГ и ФПГ в диапазоне от 1 до 30 Гц с шагом 1 Гц. Спектры вычислялись исходя из нормировки

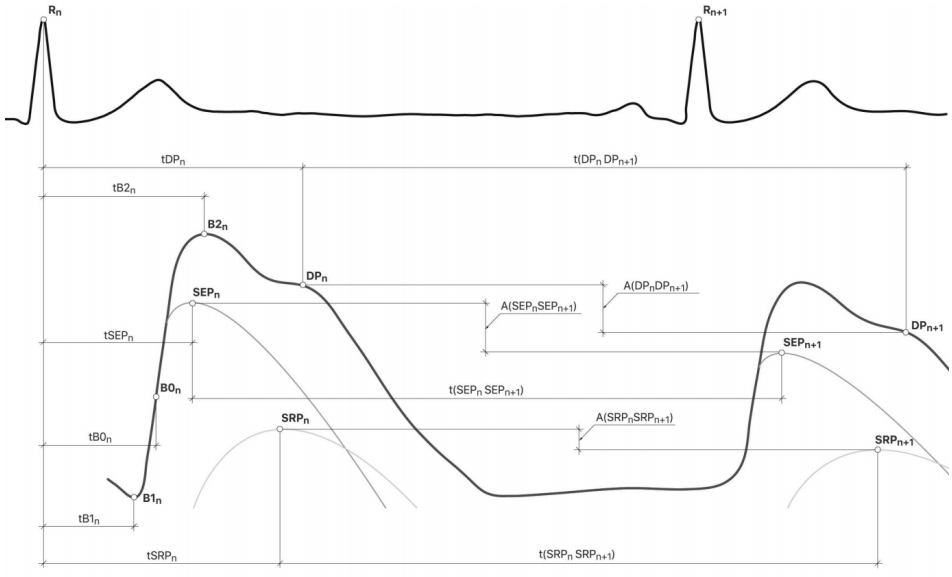

Рис. 1: Расчет признаков

времени внутри RR-интервалов на длину всего интервала. Таким образом, длина каждого интервала приравнивалась 1.

- Вторая группа содержала 32 средних и 32 дисперсии 64-точечного дискретного преобразования Фурье задержек от R-пика до минимума и максимума сигнала FPG и 62 промежуточных значений. Для этого интервал (tB1, tB1) (см. рис. 1) разбивался на 63 интервала $tB1 < tBd_i < tB2$, $i = 1 \ldots ...62$, соответствующих равному приросту сигнала ФПГ, т.е. $f(tB2) - f(tBd_{62}) = f(tBd_1) - f(tB1) = f(tBd_{j+1}) - f(tBd_j)$, где $f(t)$ – величина ФПГ сигнала в момент времени $t, j = 2 \ldots 62$.

1.2 Определения и обозначения

Назовем объектом обучающей выборки $\vec{x}_i \in \mathbb{R}^d$, объектом тестовой выборки $\vec{t}_i \in \mathbb{R}^d$ — измерение с признаками, полученными вышеописанным способом.

Пусть имеется обучающая выборка $S \in \mathbb{R}^{N \times d}$. Для каждого объекта обучающей выборки известен уровень систолического давления $y_{tr(i)} \in \mathbb{R}$.

Пусть также имеется тестовая выборка $T \in \mathbb{R}^{(M-N-1) \times d}$, отличающаяся от обучающей выборки временным промежутком. Для каждого объекта тестовой выборки

известен уровень систолического давления $y_{tt(i)} \in \mathbb{R}$.

$$S = \{(\vec{x}_i, y_{tr(i)}), i = 1 \dots N\} \ , |S| = 1600$$

$$T = \{(\vec{t}_i, y_{tt(i)}), i = N+1 \dots M\} \ , |S| = 871$$

Строим модель на обучающей выборке и делаем прогноз $\widehat{Y} \in \mathbb{R}^{M-N-1}$ на тестовой выборке, затем проверяем его качество с помощью функции R.

$$R = \sqrt{1 - \frac{\sum_{i=N+1}^{M}(y_{tt(i)} - \widehat{y}_i)^2}{\sum_{i=N+1}^{M}(y_{tt(i)} - \bar{y}_{tt(i)})^2}}, \text{ где } \bar{y}_{tt(i)} = \frac{1}{M-N-1}\sum_{i=N+1}^{M} y_{tt(i)}$$

Далее описывается метод, создающий новое признаковое пространство P, которое добавляется к исходному с целью повысить качество прогноза R.

Для применимости данного метода перейдем от сложной задачи прогнозирования к более простой задаче распознавания. Будем определять два класса:

$$y'_i = \begin{cases} 0, & \text{если } y_i < y_0; \\ 1, & \text{если } y_i \geqslant y_0 + 15. \end{cases} \text{, где } y_0 \text{ некоторый параметр.}$$

2 Описание метода

Метод определяет зависимость между бинарным целевым признаком $y \in \{0, 1\}$ и парой описательных (x^k, x^l). Он основан на построении оптимальных разбиений интервала допустимых значений отдельных признаков и двумерных совместных областей значений пары признаков.

Пусть имеется произвольная обучающая выборка $S \in \mathbb{R}^{N \times d}$. По ней строится разбиение, оптимизируя функционал Q.

$$S = \{(\vec{x}_i, y_i), i = 1 \dots N\}, \vec{x}_i \in \mathbb{R}^d, y_i \in \{0, 1\}$$

$$Q = \max_{l=1\dots4}(\nu_l - \nu_0)^2 m_l$$

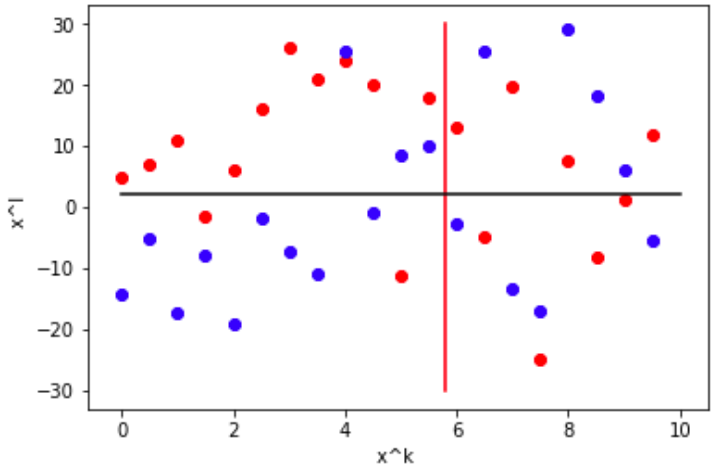

Рис. 2: Оптимальное разбиение выборки

, где $\nu_l = \dfrac{|m_{1l}|}{|m_l|}, \nu_0 = \dfrac{|m_1|}{N}$,

$m_l = \{s_i : \vec{x}_i \in q_l, i = 1 \dots N\}$,

$m_1 = \{s_i : y_i = 1, i = 1 \dots N\}$,

$m_{1l} = \{s_i : \vec{x}_i \in q_l, y_i = 1, i = 1 \dots N\}$.

ν_l - доля объектов выбранного класса в l-м квадранте.

ν_0 - доля объектов выбранного класса во всей обучающей выборке.

Данный функционал имеет геометрическую интерпретацию: максимизизация количества объектов одного класса в каждом квадранте(см. рис. 2).

Для проверки значимости закономерности функционала будем проводить перестановочные тесты. Назовем случайной перестановкой на множестве X биективное отображение $\pi : X \to X$. Пусть $X = \{1, \dots, N\}$.

$$\pi : \begin{pmatrix} 1 & 2 & 3 & \dots & N \\ \pi(1) & \pi(2) & \pi(3) & \dots & \pi(N) \end{pmatrix}$$

Пусть k — число перестановок. Применим случайные перестановки $\pi_j, j = 1 \dots k$ к нашему исходному множеству $Y \in \{0, 1\}^N$.

$$\pi_j : \begin{pmatrix} y_1 & y_2 & \dots & y_N \\ y_{\pi_j(1)} & y_{\pi_j(2)} & \dots & y_{\pi_j(N)} \end{pmatrix}$$

Сопоставим данным перестановкам выборки $\widetilde{S}_j$.

$$\widetilde{S}_j = \{(\vec{x}_i, y_{\pi_j(i)}), i = 1 \dots N\}, \vec{x}_i \in \mathbb{R}^d, y_{\pi_{j(i)}} \in \{0, 1\}, j = 1 \dots k$$

Тогда оптимизируемый функционал $\widetilde{Q}_j$ для обучающей выборки $\widetilde{S}_j$ будет выглядеть следующим образом:

$$\widetilde{Q}_j = \max_{l=1\dots4} (\widetilde{\nu}_l - \widetilde{\nu_0})^2 m_l$$

, где $\widetilde{\nu}_l = \dfrac{|\widetilde{m}_{1l}|}{|m_l|}$, $\widetilde{\nu_0} = \dfrac{|\widetilde{m}_1|}{N}$,

$\widetilde{m}_1 = \{s_i : y_{\pi_j(i)} = 1, i = 1 \dots N\}$,

$\widetilde{m}_{1l} = \{s_i : \vec{x}_i \in q_l, y_{\pi_j(i)} = 1, i = 1 \dots N\}$.

Введем множество U:

$$U = \{\pi_j | \widetilde{Q}_j > Q, j = 1 \dots k\}$$

То есть, это такие перестановки, при которых значение функционала Q больше на перестановчном множестве Y, нежели на исходном.

Назовем p-значением, h-значением — величины, проверяющие статистическую значимость некоторых гипотез. Так, данные величины могут определить случайность значения исходного функционала.

$$p = \frac{|U|}{k} \text{ , } h = \frac{Q}{\max_j \widetilde{Q}_j}$$

Интерпретация следующая:

- чем ближе p-значение к 0, тем более неслучайно значение нашего функционала;
- чем выше h-значение, тем более неслучайно значение нашего функционала.

У h-значения есть преимущество над p-значением : если обе гипотезы, которые сравниваем, имеют значения $p = 0$, то мы не можем корректно сравнить их, в то время как h-значение даст определить лучшую гипотезу, варьируя некоторый порог. Поэтому, в дальнейшем будем использовать h-значение.

3 Применение метода

Теперь применим вышеописанный метод к нашей исходной задаче. Напомним ее. Имеется обучающая и тестовая выборки с известными ответами. По обучающей выборке строим модель, по тестовой прогнозируем $\widehat{Y}$.

$$S = \{(\vec{x}_i, y_{tr(i)}), i = 1 \dots N\}$$

$$T = \{(\vec{t}_i, y_{tt(i)}), i = N+1 \dots M\}$$

Проверяем качество прогноза функцией R.

$$R = \sqrt{1 - \frac{\sum_{i=N+1}^{M}(y_{tt(i)} - \widehat{y}_i)^2}{\sum_{i=N+1}^{M}(y_{tt(i)} - \bar{y}_{tt(i)})^2}}, \text{ где } \bar{y}_{tt(i)} = \frac{1}{M-N-1}\sum_{i=N+1}^{M} y_{tt(i)}$$

Требовалось расширить признаковое пространство таким образом, чтобы увеличить качество прогноза.

Перешли от сложной задачи прогнозирования к более простой задаче распознавания, разбив систолическое давление объектов на два класса:

$$y'_i = \begin{cases} 0, & \text{если } y_i < y_0; \\ 1, & \text{если } y_i \geqslant y_0 + 15. \end{cases}, \text{ где } y_0 \text{ некоторый параметр.}$$

Так как $|S| = 1600$, то процедура поиска достоверных разбиений была трудоёмкой. Поэтому разбили обучающую выборку на две случайные части.

$$S_s = \frac{1}{4}S$$

$$S_o = \frac{3}{4}S$$

По S_s искали достоверные разбиения, по S_o h-значения.

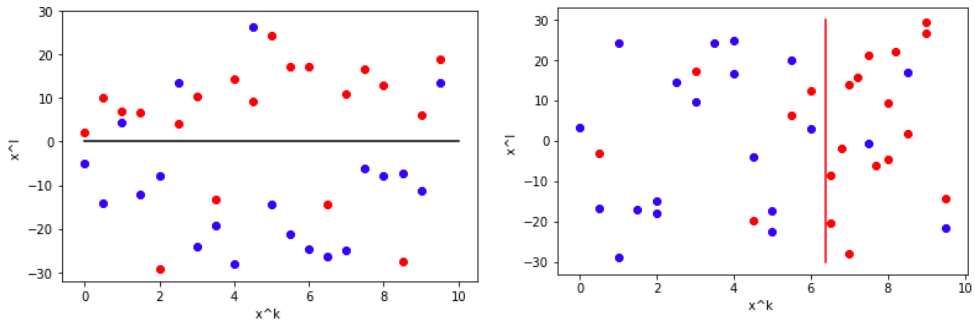

Рис. 3: Тестирование при фиксированной оси

Будем отбирать во множество P информационные признаки с помощью следующей величины:

$$h(x^k) = \frac{Q_o}{\max_j \widetilde{Q}_{o(j)}(x^k)}$$

$$h(x^l) = \frac{Q_o}{\max_j \widetilde{Q}_{o(j)}(x^l)}$$

При пороге $\alpha = 1.5$

$$P = \{(x^k, x^l) | (h(x^k) > \alpha) \& (h(x^l) > \alpha)\}$$

То есть, для поиска $h(x^k)$ фиксируем значение x^l и применяем случайные перестановки ко множеству, лежащему выше этого значения, затем ко множеству, лежащему ниже этого значения(см. рис. 3(левый)).

А для поиска $h(x^l)$ фиксируем значение x^k и применяем случайные перестановки ко множеству, лежащему левее этого значения, затем ко множеству, лежащему правее этого значения(см. рис. 3(правый)).

Далее, для каждого объекта обучающей и тестовой выборок создавался новый четырехзначный признак, определяемый систалическим давлением по обучающей выборке.

$$sys_l = \frac{1}{|m_l|} \sum_{i \in m_l} y_{tr(i)}$$

$$x_i^{new} \in \{sys_l, l = 1 \dots 4\} \ , \ i = 1 \dots N$$

$$t_i^{new} \in \{sys_l, l = 1 \dots 4\} \ , \ i = N + 1 \dots M$$

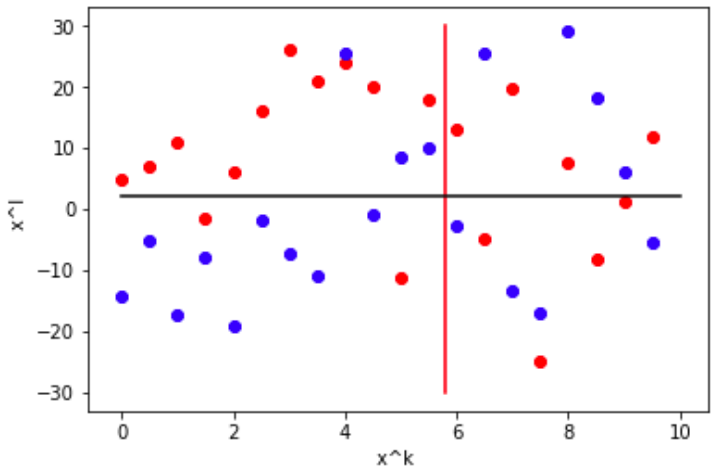

Рис. 4: Оптимальное разбиение выборки

Значит, любой объект обучающей или тестовой выборки, попавший в один из четырех квадрантов имеет новый признак — среднее систолическое давление по объектам обучающей выборки в соотвествующем квадранте.

Если нашлось такое оптимальное разбиение, при котором хотя бы один из квадрантов не имел объектов обучающей выборки, то в данном квадранте ставилось среднее систолическое давление по смежным квадрантам.

$$sys_l = \frac{1}{|m_l^{rt} + m_l^{lt}|} \sum_{i \in m_l^{rt} \bigcup m_l^{lt}} y_{tr(i)}$$

Можно догадаться, что количество новых признаков — $|P|$.

Теперь преобразуем нашу исходную обучающую и тестовую выборки, добавив новые признаки для каждого объекта.

$$\bar{x}_i = \vec{x}_i \bigcup_{p \in P} x_p^{new}, i = 1 \dots N$$

$$\bar{t}_i = \vec{t}_i \bigcup_{p \in P} t_p^{new}, i = N+1 \dots M$$

Тогда преобразованные выборки будут иметь вид:

$$S' = \{(\bar{x}_i, y'_{tr(i)}), i = 1 \dots N\}, \bar{x}_i \in R^{d+|P|}, y'_{tr(i)} \in \{0, 1\}$$

$$T' = \{(\bar{t}_i, y'_{tt(i)}), i = N+1 \dots M\}, \bar{t}_i \in R^{d+|P|}, y'_{tt(i)} \in \{0, 1\}$$

4 Вычислительные эксперименты

Рассмотрим выборки $S'_k(k = 1, 2, 3), T'_k(k = 1, 2, 3)$, отличающиеся $y_{0(k)} : y_{0(1)} = 115, y_{0(2)} = 120, y_{0(3)} = 125$. Напомним, это такое пороговое значение, при котором мы переходили от исходной задачи прогнозирования к задачам распознования нормального и повышенного систолического давления.

Будем подавать исходную и преоборазованную выборки на вход моделям RandomForestClassifier, LogisticRegression, Multi-layer Perceptron. И сравним качество прогноза по метрике AUC-ROC.

4.1 Первая выборка

$$S = \{(\vec{x}_i, y'_{tr(i)}), i = 1 \dots N\}, \vec{x}_i \in R^d, y'_{tr(i)} \in \{0, 1\}$$

$$T = \{(\vec{t}_i, y'_{tt(i)}), i = N + 1 \dots M\}, \vec{t}_i \in R^d, y'_{tt(i)} \in \{0, 1\}$$

$$S'_1 = \{(\bar{x}_i, y'_{tr(i)}), i = 1 \dots N\}, \bar{x}_i \in R^{d+|P|}, y'_{tr(i)} \in \{0, 1\}$$

$$T'_1 = \{(\bar{t}_i, y'_{tt(i)}), i = N + 1 \dots M\}, \bar{t}_i \in R^{d+|P|}, y'_{tt(i)} \in \{0, 1\}$$

где

$$y'_i = \begin{cases} 0, & \text{если } y_i < 125; \\ 1, & \text{если } y_i \geqslant 140. \end{cases}$$

Относили измерения, у которых систолическое давление меньше 125 к нормальной группе, а у которых не меньше 140 к повышенной. В таком случае объем обучающей выборки составил 832 измерений, тестовой 495. А дополнительное признаковое пространство состояло из 1718 признаков.

В первой таблице представлены средние значения качества прогноза AUC-ROC без признакового пространства и вместе с ним. Во второй таблице представлены средние значения качества прогноза AUC-ROC с дополнительными признаковым пространством и с методом отбора признаков — ElasticNet с параметрами: alpha=0.1, l1_rate. Sel_feats — количество отобранных признаков.

Таблица 1: 125(1)

Add_feat	Random_forest	Logreg	MLP
No	0.775	0.694	0.78
Yes	0.815	0.738	0.71

Таблица 2: 125(2)

l1_rate	Sel_feats	Random_forest	Logreg	MLP
0.1	251	0.776	0.736	0.763
0.2	142	0.768	0.77	0.764
0.3	109	0.769	0.769	0.773
0.5	92	0.762	0.74	0.757
0.6	77	0.76	0.729	0.767
0.7	69	0.755	0.724	0.761
0.9	61	0.763	0.746	0.766

Как видно из первой таблицы, дополнительное признаковое пространство увеличило качество прогноза у моделей LogisticRegression и RandomForestClassifier, но уменьшило у нейросети. Это связано с тем, что при добавлении новых признаков количество гиперпараметров значительно увеличилось, то есть возникало переобучение. Максимальная точность была достигнута у модели RandomForestClassifier с дополнительным признаковым пространством — 0.815. Что оказалось выше, чем максимальное значение любой модели без признакового пространства(на 0.035 выше в сравнении с RandomForestClassifier).

4.2 Вторая выборка

$$S = \{(\vec{x}_i, y'_{tr(i)}), i = 1 \dots N\}, \vec{x}_i \in R^d, y'_{tr(i)} \in \{0, 1\}$$

$$T = \{(\vec{t}_i, y'_{tt(i)}), i = N + 1 \dots M\}, \vec{t}_i \in R^d, y'_{tt(i)} \in \{0, 1\}$$

$$S'_2 = \{(\bar{x}_i, y'_{tr(i)}), i = 1 \dots N\}, \bar{x}_i \in R^{d+|P|}, y'_{tr(i)} \in \{0, 1\}$$

$$T_2' = \{(\bar{t}_i, y'_{tt(i)}), i = N+1 \dots M\}, \bar{t}_i \in R^{d+|P|}, y'_{tt(i)} \in \{0,1\}$$

где

$$y'_i = \begin{cases} 0, & \text{если } y_i < 120; \\ 1, & \text{если } y_i \geqslant 135. \end{cases}$$

Относили измерения, у которых систолическое давление меньше 120 к нормальной группе, а у которых не меньше 135 к повышенной. В таком случае объем обучающей выборки составил 721 измерений, тестовой 424. Дополнительное признаковое пространство состояло из 3803 признаков.

Таблица 3: 120(1)

Add_feat	Random_forest	Logreg	MLP
No	0.81	0.71	0.8
Yes	0.81	0.81	0.66

Таблица 4: 120(2)

l1_rate	sel_feats	Random_forest	Logreg	MLP
0.1	251	0.805	0.785	0.79
0.2	142	0.79	0.736	0.79
0.3	109	0.795	0.727	0.78
0.5	92	0.805	0.716	0.76
0.6	77	0.8	0.722	0.77
0.7	69	0.8	0.744	0.77
0.9	61	0.79	0.74	0.785

В данном случае новые признаки оказались существенно информативными для LogisticRegression, они подняли качество прогноза на 0.1, но при этом снизили у нейросети. RandomForestClassifier осталась неизменной.

4.3 Третья выборка

$$S = \{(\vec{x}_i, y'_{tr(i)}), i = 1 \dots N\}, \vec{x}_i \in R^d, y'_{tr(i)} \in \{0, 1\}$$

$$T = \{(\vec{t}_i, y'_{tt(i)}), i = N+1 \dots M\}, \vec{t}_i \in R^d, y'_{tt(i)} \in \{0, 1\}$$

$$S'_3 = \{(\bar{x}_i, y'_{tr(i)}), i = 1 \dots N\}, \bar{x}_i \in R^{d+|P|}, y'_{tr(i)} \in \{0, 1\}$$

$$T'_3 = \{(\bar{t}_i, y'_{tt(i)}), i = N+1 \dots M\}, \bar{t}_i \in R^{d+|P|}, y'_{tt(i)} \in \{0, 1\}$$

где

$$y'_i = \begin{cases} 0, & \text{если } y_i < 115; \\ 1, & \text{если } y_i \geqslant 130. \end{cases}$$

Относили измерения, у которых систолическое давление меньше 115 к нормальной группе, а у которых не меньше 130 к повышенной. В таком случае объем обучающей выборки составил 822 измерений, тестовой 434. Дополнительное признаковое пространство состояло из 6065 признаков.

Таблица 5: 115(1)

Add_feat	Random_forest	Logreg	MLP
No	0.795	0.689	0.785
Yes	0.77	0.769	0.66

Таблица 6: 115(2)

l1_rate	sel_feats	Random_forest	Logreg	MLP
0.01	1472	0.795	0.77	0.68
0.1	321	0.795	0.765	0.78
0.2	190	0.805	0.73	0.785
0.3	125	0.79	0.699	0.775
0.5	83	0.785	0.726	0.785
0.6	74	0.785	0.744	0.77
0.7	70	0.785	0.742	0.76
0.9	56	0.788	0.743	0.79

В третьем случае дополнительное признаковое пространство ухудшило модель RandomForestClassifier (снизило на 0.025), но также хорошо улучшило LogisticRegrission (повысило на 0.08). Нейросеть оставляет желать лучшего.

4.4 Прогноз систалического давления с использованием дополнительного признакового пространства

Итак, соберем все наши дополнительные признаковые пространства, созданные вышеописанным методом, в котором мы разбили исходную задачу прогнозирования на три разные задачи распознавания, отличающиеся пороговым значением систолического давления в каждой группе. В первой задаче количество новых признаков составило — 1718, во второй — 3803, в третьей — 6065. И проверим качество прогноза функцией R. Так как модель RandomForestClassifier в задачах распознавания оказалась наилучшей, то будем использовать RandomForestRegressor.

Таблица 7: Итоговая

Feat_space	Random_forest
No	0.41
1	0.425
2	0.42
3	0.425
1+2+3	0.43

Исходя из значений приведенных в таблице 7 можно сделать вывод, что дополнительное признаковое пространство увеличило качество прогноза исходной задачи, но на незначительную величину.

5 Заключение

В ходе работы были достигнуты следующие результаты:

- Был разработан новый способ повышения обобщающей способности на основании расширения признакового пространства, полученного с использованием двумерных моделей оптимальных разбиений.
- Разработанный метод позволил повысить точность распознавания групп с повышенным и нормальным систолическим давлением по синхронным сигналам, полученных с ЭКГ и ФПГ.

Список литературы

[1] О. В. Сенько, В. Я. Чучупал, А. А. Докукин *Неинвазивное оценивание уровня артериального давления с помощью кардиомонитора CardioQvark*// Матем. биология и биоинформ., 2017, том 12, выпуск 2, 536–545.

[2] О. В. Сенько *Перестановочный тест в методе оптимальных разбиений*// 2003, том 43, 1438-1447.

[3] Takayuki Sato T., Nishinaga M., Kawamoto A., Ozawa T., Takatsuji H. *Accuracy of a Continuous Blood Pressure Monitor Based on Arterial Tonometry*// Hypertension. 1993. V. 21. No. 6. P. 666–874.

[4] Chou P. *Optimal partitioning for classification and regression trees*// IEEE Trans. Pattern Analys. and Mach. Intelligente. 1991. V. 13. P. 340–354.

[5] О. В. Сенько , А. М. Морозов , А. В. Кузнецова , Л. Л. Клименко *Оценка эффекта множественного тестирования в методе оптимальных достоверных разбиений*//Машинное обучение и анализ данных. 2016, т.2, № 1.

[6] R. Tibshirani *Regression shrinkage and selection via the lasso*//J. R. Stat. Soc. 58, 267–288 (1996).

[7] Trevor Hastie, Robert Tibshirani, and Jerome Friedman *The Elements of Statistical Learning*//Springer New York Inc., New York, NY, USA, 2001.

[8] Lior Rokach, Oded Maimon *Data Mining With Decision Trees: Theory and Applications*//World Scientific Publishing Co., Inc. River Edge, NJ, USA, 2014.

[9] Fahimeh Ghasemian, Kamran Zamanifar, Nasser Ghasem-Aqaee, Noshir Contractor *Toward a better scientific collaboration success prediction model through the feature space expansion*//Scientometrics, (1-25). doi: 10.1007/s11192-016-1999-x, 2016.