УДК 159.9.072.53

Летков Юрий Владимирович, полиграфолог, разработчик СППРП «Сокол»
г. Вологда, math-polygraph@yandex.ru

Калафати Александр Юрьевич, полиграфолог, руководитель проекта полиграф «Триумф» г. Москва, psy.akalafati@gmail.com

Алгоритм классификации полиграмм Out_LIEr

Аннотация

В статье описаны математические основы работы алгоритма классификации полиграмм Out_LIEr с поисковыми тестами, тестами знания виновного и тестами вопросов сравнения, описаны результаты испытания алгоритма на рандомизированной выборке из архива подтверждённых дел Национального центра оценки достоверности информации (США) и на выборке полиграмм российских полиграфологов в сравнении с алгоритмом «Сокол» и алгоритмом «ChanceCalc».

Ключевые слова

Тест, дискриминантный анализ, диагностика лжи, классификация полиграмм, чувствительность, специфичность, ошибка классификации, вероятность правильной диагностики лжи, неопределённый результат.

Введение

Авторы исходят из того, что тестирование на полиграфе является диагностическим измерением [5]. А любая диагностика тесно связанна с классификацией. Отнести полиграмму, полученную в результате теста на полиграфе, к той или иной группе означает диагностировать требуемое свойство. Задача, которую перед полиграфологом ставит заказчик – определить правдивый или лживый ответ даёт опрашиваемый на вопросы заказчика. Валидизированные полиграфные тесты [13, 7] предполагают, что вопросы вводятся таким образом, что на них возможен либо утвердительный, либо отрицательный ответ, а сами проверочные вопросы должны быть предельно однозначны, поэтому полиграфолог должен диагностировать такое дихотомическое качественное свойство, даваемого ответа на проверочный вопрос (ПВ), как его правдивость/лживость. Мы исходим из следующей модели результатов теста: вопрос запускает у опрашиваемого непроизвольную оценку его значимости, то есть оценку насколько данный вопрос наиболее близок к его актуальной потребности, а в ходе проверки это, как правило, потребность в безопасности и в этом мы полностью согласны с Поповичевым С.В. [6,7].[1] Значит, в ситуации, когда опрашиваемый лжет на проверочный(е) вопрос(ы) его реакции должны быть в среднем сильнее чем реакции на служебные вопросы. А в ситуации, когда опрашиваемый отвечает правдиво на проверочные вопросы – его реакции должны быть в среднем слабее, чем реакции на служебные вопросы. Т.е. для лжеца самыми значимыми вопросами будут являться проверочные вопросы (ПВ), а для правдивого – служебные. Следовательно, диагностируя лживость/правдивость ответа полиграфолог одновременно диагностирует значимость/незначимость вопроса для опрашиваемого. Но между качественными свойствами вопроса – значимостью и лживостью есть отличие, которое заключается в том, что лживость ответа можно установить объективным способом (помимо полиграфа). А вот со значимостью гораздо сложнее, т.к. мысли людей мы пока читать не умеем, имея только косвенные инструменты в виде психологических тестов или субъективной оценки со стороны самих проверяемых. Получается, что в контексте полиграфной проверки объективно установить значимость можно только, установив из независимых источников информации правдивость/лживость ответа. Таким образом, все объективно подтвержденные полиграммы делятся на два класса: полиграммы тех тестов, в которых давались правдивые ответы на ПВ, и на полиграммы тех тестов, в ходе которых давались лживые ответы на ПВ. И в этом

[1] Мы хорошо понимаем, что в лабораторных исследованиях актуальная потребность будет отличаться, от реальных тестирований, но принципиального отличия наблюдаться не будет. А разница должна быть только в силе влиянии посторонних факторов и с какой легкостью они будут уводить внимание опрашиваемого от проверки.

случае имеется независимый способ, объективного разделения выборки полиграмм на две группы.

Одной из статистических процедур дискриминантного анализа является классификация наблюдений по группам. Методы статистической классификации предполагают получение одной или нескольких функций, которые дают возможность отнести конкретный объект к определённой группе. Эти функции называют классифицирующими. Они зависят от значений наблюдаемых переменных. В целях классификации полиграмм производят измерение, так называемых, метрических данных – показателей признаков физиологических реакций. После чего встаёт задача отнести полиграмму к одному из классов, произведя таким образом диагностику свойства лживости ответа на ПВ. Для решения этой задачи в рамках дискриминантного анализа необходимо найти классифицирующую функцию. В детекции лжи каждый из двух классов можно рассматривать как генеральную совокупность, а произведённые измерения, как случайное наблюдение над конкретной генеральной совокупностью. И тогда встаёт вопрос, из какой генеральной совокупности взяты наблюдения – метрические данные, измеренные на записанной полиграмме. В результате имеем две конкурирующие статистические гипотезы. Для решения задачи классификации определим понятие класса как генеральной совокупности, которая описывается некоторой случайной величиной, с некоторым законом распределения вероятностей и, соответственно, плотностью распределения вероятностей. В алгоритме Out_LIEr[2] (OL) классифицирующая функция получила название ‘силы реакции’ (СР), которая, являясь случайной величиной, для каждого класса имеет свой закон распределения вероятностей. СР определяется на основе модели предметной области, в которую составной частью входит решение первой задачи дискриминантного анализа – определения признаков реакций, которые дают возможность наилучшим образом дискриминировать результат полиграфного теста.

[2] Outlier в переводе с английского языка означает отклонение, отклоняющийся результат. Дословно, лежащий (lier) во вне (out). В тоже время слово lier – это ещё и лжец. И тогда название можно перевести как «Лжец – вон!».

Модель предметной области

1. Вегетативные реакции при ложном и правдивом ответе на вопросы теста различаются друг от друга [5];

2. Наиболее адекватно отражают физиологические процессы, связанные с типом даваемых ответов, так называемые «параметры Кирчера» [11] [12], и уменьшение амплитуды ФПГ;

3. **Out_LIEr CQT**

Статистически реакции, сопровождающие ложный ответ на проверочный вопрос, превышают реакции на среднее значение реакций на специальные (служебные) вопросы, используемые для сравнения; а реакции, сопровождающие правдивый ответ, меньше среднего значения реакций на вопросы сравнения.

4. **Out_LIEr CIT**

Статистически реакции, сопровождающие ложный ответ на стимул в ряду однотипных стимулов, превышают реакции на эти стимулы, и не отличаются от них при правдивом ответе.

Вычислительные процедуры Out_LIEr CIT

В тестах всех видов алгоритм для каждого, участвующего в анализе стимула, вычисляет значение СР, которая обладает следующими свойствами:

1. Зависит от «параметров Кирчера», которые включены в анализ;

2. Является монотонно возрастающей по каждой переменной в отдельности. Чем сильнее реакция по «параметру Кирчера», тем большее значение она принимает;

3. Всегда неотрицательна.

Значение СР вычисляется с использованием неравенства Чебышева, которое гласит, что для любой случайной величины вероятность её отклонения от своего среднего значения на величину большую чем некоторое число, не превосходит отношения её дисперсии к квадрату числа.

$$P(|z - Mz| \geq t) \leq \frac{Dz}{t^2} \quad (1)$$

Чем больше амплитуда КГР, подъём АД, меньше длина линии дыхания и амплитуда ФПГ, тем большее значение принимает эта случайная величина. В поисковых тестах и тестах знания виновного значение вычисляется для каждого стимула и производится ранжирование. Причем стимул с наибольшим значением СР ставится на первое место, пример можно видеть на рисунке 1.

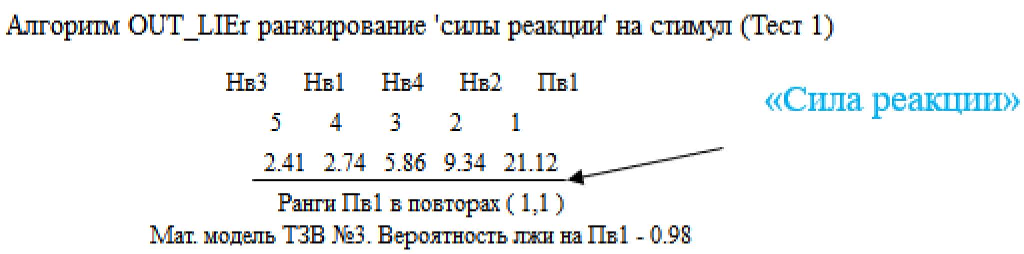

Рисунок 1. Ранжирование по силе реакции

Для теста знания виновного (ТЗВ) алгоритм вычисляет оценку вероятности ложности ответа/значимости стимула на проверочный вопрос. Для поискового теста (ПТ) вероятность лжи/значимости стимула на вопрос, который занял первое место в ранжированном ряду. Хотя ПТ и ТЗВ могут иметь схожую структуры и могут считаться близкими по своей логике, но с точки зрения информационного наполнения и принятия решения, основывающегося на их результатах – это разные тесты. В случае ПТ и ТЗВ они представляют ряд, в определённом смысле, однородных стимулов. Но в случае ПТ полиграфолог предполагает, что в контексте темы проверки для опрашиваемого один из стимулов теста является значимым. Полиграфолог не знает какой из стимулов является значимым, и в реальных, не лабораторных условиях, не может знать достоверно, что такой стимул имеется среди стимулов теста. Полиграфолог может только с большой долей вероятности полагать, что такой стимул включён в тест. Причина неопределённости может заключаться в том, что опрашиваемый может оказаться непричастным к тому событию, по которому производится проверка. Кроме того, опрашиваемый может забыть ту деталь события, которую полиграфолог пытается определить, или же может он не обратил внимания на эту деталь, и она отсутствует в долговременной памяти. С другой стороны, полиграфолог может допустить ошибку и не включить действительно значимый стимул в ряд, может допустить ошибку, основываясь на неверной исходной информации. Решение о значимости стимула полиграфологом принимается, если он видит достаточно сильную и повторяющуюся реакцию на определённый стимул. Категории «достаточно сильная» и «повторяющаяся» при экспертном принятии решения являются интуитивными и не определёнными, в смысле отсутствия критериев. В ТЗВ полиграфолог знает тот стимул, который действительно связан с событием, по которому производится разбирательство, и включив его в ряд подобных, проверяет известен ли он опрашиваемому лицу. Если реакция на проверочный вопрос стабильно превышает реакции на вопросы в ряду, принимается решение о значимости проверочного вопроса, и, соответственно, о том, что опрашиваемый скрывает факт своей информированности о детали, связанной с рассматриваемым событием. Как и в случае с ПТ опрашиваемый причастный к рассматриваемому событию может не обратить внимания на деталь, забыть её. Тест может быть составлен на основании ошибочной информации. Но в случае ТЗВ полиграфолог не оценивает априорно причастность опрашиваемого к событию как высокую, а с помощью теста пытается это установить апостериорно.

Таким образом классически ПТ используют для получения информации о важной детали события, предполагая высокую вероятность причастности опрашиваемого, а в случае ТЗВ определяют причастности опрашиваемого, основываясь на известной информации о событии. Можно говорить, что в некотором смысле, с помощью ПТ и ТЗВ решаются противоположные задачи. Поэтому математические модели этих тестов, которые можно использовать для построения правил решения, должны отличаться друг от друга. В частности, в ТЗВ, в отношении причастного, ошибка может произойти в том случае, если реакция на проверочный вопрос (ПВ) не будет наибольшей в ряду, в ПТ ошибка в отношении причастного может быть двух видов. В первом случае ошибка может произойти, когда полиграфолог не сможет уверенно выделить реакцию на действительно значимый стимул из ряда – ошибка «пропуска цели». Во-втором случае, при наличии в ряду действительно значимого стимула, полиграфолог может выделить как имеющий наибольшую реакцию, в действительности незначимый стимул – ошибка «ложной цели». В отношении непричастного опрашиваемого ошибка «пропуска цели» невозможна, в отличии от ошибки «ложной цели».

Вероятностно-статистическая модель поискового теста

f – наблюдаемое значение силы реакции на стимул с наибольшей силой реакции в ряду

A – событие «стимул значим» – стимул с наибольшей силой реакции

$\bar{A}$ – событие «стимул незначим», противоположное A

B – опрашиваемый причастен к событию

$\bar{B}$ – опрашиваемы не причастен к событию, противоположное B

Найдём вероятность того, что в результате теста имеется событие f[3] и событие A

$$P(A \cdot f) = P(f)P(A|f) = P(A)P(f|A) \tag{2}$$

Учитывая, что

$$P(f) = P(f \cdot A) + P(f \cdot \bar{A}) = P(A)P(f|A) + P(\bar{A})P(f|\bar{A}) \tag{3}$$

получим интересующую полиграфолога вероятность того, что стимул с наибольшей реакцией значим для опрашиваемого, при условии того, что произошло f.

[3] Под f необходимо понимать событие наблюдения «силы реакции» в некотором малом промежутке

$$P(A|f) = \frac{P(A)P(f|A)}{P(A)P(f|A)+P(\bar{A})P(f|\bar{A})} \quad (4)$$

Далее получим вероятность A

$$P(A) = P(A \cdot B) = P(B)P(A|B) = \alpha \cdot Ch \quad (5)$$

где $P(B) = \alpha$ – вероятность того, что опрашиваемый причастен к событию (входной параметр алгоритма – задаваемый «извне»), $P(A|B) = Ch$ – вероятность того, что вопрос с наибольшей СР действительно значим для причастного опрашиваемого.

Получим вероятность $\bar{A}$

$$P(\bar{A}) = P(B \cdot \bar{A}) + P(\bar{B} \cdot \bar{A}) = P(B \cdot \bar{A}) + P(\bar{B}) = P(B)P(\bar{A}|B) +$$
$$+(1-\alpha) = \alpha(1-Ch) + (1-\alpha) = 1 - \alpha \cdot Ch \quad (6)$$

Окончательно будем иметь

$$P(A|f) = \frac{\alpha \cdot Ch \cdot P(f|A)}{\alpha \cdot Ch \cdot P(f|A) + (1-\alpha \cdot Ch) \cdot P(f|\bar{A})} \quad (7)$$

Где $P(f|A)$ – вероятность f при условии того, что стимул значим, $P(f|\bar{A})$ – вероятность f при условии, что стимул незначим. Оценки этих вероятностей, так же, как и оценку Ch, можно получить из эмпирических исследований. Так как в тестах может быть использовано разное количество «параметров Кирчера» и разное количество стимулов, наблюдаемые значения f можно перевести в выборочные z-значения.

Вероятностно-статистическая модель теста знания виновного

f – наблюдаемое значение силы реакции на проверочный стимул

А – событие «проверочный стимул значим» – стимул с наибольшей силой реакции

$\bar{A}$ – событие «проверочный стимул незначим», противоположное А

B – опрашиваемый причастен к событию

$\bar{B}$ – опрашиваемы не причастен к событию, противоположное B

В условиях ТЗВ события А и B (как и противоположные им) являются эквивалентными: $A = B$, $\bar{A} = \bar{B}$, поэтому производя преобразования 2 – 3, аналогично случаю ПТ, получим

$$P(A|f) = \frac{P(B)P(f|B)}{P(B)P(f|B)+P(\bar{B})P(f|\bar{B})} \tag{8}$$

$$P(A|f) = \frac{\alpha P(f|B)}{\alpha P(f|B)+(1-\alpha)\cdot P(f|\bar{B})} \tag{9}$$

где $P(B) = \alpha$ – вероятность того, что опрашиваемый причастен к событию (входной параметр алгоритма). $P(f|A)$ – вероятность f при условии, что опрашиваемый причастен, $P(f|\bar{A})$ – вероятность f при условии, что опрашиваемый непричастен. Оценку двух последних вероятностей можно получить из эмпирических исследований, а значения f перевести в выборочные z-значения.

Получение эмпирических оценок вероятностей

Для получения оценок вероятностей в распоряжении авторов имелись 335 результатов тестов с карточками, которые проводились в качестве стимулиционно-адаптационного теста (САТ) в полевых условиях. Тест проводился следующим образом – заранее подготавливались 6 карточек с изображенными на них различными цветными фигурами, цифрами и словами. Пример карточек представлен на рисунке 2. Инструкция проверяемому давалась следующая – в рамках теста необходимо скрывать от полиграфолога своё знание карточки, т.е. отвечать на все вопросы, касающиеся карточки – нет. Даже если при этом будет названа та карточка, которая в настоящий момент находится у него.

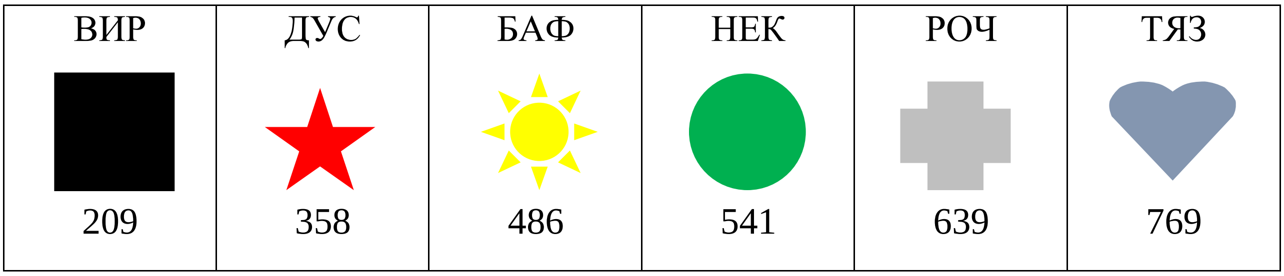

Рисунок 2. Карточки для САТ

При этом проверяемый мотивировался на тест следующим образом – если невозможно будет определить вытащенную им карточку, то тестирование будет перенесено на следующий день.

1. Вы готовы отвечать на мои вопросы?
2. На карточке, которую Вы вытащили, изображена фигура – Солнце?
3. На карточке, которую Вы вытащили, изображена фигура – Крест?
4. На карточке, которую Вы вытащили, изображена фигура – Круг?
5. На карточке, которую Вы вытащили, изображена фигура – Сердце?
6. На карточке, которую Вы вытащили, изображена фигура – Квадрат?
7. На карточке, которую Вы вытащили, изображена фигура – Звезда?

Для контрольной группы – ситуации, когда человек не владеет уликовой информацией было сделано следующее. Проверяемому сначала показывались 6 карточек, рубашкой вверх, так чтобы он не мог увидеть какую-либо из карточек. Далее объяснялось, что для проверки реакций, когда он говорит заведомую правду, ему нужно будет честно отвечать на все вопросы, касающиеся того, что может быть изображено на показанных только что карточках. Т.к. он карточки видел только рубашками вверх, то он про них сейчас ничего не знает, следовательно, может честно давать на все вопросы ответы – нет. И далее проводился точно такой же тест, как и для лжецов, как правило в 2 повтора – сначала назывались фигуры, а потом цвета.

Всего было собрано 233-х теста, где опрашиваемый лгал, отвечая на вопрос о том, что изображено на карточке, которую он выбрал. В 102 тестах опрашиваемые не лгали ни на один вопрос. Результаты этих тестов были использованы для оценки распределения, нормализованной СР методом ядерной оценки плотности с ядром Гаусса, и определения параметров распределения, если устанавливалось, что изучаемая случайная величина распределена нормально. Так же имелись результаты ТЗВ, проведённых по реальным уголовным делам, с результатом известным из независимых достоверных источников[4]. Из 215 таких тестов 88 тестов были проведены с лицами причастными к преступлению, и 128 с непричастными. Тесты с карточками включали в себя шесть проверочных вопросов и начинались с нейтрального вопроса, не содержащего проверочный стимул. Так как при такой структуре ПТ на первый проверочный вопрос в большинстве случаев наблюдается усиленная реакция, то в алгоритме предусмотрено «подавление» реакции – автоматическое её уменьшение путём деления наблюдаемого значения. Все тесты с карточками повторялись два раза. Полиграммы тестов CAT с карточками записывались на полиграфах «Триумф 2» и «Акикат» с использованием каналов:

1. Дыхательная активность (пневмодатчик)
2. Электродермальная активность ЭДА (многоразовые электроды)
3. Фотоплетизмограмма ФПГ
4. Оценка изменения кровотока в руке (манжета)

А также на полиграфе «Lafayette LX5000» с регистрацией признаков реакций в каналах:

1. Дыхательная активность (пневмодатчик)
2. ЭДА (многоразовые электроды)
3. ФПГ
4. Оценка изменения кровотока в руке (манжета)

[4] Из личного архива первого автора статьи

Тесты, проведённые по уголовным делам, не готовились специальным образом к проведению исследований. Среди них были тесты с одним, двумя и тремя повторами, с количеством проверочных вопросов в них от 4-х до 8-ми. Первый по смыслу проверочный вопрос в этих тестах обозначался как нейтральный, поэтому при их анализе «подавление» реакции не использовалось. Полиграммы этих тестов записывались на полиграфе «Диана 4» с использованием каналов:

1. Дыхание (тензорезисторы)

2. КГР (фильтр 1го порядка от ЭДА)

3. ФПГ

Результаты тестов с карточками использовались в качестве обучающей выборки для алгоритма, а выборка ТЗВ по уголовным делам использовалась в качестве тестовой.

Поисковые тесты

Зависимость результатов Out_LIEr CIT от признаков реакций в сердечно-сосудистой системе и количества повторов теста для тестов с причастными опрашиваемыми.

Выборки с ложными ответами

На первом этапе изучалась зависимость ранга, присваиваемого алгоритмом значимому вопросу по СР от используемых признаков реакций в сердечно-сосудистой системе (ССС). 233 теста с карточками, где присутствовал обман пересчитывались несколько раз. Каждый раз использовался признак реакции в канале Дыхание – грудное дыхание и признак в канале ЭДА – амплитуда ЭДА. В остальных случаях варьировались признаки реакций в сердечно-сосудистой системе следующим образом: подъём АД и спазм ФПГ, подъём АД и длина линии ФПГ, спазм ФПГ, длина линии ФПГ, подъём АД, подъём АД и спазм ФПГ с визуализацией «ФПГ вид КГР»[5]. В таблице 1 представлены результаты, полученные при пересчёте данных.

Таблица 1. Зависимость результатов ранжирования от используемых каналов

Ранг	АД + спазм ФПГ	АД + длина ФПГ	Спазм ФПГ	Длина ФПГ	АД	АД + ФПГ «вид КГР»
1	204(87,6%)	200(85,8%)	200(85,8%)	204(87,6%)	202(86,7%)	205(88,0%)
2	15	21	20	21	16	13
3	6	8	7	6	7	7
4	5	1	4	1	4	3
5	3	3	1	0	3	5
6	0	0	1	1	1	0

Статистически значимой разницы в зависимости ранжирования от используемых признаков реакции в сердечно-сосудистой деятельности не обнаруживается (Хи2, p-значение = 0,80). На рисунке 3 представлена непараметрическая оценка плотности распределения, нормализованной СР на стимул с рейтингом один для каждого случая. В таблице 2 помещены p-значения, вычисленные в результате тестов Колмогорова-Смирнова, которые проводились при попарном сравнении распределений. Тест Колмогорова – Смирнова обнаружил статистически значимое отличие на уровне 0,05 между распределениями только в двух парах: «Длина ФПГ» vs «АД + спазм ФПГ» и «Длина ФПГ» vs «АД + ФПГ “вид КГР”». В остальных случаях разница между распределениями если и имеется, то она мала, и с практической точки зрения, в задаче классификации/диагностирования эти распределения можно считать неразличимыми. Для использования в алгоритме была выбрано распределение «АД + ФПГ “вид КГР”» по причине лучшего результата в ранжировании, и по причине наиболее сглаженной, «естественной» непараметрической оценке распределения. Таким образом получена оценка распределения плотности вероятности случайной величины Fl = $P(f|A)$, которая используется в алгоритме для получения дискриминирующего правила.

[5] В этом виде для каждого отсчета по каналу ФПГ вычисляется амплитуда спазма, а дальше полученные значения инвертируются и представляются в виде графика, где спазм по ФПГ отображается в виде подъема графика. В этом случае оценка сила реакции происходит как амплитуда подъёма, т.е. представляет собой отношение между начальным условиями до реакции и максимальным уровнем реакции – подобно каналу ЭДА или КГР.

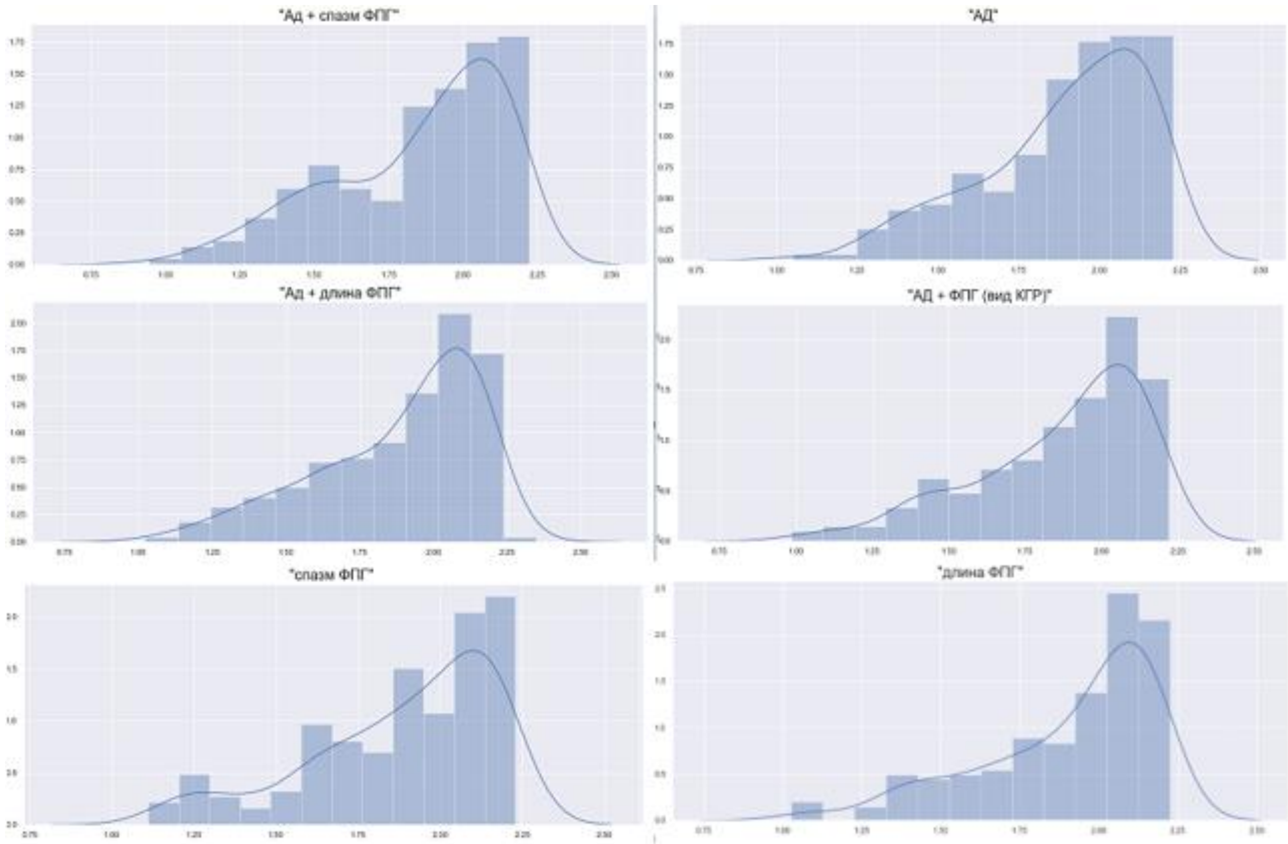

Рисунок 3. Оценка распределения нормализованной СР с разными признаками реакций в ССС

Таблица 2. p-значения, полученные в тестах Колмогорова – Смирнова при попарном сравнении распределений нормализованной СР с разными признаками реакции в ССС

	АД + спазм ФПГ	АД + длина ФПГ	Спазм ФПГ	Длина ФПГ	АД	АД + ФПГ «вид КГР»
АД + спазм ФПГ	1	0,48	0,31	0,04	0,28	0,59
АД + длина ФПГ		1	0,63	0,39	0,85	0,41
Спазм ФПГ			1	0,31	0,64	0,16
Длина ФПГ				1	0,20	0,02
АД					1	0,58
АД + ФПГ «вид КГР»						1

Учитывая результаты теста Колмогорова-Смирнова, при использовании алгоритма Out_LIEr CIT для оценки достоверности результата не рекомендуется использовать в качестве единственного признака реакции в ССС длину ФПГ. Для оценки распределения, нормализованной СР в тестовой выборке (тестов, проведённых с действительно причастными лицами) из неё, были взяты данные тестов, которые были повторены не менее 2-х раз с количеством стимулов в них не менее пяти. Таких тестов оказалось 64. Из этих тестов были отобраны те, в которых вычисленная СР на проверочный вопрос получила 1 место (наибольшая СР в ряду), таких тестов оказалось 52. На рисунке 4 можно видеть непараметрическую оценку распределения нормализованной СР, полученную на данных этих тестов, в сравнении с распределением обучающей выборки.

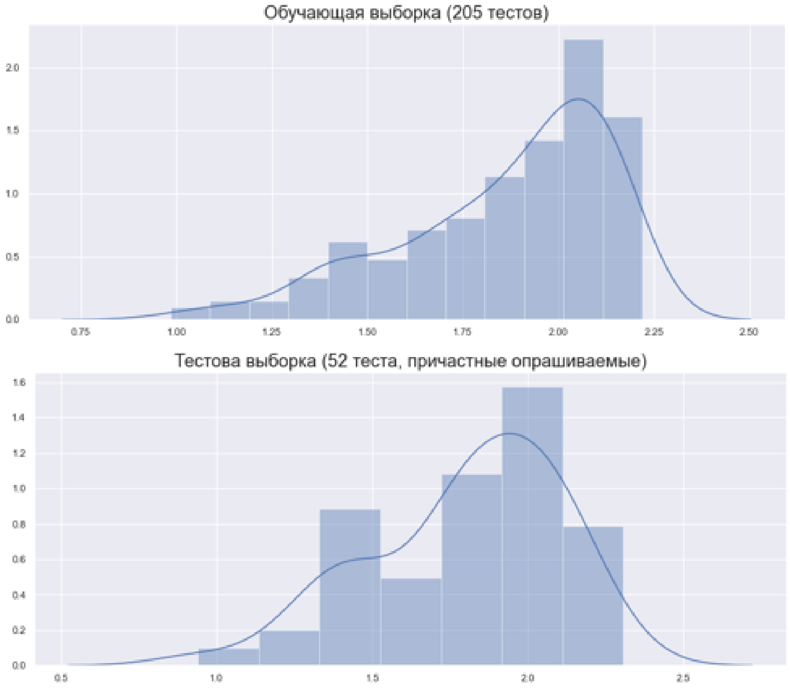

Рисунок 4. Оценка распределения нормализованной СР на значимые стимулы в обучающей и тестовой выборках

Тест Колмогорова-Смирнова не обнаружил статистически значимого различия между распределениями нормализованной СР на значимый стимул в обучающей и тестовой выборке (p-значение = 0,103). В таблице 3 размещены частоты полученных рангов СР на значимые стимулы в тестах с карточками, в которых присутствовал обман, и на проверочные вопросы в ТЗВ с причастными опрашиваемыми.

Таблица 3. Частоты рангов СР в выборках для тестов с повторами

	1	2	3	4	5	6	Всего
Обучающая выборка	205 (88%)	13	7	3	5	0	233
Тестовая выборка	52 (81%)	6	3	1	2	0	64

Тест Хи2 статистически значимой разницы в распределении рангов в обучающей и тестовой выборке не обнаружил (p-значение = 0,367). Частоты распределения рангов в тестовой и обучающей выборке были пересчитаны для одного повтора теста, результаты пересчёта в таблице 4.

Таблица 4. Частоты рангов СР для одного повтора теста

	1	2	3	4	5	6	Всего
Обучающая выборка	355 (70%)	82	31	22	12	8	510
Тестовая выборка	115(72%)	23	13	5	4	0	160

Тест Хи2 статистически значимой разницы в распределении рангов в обучающей и тестовой выборке не обнаружил (p-значение = 0,536). Из таблиц 3 и 4 можно видеть, что в случае, когда тесты повторялись (более одного повтора), и в обучающей, и в тестовой выборках увеличивался процент получения СР на значимый стимул 1 ранга с 70-72% до 81-88%. Учитывая полученный результат, необходимо отметить, что при использовании алгоритма Out_LIEr CIT для оценки достоверности результата рекомендуется делать не менее двух повторов ПТ.

Выводы:

1. При использовании алгоритма Out_LIEr результаты ранжирования по СР в тестах МВСИ от используемых признаков реакции в ССС (от их комбинации) не зависят;
2. Оценку распределения плотности вероятности Fl можно получить из ядерной оценки. Оптимальным вариантом будет использовать для этого данные с показателями таких признаков реакции в ССС как «АД» и «ФПГ вид КГР»;
3. Между распределениями вероятностей нормализованной силы реакции на обучающей выборке и тестовой выборке статистически значимого различия нет;
4. Если делать выводы только по рангу СР, т.е. определять как значимый стимул, занявший первое место по СР, то примерно в 85% случаев, когда

тест проводится с причастным опрашиваемым, будет правильный результат и в 15% случаев будет принято ошибочное решение (оценки получены путём аппроксимации частотного ряда степенной функцией);

5. При использовании алгоритма Out_LIEr для оценки достоверности диагностического заключения в тестах МВСИ не рекомендуется использовать признак «длина ФПГ» как единственный признак реакции в ССС;
6. При использовании алгоритма Out_LIEr для оценки достоверности диагностического заключения необходимо тест МВСИ повторять не менее двух раз.

Выборка с правдивыми ответами

Данные 102 тестов с карточками, в которых отсутствовали ложные ответы, были использованы в качестве обучающей выборки алгоритма. В них отсутствовал реальный проверочный и значимый стимул, поэтому распределение частот ранга проверочного-значимого стимула изучалось на тестовой выборке. Из имеющейся выборки, которая состояла из тестов с разным количеством повторов, и разным количеством стимулов, были отобраны отдельные повторы тестов как самостоятельные тесты. В этой выборке имеется большая доля тестов, которые проводились без повторов. Объясняется это тем, что во-первых, нередко опросы проводились в условиях дефицита времени; во-вторых, нередко для проверки подготавливалось несколько ТЗВ, с помощью каждого из которых решалась одна задача – определение причастности; в третьих, с опрашиваемым проводился тест вопросов сравнения (ТВС), и при высокой достоверности его результата не имело смысла увеличивать количество повторов ТЗВ. В таблицах 5,6,7 находятся частоты рангов проверочного вопроса (ПВ) для тестов с 4, 5 и 6 – ю стимулами. Предполагалось, что из-за эффекта «лесенки» (наличие этого эффекта показано в Приложении А) – опрашиваемый сильнее реагирует на первые стимулы в ряду с последующим убыванием СР, и по причине того, что ПВ чаще ставится ближе к середине ряда и никогда с краёв, распределение частот рангов СР на ПВ у правдивых будет отличатся от равномерного. У правдивых реже должен наблюдаться максимальный (1-ый) ранг и минимальный ранг, чем промежуточные ранги.

Таблица 5. Частоты рангов ПВ тестов с одним повтором и 4-мя стимулами

Ранги	1	2	3	4	Всего
Частоты	7	9	12	11	39

Таблица 6. Частоты рангов ПВ тестов с одним повтором и 5-ю стимулами

Ранги	1	2	3	4	5	Всего
Частоты	10	34	25	35	18	122

Таблица 7. Частоты рангов ПВ тестов с одним повтором и 6-ю стимулами

Ранги	1	2	3	4	5	6	Всего
Частоты	5	5	6	7	15	1	39

Для теста с 4-мя стимулами Хи2 не отверг предположение о равномерном распределении рангов ПВ (p-значение = 0,679). Объяснить это можно тем, что количество стимулов в тесте мало для того, чтобы эффект «лесенки» проявил себя достаточно сильно. Но для тестов с 5-ю и 6-ю стимулами предположение о равномерном распределении отвергается на уровне значимости <0,001 и 0,025 соответственно. Далее были отобраны данные тестов с двумя повторами с 4-мя и 5-ю стимулами. Тестов с двумя повторами и шестью стимулами в выборке оказалось всего 7, что недостаточно для проведения статистического теста. В таблицах 8 и 9 частоты рангов ПВ тестов с двумя повторами. Для таких тестов ожидается, что распределение рангов на ПВ вопрос будет равномерным, так как порядок стимулов при повторе меняется. Как правило те, что находились в конце ряда, переставляются на первые места. В ПТ наиболее вероятный стимул и действительно значимый, либо ставится ближе к середине ряда, либо может с равной вероятностью оказаться на любом месте ряда в обоих повторах теста. Что так же должно компенсировать эффект «лесенки».

Таблица 8. Частоты рангов ПВ тестов с двумя повторами и 4-мя стимулами

Ранги	1	2	3	4	Всего
Частоты	1	6	4	4	14

Таблица 9. Частоты рангов ПВ тестов с двумя повторами и 5-ю стимулами

Ранги	1	2	3	4	5	Всего
Частоты	6	12	7	7	7	39

Результаты статистических тестов позволили не отвергнуть предположение о компенсации «лесенки» с помощью повтора. p-значение = 0,571 для тестов с 5-ю стимулами и p-значение = 0,796 с 4-мя стимулами. Факт равномерного распределения ранга СР на ПВ означает, что если опрашиваемый отвечает правдиво на все вопросы в ряду, то ПВ с одинаковой вероятностью может занять по СР любое место от первого до равного количеству стимулов в ряду. Этот факт означает, что вероятность того, что у правдивого опрашиваемого проверочный вопрос будет иметь наибольшую силу реакции равна 1/n. Например, для теста с 4 стимулами эта вероятность равна ¼ - 0,25; с 5 стимулами – 1/5 (0,2) и т.д. Этот факт используется в СППРП «Сокол», поэтому при её использовании для оценки достоверности диагностического заключения необходимо повторять тест не менее двух раз. Эта же рекомендация была дана выше в случае выборки с ложными ответами.

Факт равномерного распределения ранга был проверен на обучающей выборке ПТ с правдивыми ответами и с двумя повторами. В случае ПТ при отсутствии значимого стимула, при верности предположения о равномерном распределения ранга, любой из стимулов с одинаковой вероятностью может занять наивысший ранг. В обучающую выборку включены ряды с шестью стимулами. Первый задаваемый вопрос в ряду (R1), который во всех повторах должен быть одним и тем же, является «бросовым» – подбирается так, чтобы заранее было известно, что он не может быть значимым. В ряду он обозначался как релевантный и включался в анализ с «подавлением реакции». Если такой вопрос получает наивысший ранг, то можно диагностировать, что ни один стимул ряда не является значимым. В силу особенности этого стимула он исключался из подсчёта. Для остальных стимулов ряда было подсчитано сколько раз каждый из них получал наивысший ранг (имел наибольшую силу реакции). В 6 случаях наивысший ранг получил R1, в остальных 96 случаях наивысший ранг получили не «бросовые» стимулы. Результат подсчёта находится в таблице 10.

Таблица 10. Частоты получения наивысшего ранга стимулами ряда

Стимул	R2	R3	R4	R5	R6	Всего
Частота	21	21	17	19	18	96
Теоретическая частота	19,2	19,2	19,2	19,2	19,2	96

Результат теста Хи2 (p-значение = 0,955) не отверг предположение о равномерном распределении первого ранга среди стимулов ряда. Таким образом вероятность того, что любой из стимулов (кроме «бросового») в ПТ получить наивысший ранг (будет иметь наибольшую СР) равна 0,2. P= 1/5, 96·0,2 = 19,2. Далее была получена ядерная оценка плотности распределения вероятностей на выборке из 102 тестов (обучающая выборка) с правдивыми ответами. Tr – обозначение соответствующей случайной величины. На рисунке 5 изображена ядерная оценка плотности Tr. На рисунке 6 график «Квантиль-квантиль» Tr – нормальное распределение.

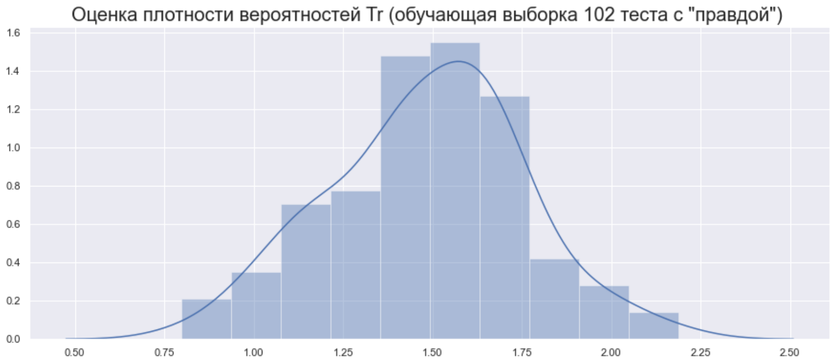

Рисунок 5. Оценка плотности распределения Tr (обучающая выборка 102 теста)

Рисунок 6. График «квантиль-квантиль» (обучающая выборка 102 теста)

Визуальный анализ позволяет сделать предположение о том, что Tr имеет нормальное[6] распределение. Результат теста Шапиро-Уилка на нормальность (p = 0,787), и теста Андерсона-Дарлинга (статистика теста – 0,385) на нормальность позволили не отвергнуть предположение о принадлежности Tr (обучающая выборка) к семейству нормальных распределений. Также была получена ядерная оценка плотности распределения вероятностей на выборке из 28 тестов (обучающая выборка, значимый стимул не получил первое место) из 233 тестов с присутствием ложного ответа. Tr28 –обозначение соответствующей случайной величины. На рисунке 7 –ядерная оценка плотности, на рисунке 8 – график «квантиль-квантиль».

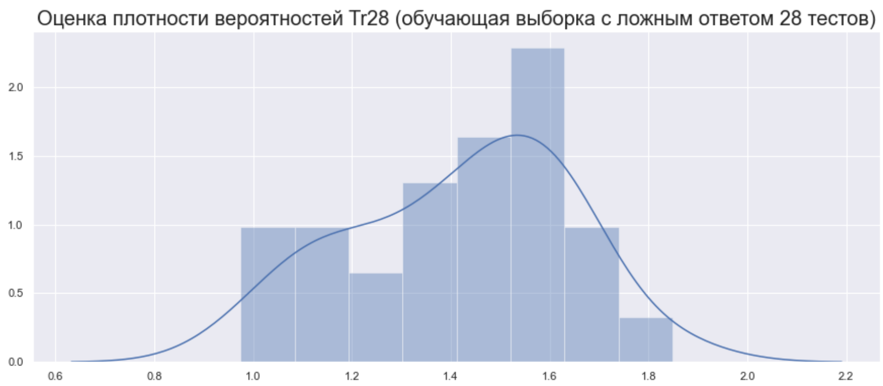

Рисунок 7. Оценка плотности распределения Tr28

[6] Здесь и далее в главе о поисковых тестах речь идёт об усечённых нормальных распределениях

Рисунок 8. График «квантиль-квантиль» (обучающая выборка 28 тестов)

Результаты теста выборки из Tr28 Шапиро-Уилка на нормальность (p = 0,402) и тест Андерсона-Дарлинга (статистика теста – 0,435) позволяют предположить, что Tr28 так же, как и Tr принадлежит семейству нормальных распределений. Результаты тестов Колмогорова-Смирнова на однородность (p = 0,420) и t-теста Стьюдента на равенство генеральных средних (p = 0,233) позволяют предполагать, что если распределения Tr и Tr28 и имеют различные параметры, то различие это мало – не обнаруживается статистическими тестами, и в целях классификации этим различием можно пренебречь. В соответствии с вероятностно-статистической моделью ПТ, с целью построения распределения случайной величины Tr(f) = $P(f|\bar{A})$, и использования его в дискриминирующем правиле алгоритма, к выборке тестов с правдивыми ответами были добавлены результаты тестов из выборки с ложными ответами, в которых значимый стимул не получал 1-ый ранг, и таким образом были получены оценки параметров распределения. Для оценки распределения случайной величины Tr на тестовой выборке из выборки ТЗВ проведённых с непричастными были взяты 38 тестов с двумя повторами и пятью стимулами. Из выборки ТЗВ с причастными были взяты данные 7-ми тестов, в случае которых проверочный стимул не получил 1-ый ранг. На рисунке 9 ядерная оценка плотности вероятности Tr, на рисунке 10 график «квантиль-квантиль».

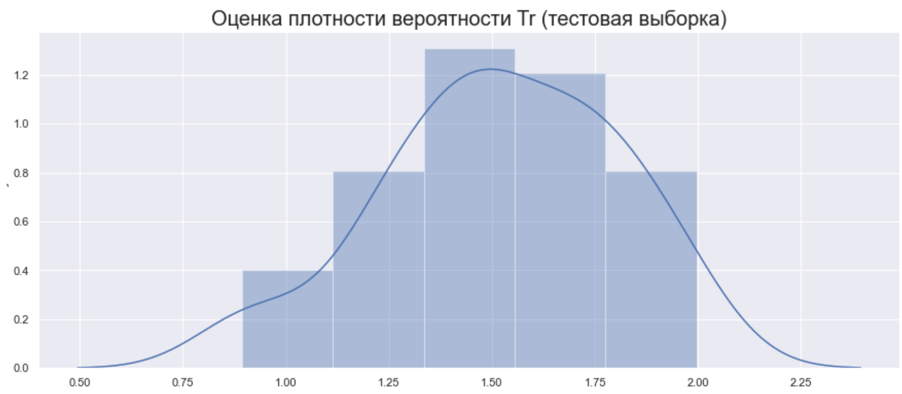

Рисунок 9. Оценка плотности вероятности Tr (тестовая выборка)

Рисунок 10. График «квантиль-квантиль» (тестовая выборка)

Тесты Шапиро-Уилка на нормальность (p = 0,209), Андерсона-Дарлинга (статистика теста – 0,386) на нормальность позволили не отвергнуть предположение о принадлежности Tr (тестовая выборка) к семейству нормальных распределений. Результат t-теста (критерий Стьюдента, p = 0,277) даёт возможность полагать равенство генеральных средних случайных величин Tr (обучающая выборка) и Tr (тестовая выборка). Результат t-теста и теста Колмогорова-Смирнова на однородность (p = 0,292) позволяют сделать вывод, что если параметры этих распределений и имеют различие, то различие это мало и статистически не значимо. Поэтому в целях классификации ими можно пренебречь.

Выводы:

1. Для непричастных в одном повторе ТЗВ распределение 1-го ранга среди стимулов ряда отличается от равномерного. Имеется эффект «лесенки» – 1-вый ранг чаше получают первые реализуемые стимулы в ряду. При повторе теста происходит «сглаживание» и распределение 1-го ранга становится равномерным;
2. Для непричастных в ПТ при двух повторах теста все стимулs получают 1-ый ранг с одинаковой вероятностью, равной величине обратной к числу стимулов в ряду, без учёта «бросового», если он включён в анализ с использованием «подавления»;
3. Функция распределения вероятностей нормализованной силы реакции на обучающей выборке непричастных (Tr) с данными тестов из обучающей выборки причастных, в которых значимый вопрос не получил 1-го ранга, принадлежит семейству нормальных распределений;
4. Функция распределения вероятностей нормализованной силы реакции на тестовой выборке непричастных с данными тестов из тестовой выборки причастных, в которых значимый вопрос не получил 1-го ранга, принадлежит семейству нормальных распределений;
5. Различия в вышеуказанных функциях распределения, если и имеют место, то не являются значимыми в целях классификации;
6. Поисковые тесты необходимо повторять не менее двух раз;
7. В тестах должно быть 5 (без «подавления») или 6 стимулов (с «подавлением»).

Испытание алгоритма на поисковых тестах

На рисунке 11 изображены оценки функций плотности распределения вероятностей, которые в алгоритме используются для классификации полиграмм поисковых тестов.

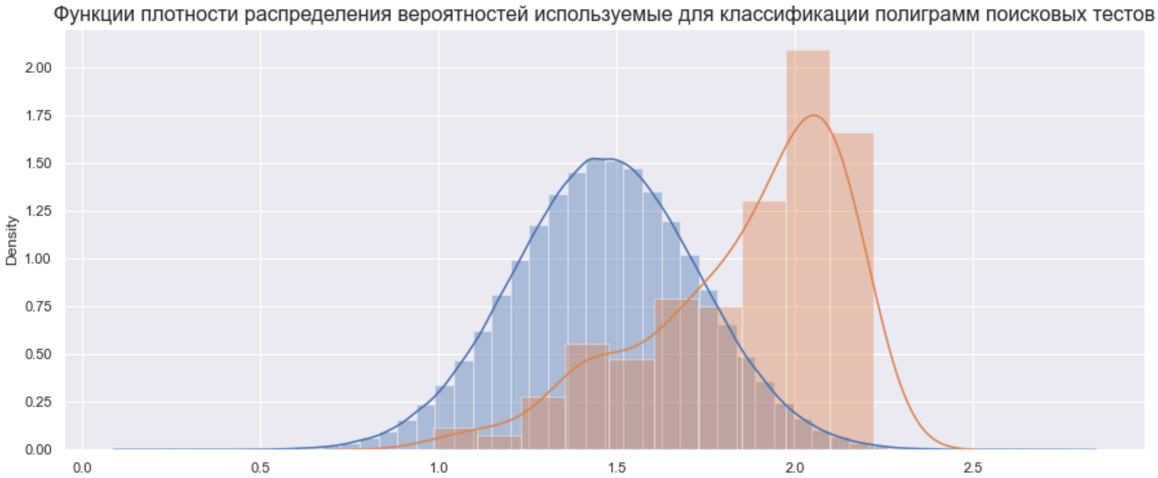

Рисунок 11. Оценки функций плотности распределения вероятностей, используемых для классификации полиграмм поисковых тестов

Из рисунка можно видеть, что моды, и центры распределений расположены относительно близко друг к другу, и поэтому стоит ожидать большой процент ошибочных случаев классификации, который необходимо уменьшить за счёт увеличения процента неопределённых результатов. Испытание алгоритма производилось на тестовых выборках. ТЗВ при этом рассматривались как ПТ. Результаты испытаний на выборках причастных и непричастных, а процентном соотношении представлены в таблицах 11, 12 соответственно. В выборки были включены тесты с не менее чем с пятью стимулами и двумя повторами. Объём выборки причастных 65 тестов, объём непричастных – 48 тестов. Принять решение алгоритм рекомендует при превышение оценки вероятности 0,75 принадлежности к классу. Это решение об отнесении полиграммы к классу полиграмм, в которых стимул с максимальным рангом (1-ым) является значимым для опрашиваемого. Низкий порог принятия решения должен обеспечить определённый баланс между процентом ошибок классификации и процентом неопределённых результатов.

Принятие решения по рангу означает определение значимого вопроса по наибольшей реакции на стимул. В этом случае отсутствуют неопределённые результаты, когда алгоритм даёт рекомендацию: «Значимость ни одного из стимулов не установлена». В случае причастных будет велика ошибка ложной цели – порядка 12-19%, а в случае непричастных ошибка будет составлять все 100%.

Таблица 11. Результаты алгоритма на выборке причастных

Причастные		**Правильное решение %**	**Ошибка «ложной цели» %**	**Неопределённый результат %**
Принятие решения по рангу		**81,5**	**18,5**	**0**
Априорная вероятность	**1**	**61,5**	**3,1**	**35,4**
	0,95	**60,0**	**3,1**	**36,9**
	0,9	**55,4**	**3,1**	**41,5**
	0,5	**36,9**	**1,5**	**61,5**

Таблица 12. Результаты алгоритма на выборке непричастных

Непричастные		**Правильное решение %**	**«Ошибка ложной цели» %**	**Неопределённый результат %**
Принятие решения по рангу		**0**	**100**	**0**
Априорная вероятность	**1**	**0**	**43,75**	**56,25**
	0,95	**0**	**37,5**	**62,5**
	0,9	**0**	**31,25**	**68,75**
	0,5	**0**	**10,42**	**89,58**

При использовании оптимального Байесовского решающего правила, в соответствии с которым даёт рекомендации алгоритм, в случае причастных ошибка сокращается более чем в шесть раз, но при этом возникает ситуация, когда «значимость ни одного из стимулов не установлена». Процент неопределённого результата зависит от используемой априорной вероятности (АВ). Равную единице АВ следует выбирать, когда достоверно известно, что значимый стимул имеется в тестовом ряду. Например, тест с карточками соответствует этой ситуации. Или, если достоверно известно, что опрашиваемый причастен к событию и полиграфолог включил в стимульный ряд все возможные варианты. Ситуации, когда имеются основания предполагать причастность опрашиваемого к событию, но нет информации, которая бы позволяла утверждать это с высокой долей вероятности, соответствует АВ = 0,5. При таком значении параметра ошибка с непричастными минимизируется. Проводить же ПТ с опрашиваемым, который скорее не имеет отношения к событию, чем имеет, чревато тем, что с большой долей вероятности будет допущена ошибка, если не выставить АВ близкой к нулю. Но тогда почти наверняка результат будет неопределённым. Иначе говоря, проводить ПТ с непричастными не имеет смысла. Проводить ПТ допустимо тогда, когда имеется информация, которая говорит о большой вероятности причастность опрашиваемого, например, это может быть результат ТВС. Для изучения результатов алгоритма моделировались ситуации с различным соотношением доли причастных и не причастных. Во всех случаях отбор тестов для моделирования производился случайным образом. В таблице 13 результат моделирования ситуации 50/50. В этом случае, при принятии решения по максимуму реакции, ошибка «ложной цели» в полтора раза превосходит количество принятых правильных решений. При использовании решающего правила алгоритма ситуация принимает обратный характер, но за счёт не приемлемого процента не определённых результатов. В неопределённом результате ПТ скрываются ошибка «пропуска цели», совершаемая с причастными опрашиваемыми, и те случаи, которые могли бы увеличить количество ошибок «ложной цели», совершаемой и с причастными и не причастными, если бы было принято решение. Как уже было отмечено выше ситуация 50/50 является не приемлемой для использования ПТ. В таблице 14 находятся результаты моделирования ситуации 90/10. Ситуация 90/10 более приемлема для использования ПТ. При использовании решающего правила алгоритма количество ошибок «ложной цели», относительно принятия решения по рангу, уменьшается почти в 5 раз, но процент не определённых случаев равен почти 45%. В таблице 15 можно увидеть результаты моделирования ситуации 95/5.

Таблица 13. Результаты алгоритма в ситуации 50/50

50/50		Правильное решение	Ошибка «ложной цели»	Неопределённый результат 74 / 77,08%		Всего
				Ошибка «пропуска цели»»	Избежание ошибки «ложной цели»	
Принятие решения по рангу		39	57	0	0	96
		40,6%	59,4%	0%	0%	100%
АВ	0,5	16	6	23	51	96
		16,67%	6,25%	23,96%	53,12%	100%

Таблица 14. Результаты алгоритма в ситуации 90/10

90/10		Правильное решение	Ошибка «ложной цели»	Неопределённый результат 32 / 44,44%		Всего
				Ошибка «пропуска цели»	Избежание ошибки «ложной цели»	
Принятие решения по рангу		53	19	0	0	72
		73,6%	26,4%	0%	0%	100%
АВ	0,90	36	4	17	15	72
		50,00%	5,56%	23,61%	20,83%	100%

Результаты алгоритма в этом случае несущественно отличается от ситуации 90/10. Но в целом результаты алгоритма в такой ситуации выглядят лучше. Максимально хороший результат алгоритм достигает в случае 100/0 – таблица 16. Так же была смоделирована ситуация использования параметра АВ равного единице в ситуации 90/10. Из таблицы 17 можно видеть, в этом случае результат в целом незначительно, но в лучшую сторону отличается от ситуации использования правильного значения АВ. Но почти наверняка это эффект маленьких выборок, и вопрос требует уточнения, хотя уже можно говорить, что использование значения АВ равного 1 вместо 0,9 не значительно ухудшит результат алгоритма. В целом завышение АВ увеличит процент правильных решений, но приведёт к росту ошибок «ложной цели».

Таблица 15. Результаты алгоритма в ситуации 95/5

95/5		Правильное решение	Ошибка «ложной цели»	Неопределённый результат 28 / 40,58%		Всего
				Ошибка «пропуска цели»	Избежание ошибки «ложной цели»	
Принятие решения по рангу		53	16	0	0	69
		76,81%	23,19%	0%	0%	100%
АВ	0,95	39	2	14	14	69
		56,52%	2,90%	20,29%	20,29%	100%

Таблица 16. Результаты алгоритма в ситуации 100/0

100/0		Правильное решение	Ошибка «ложной цели»	Неопределённый результат 23 / 35,38%		Всего
				Ошибка «пропуска цели»	Избежание ошибки «ложной цели»	
Принятие решения по рангу		53	12	0	0	65
		81,54%	18,46%	0%	0%	100%
АВ	1	40	2	13	10	65
		61,54%	3,08%	20,00%	15,38%	100%

Таблица 17. Результаты алгоритма в ситуации 90/10, АВ = 1

90/10		Правильное решение	Ошибка «ложной цели»	Неопределённый результат 28 / 38,89%		Всего
				Ошибка «пропуска цели»	Избежание ошибки «ложной цели»	
Принятие решения по рангу		53	19	0	0	72
		73,6%	26,4%	0%	0%	100%
АВ	1	40	4	13	15	72
		55,55%	5,56%	18,05%	20,83%	100%

Из всех рассмотренных вариантов, не беря в расчёт ситуацию 100/0,которая на практике реализуется либо крайне редко, либо только в специально организованных условиях, наилучшие результаты получены в ситуации 95/5. Учитывая, что чаще всего на практике имеется шанс, что опрашиваемый на самом деле не причастен к событию, или же допущена ошибка в подборе стимулов теста, можно рекомендовать использовать решающее правило алгоритма с АВ = 0,95.

Тест знания виновного

Выборка тестов с правдивыми ответами

Построение дискриминирующих распределений для ТЗВ отличается от той же задачи в случае ПТ. Если во втором случае это делается для стимулов с наибольшей СР, то в первом это делается для ПВ, сила реакции на который может занять любое место в ранжированном ряду – от первого до последнего. По этой причине плотность распределения вероятностей случайной величины Tr (обозначение вычисляемой нормализованной СР для правдивых опрашиваемых) отличается от нормального. На рисунке 12 оценка функции плотности распределения вероятностей Tr для ТЗВ.

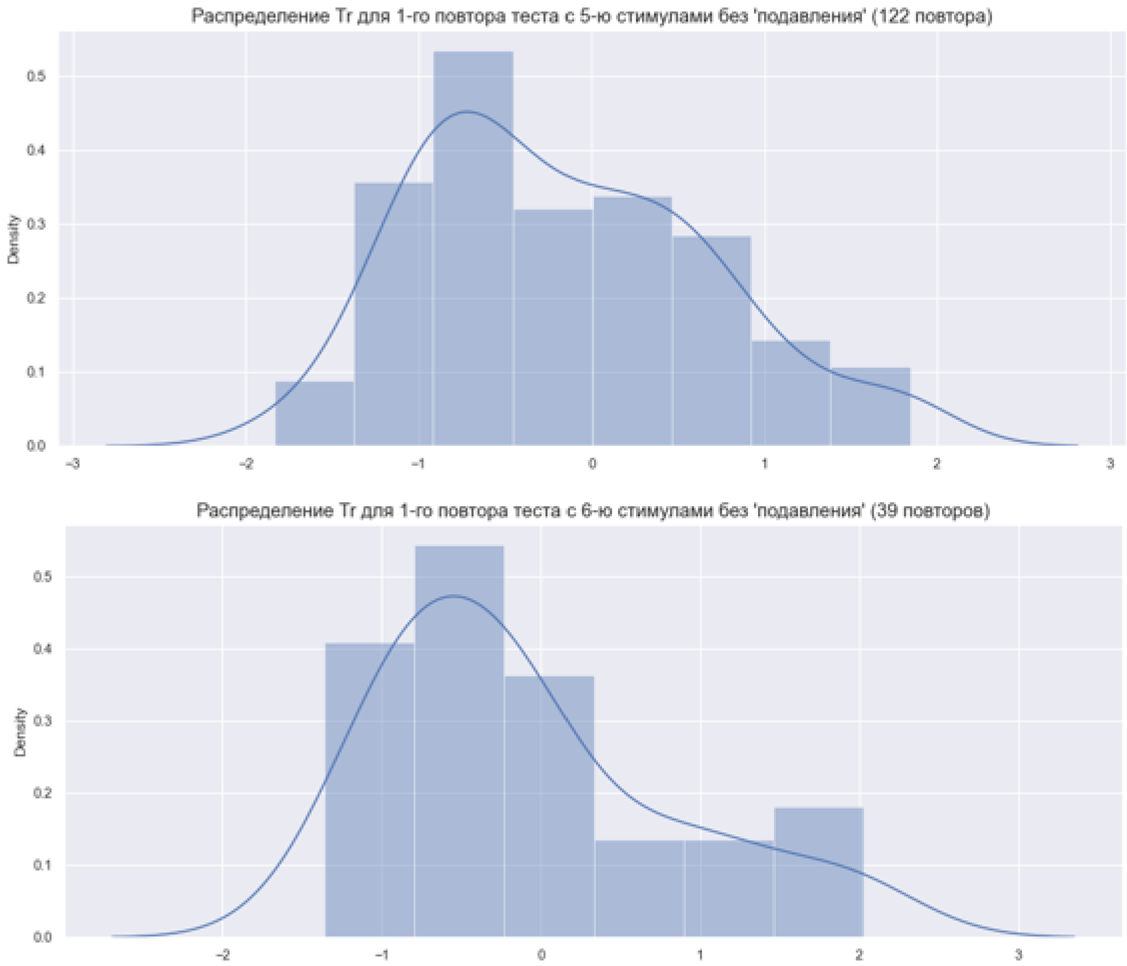

Рисунок 12. Распределения Tr для тестов с 5-ю и 6-ю стимулами

Результат теста Колмогорова-Смирнова на однородность (p = 0,506) позволяет сделать вывод о том, что Tr для тестов с 5-ю и 6-ю стимулами без ‘подавления’ распределена одинаково и данные могут быть объединены. На рисунке 13 имеется оценка распределения Tr полученная на тестах с 5-ю и 6-ю повторами.

Рисунок 13. Распределение Tr

Визуальный анализ распределения Tr показывает, что оно далеко от нормального, что подтверждается тестами Шапиро-Уилка и Андерсена-Дарлинга (p <0,001; statistic = 2,05). Этим оно отличается от распределения Tr в случае поисковых тестов, которое подчиняется нормальному закону. Поэтому встаёт вопрос о том, что определяет такую форму распределения Tr для ТЗВ. Для разрешения этой проблемы было сгенерировано несколько десятков виртуальных распределений Tr для ТЗВ из выборки поисковых тестов «на правду». Такой тест можно рассматривать как ТЗВ, если известно заранее, какую карточку взял опрашиваемый, но при этом в неё не смотрит и не знает, что на ней изображено и написано. Перед таким тестом опрашиваемый берёт любую из пяти карточек с равной вероятностью равной 1/5. Поэтому сначала генерировались виртуальные выборки Tr для ТЗВ, с назначением проверочным вопросом любого из пяти ранжированных стимулов теста (реакция на первый стимул в таком тесте ‘подавляется’ и поэтому первый стимул в тесте должен быть с заведомо правдивым ответом). При генерации виртуальной выборки первый ‘подавленный’ стимул исключался из анализа, таким образом в анализе участвовали только пять стимулов. Для представленной на рисунке 14 виртуальной выборки и реальной выборки ТЗВ тест на однородность Колмогорова-Смирнова дал p-значение меньше 0,05. Это позволяет отвергнуть предположение, что они обе извлечены из одного распределения.

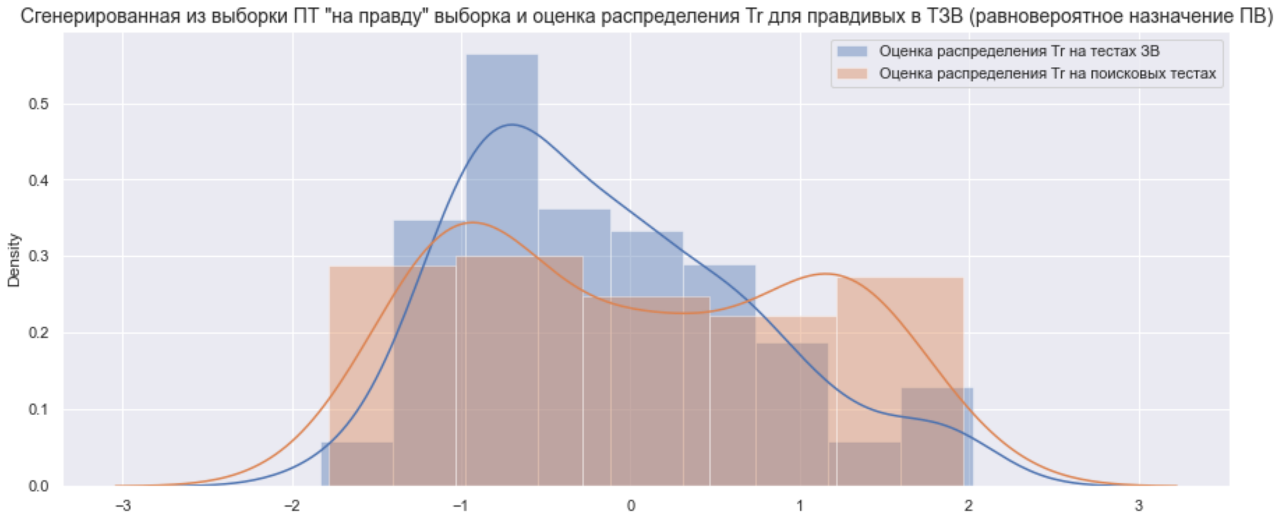

Рисунок 14. Пример сгенерированной выборки Trв ТЗВ при равновероятном назначении ПВ

Всего было сгенерировано 10 подобных распределений. При этом в пяти случаях p-значение, полученное в тесте Колмогорова-Смирнова, принимало значение меньшее чем 0,05 (в 5-ти большее 0,05). Простой расчёт показывает, что вероятность того, что возможно наблюдать подобную комбинацию p-значений, при условии, что обе выборки извлечены из одного распределения, равна 6,1e-05. Отсюда можно предположить, что СР на ПВ в тестах «на правду» не может с одинаковой вероятностью (0,2) получить ранг от 1-ого до 5-ти. И причина этого, почти наверняка, в том, что в реальных тестах полиграфолог размещает ПВ ближе к центру ряда и никогда первым или последним. На реальной выборке ТЗВ (99 повторов тестов с 5-ю стимулами) были подсчитаны частоты рангов, которые получал ПВ. Эти частоты размещены в таблице 16.

Таблица 16. Наблюдаемые частоты рангов ПВ

1	2	3	4	5
13	30	23	24	9

Исходя из значений наблюдаемых частот, было сделано предположение, что реальные частоты рангов ПВ имеют симметричное распределение, и на основании, этого, суммируя и усредняя, и с выравниванием с помощью аппроксимации параболой, которую можно увидеть на рисунке 15, были получены теоретические частоты ПВ –таблица 17

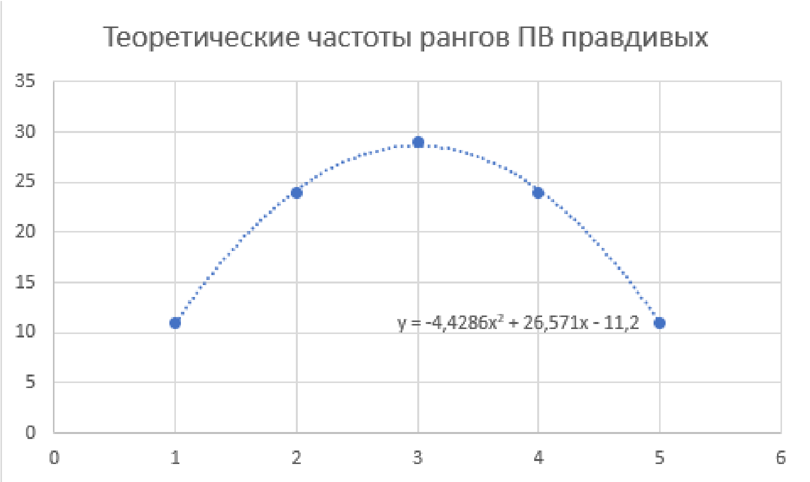

Рисунок 15. Теоретические частоты рангов ПВ правдивых

Таблица 17. Теоретические частоты рангов ПВ

1	2	3	4	5
11	24	29	24	11

Результат Хи2 теста: p = 0,483 – позволяет оставить это предположение в силе. Далее на основе полученных теоретических частот рангов ПВ, из выборки поисковых тесов «на правду» генерировались виртуальные распределения Tr для ТЗВ. На рисунке 16 можно видеть оценку типичного, полученного таким образом, распределения вероятностей. P-значение полученное в тесте Колмогорова-Смирнова на однородность с реальной выборкой, для распределения с рисунка 11 равняется 0,771. Всего, таким же образом было сгенерировано 10 виртуальных выборок, с каждой из которых проводился тест Колмогорова-Смирнова на однородность с реальной выборкой. Во всех случаях p-значение было больше 0,05. Отсюда вероятность наблюдать подобную комбинацию, при условии, что обе выборки извлечены из одного распределения равна 0,599. На основании критерия отношения правдоподобия можно сделать вывод, что гипотеза об определении параметров распределения Tr в ТЗВ является вполне правдоподобной в отличии от предположения равновероятного ранга ПВ в ТЗВ.

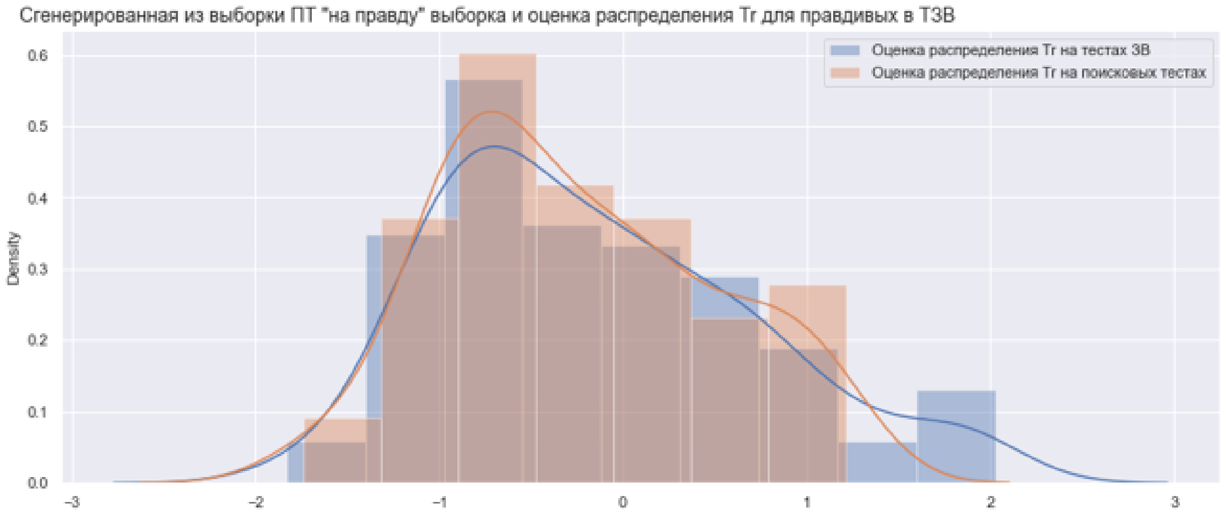

Рисунок 16. Пример сгенерированной из выборки ПТ «на правду» выборки Trв ТЗВ при назначении ПВ в соответствии с теоретическими частотами рангов ПВ

Для использования в дискриминирующем правиле для ТЗВ, из выборки ПТ была сгенерирована выборка Tr для ТЗВ (обучающая выборка) – рисунок 17.

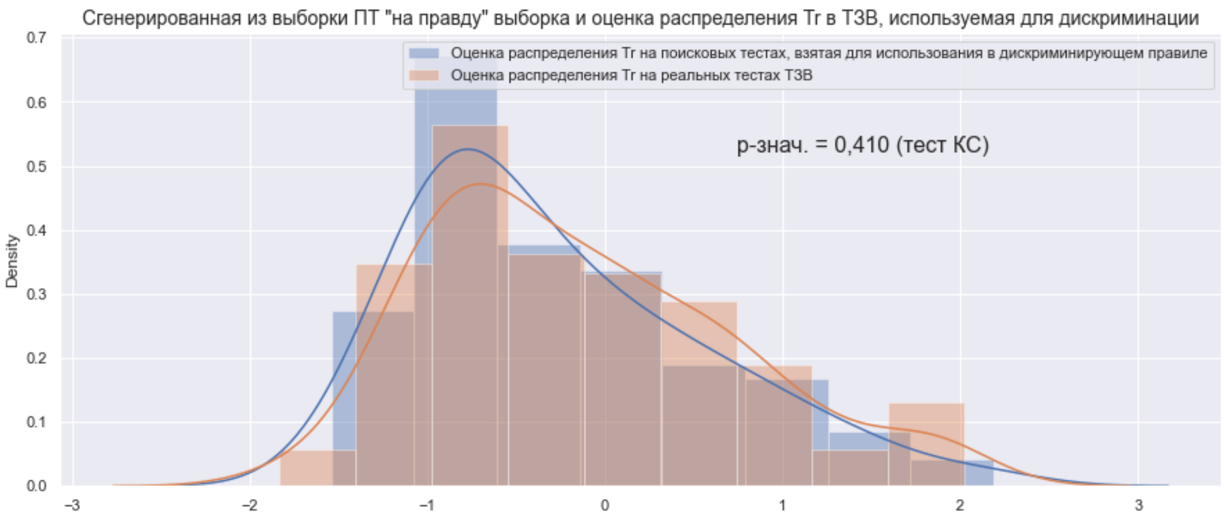

Рисунок 17. Виртуальная выборка Trв ТЗВ, используемая для дискриминации результатов ТЗВ.

Выводы:

1. Оценки распределения вероятностей, нормализованной СР «правдивых» в ТЗВ с пятью и шестью стимулами, могут быть получены на объединённой выборке. Соответствующие выборки в смысле однородности не имеют статистически значимого различия;
2. Закон распределения вероятностей нормализованной СР «правдивых» в ТЗВ определяется тем, что в реальных тестах полиграфолог размещает ПВ ближе к центру ряда и почти никогда первым или последним.
3. Выборки, нормализованной СР «правдивых» полученных в ТЗВ без ‘подавления’ и в ТЗВ с ‘подавлением’, не имеют статистически значимого различия в смысле однородности, и могут быть использованы для дискриминации реакций на ПВ в ТЗВ в равной мере.

Выборка тестов с ложными ответами

Для построения дискриминирующего(их) распределений Fl (нормализованной СР «лживых») для ТЗВ были использованы выборочные данные трёх видов тестов, два из них: стимулирующие тесты на «имя», тесты, проведённые по УД (уголовным делам); все тесты с известным решением – дела раскрыты, вынесено судебное решение. И третий вид тестов – тесты с известным решением – с карточками. Объём выборки тестов «на имя» 698 тестов, в каждом тесте пять стимулов и один «бросовый» – первый вопрос с именем, каждый тест в выборке повторялся только один раз. В выборке с тестами по УД 156 повторов тестов (так как будто каждый тест проводился отдельно и только один раз). В каждом тесте пять стимулов с одним «бросовым». На рисунке 18 изображены оценки соответствующих распределений. Тест Колмогорова-Смирнова на однородность (p = 0,180) подтверждает визуальную оценку о статистической незначимости различий между распределениями. Выпуклость в оценке плотности распределения вероятностей Fl «на имя», там, где у тестов по УД –вогнутость, почти наверняка, возникла из-за человеческого фактора – ошибок при записи в анализируемую таблицу рангов ПВ. В силу большого объёма выборки проверка этого обстоятельства не делалась.

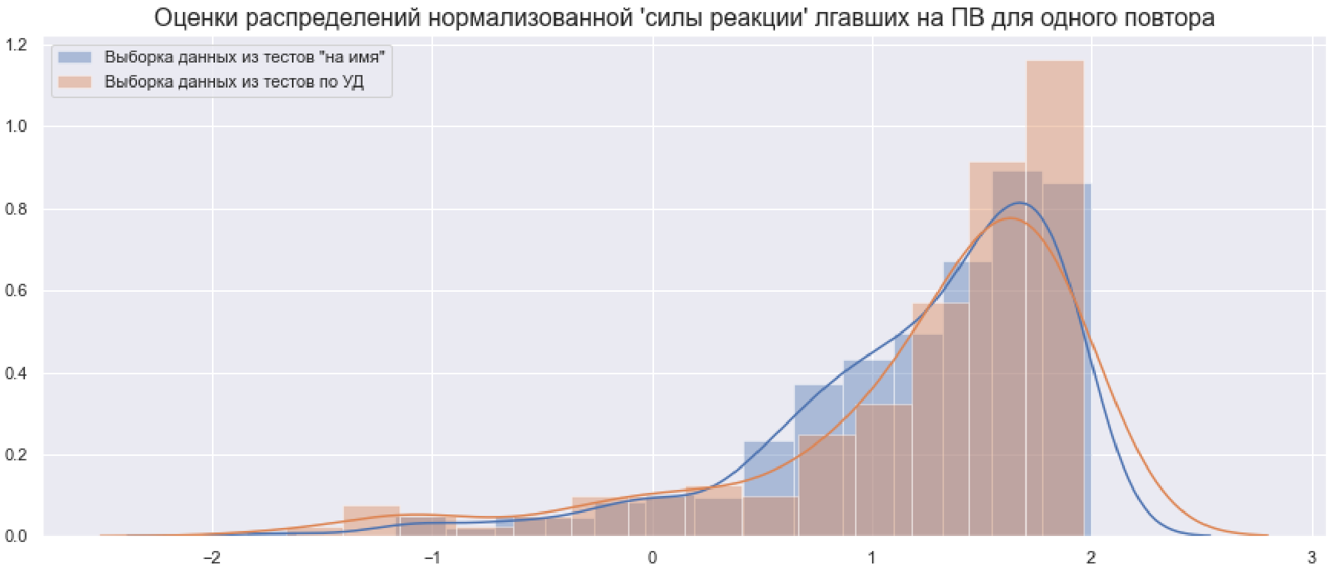

Рисунок 18. Оценки распределений Fl тестов «на имя» и по УД

Из выборки тестов «на имя» была сгенерирована случайная подвыборка объёмом 166 элементов, примерно равная объёму выборки тестов по УД. На рисунке 19 можно увидеть оценку распределения нормализованной СР на этой выборке.

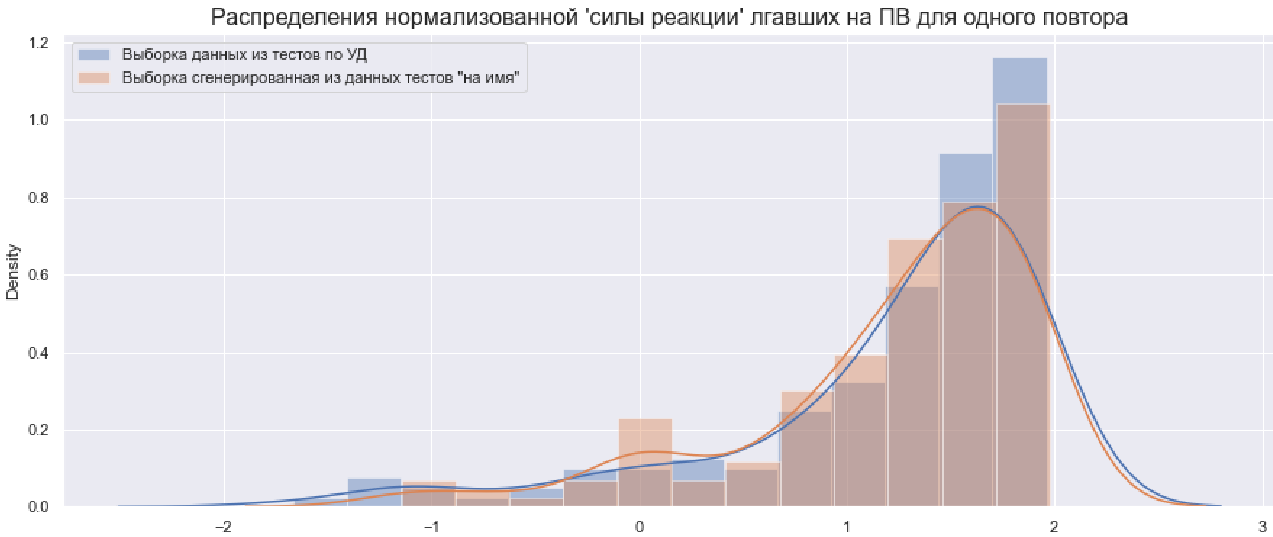

Рисунок 19. Оценки распределений Fl подвыборки тестов «на имя» и по УД

Р-значение, полученное в тесте Колмогорова-Смирнова для этих выборок равно 0,923. Подвыборка тестов «на имя» используется в алгоритме для дискриминации реакций на ПВ (обучающая выборка), а тесты по УД используются для испытания алгоритма (тестовая выборка). В третьей выборке тестов с карточками находилось 383 повторов тестов. В целях построения алгоритма эти тесты рассматривались как ТЗВ, так как после теста, тот стимул, на который лгал опрашиваемый, становился известен. На этой выборке было проведена оценка распределения Fl для тестов с шестью стимулами, первый из которых ‘подавлялся’, но брался в анализ. На рисунке 20 изображены оценки Fl для тестов с пятью стимулами и одним «бросовым» и для тестов с шестью стимулами, первый в повторе из которых «подавлялся».

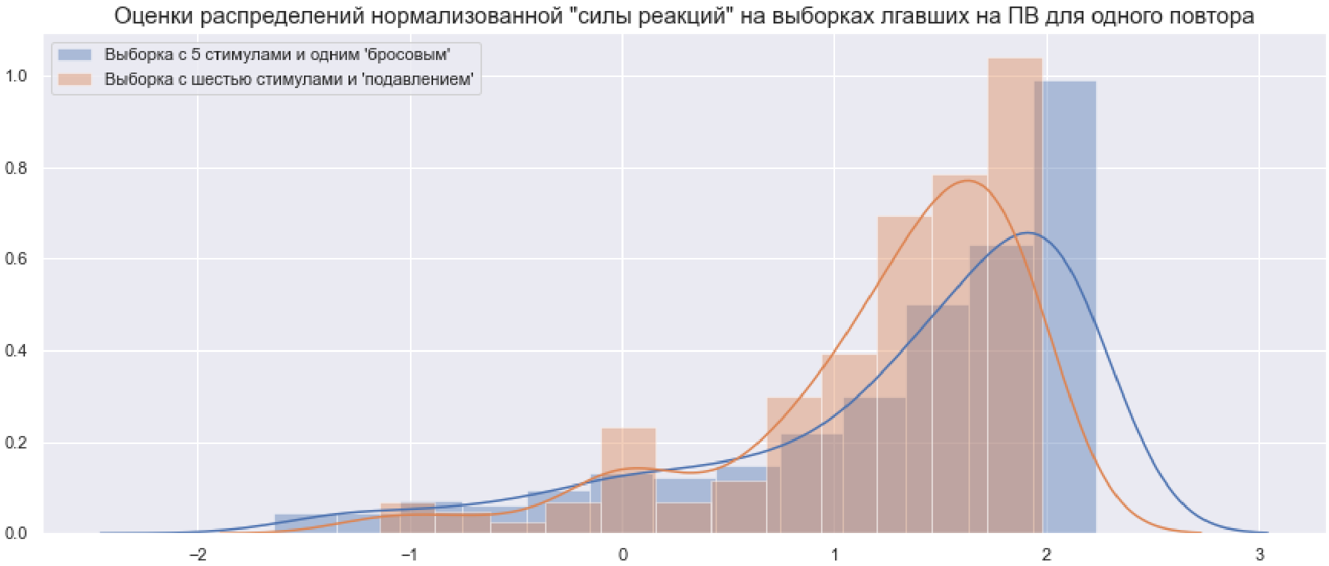

Рисунок 20. Оценки распределений Fl тестов «лживых»

Визуальный анализ предполагает неоднородность этих выборок, что находит подтверждение в тесте Колмогорова-Смирнова (p = 5,05e-08). Обе эти выборки используются для дискриминации реакций на ПВ в тестах ЗВ. Выборка тестов по раскрытым уголовным делам используется для испытания алгоритма (тестовая выборка).

Выводы:

Закон распределения Fl, в отличии от Tr, для ТЗВ с пятью стимулов и одним «бросовым» и ТЗВ с шестью стимулами и «подавлением» различаются.

Испытания алгоритма на ТЗВ

На рисунке 21 изображены оценки функций плотности распределения вероятностей, которые в алгоритме используются для классификации полиграмм поисковых тестов в алгоритме OL.

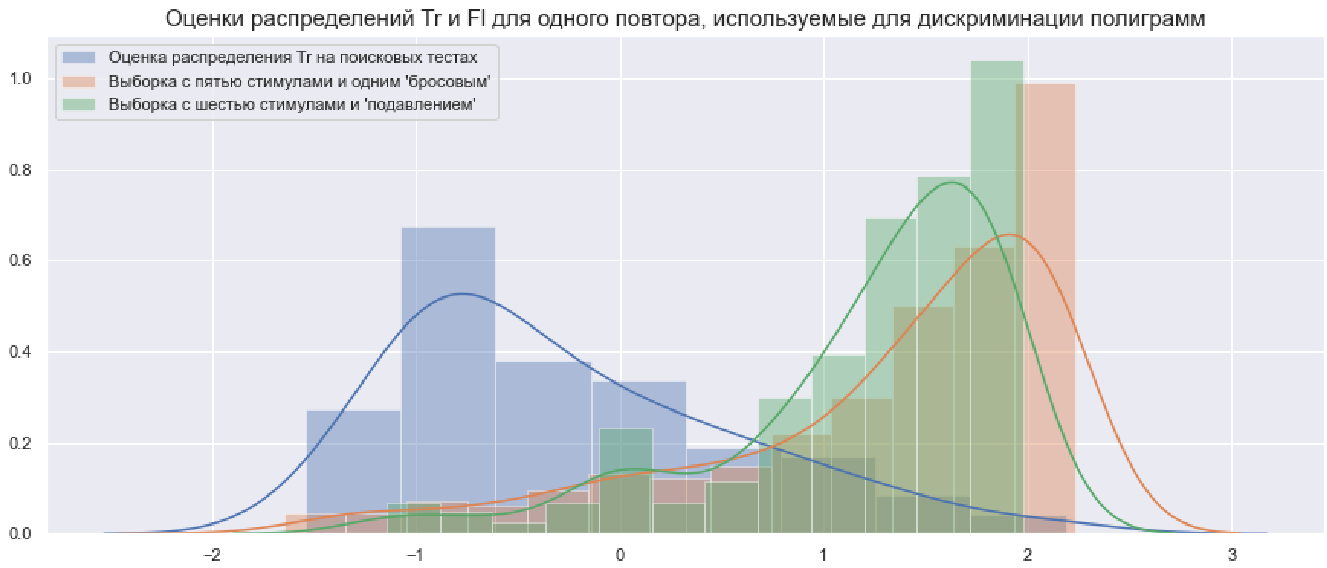

Рисунок 21. Оценки распределений Tr и Fl используемых для классификации полиграмм

Как можно видеть моды распределений находятся на существенно большем расстоянии друг от друга чем в случае поисковых тестов. Стоит ожидать, что в данном случае дискриминирующие свойства этих распределений будут существенно лучше, что будет приводить к меньшему проценту ошибочных и неопределённых результатов, при использовании значения априорной вероятности равной 0,5.

Испытание алгоритма OL на ТЗВ производилась на тестах, проведённых по реальным уголовным делам при значении априорной вероятности равной 0,5. В тестовую выборку входили 77 тестов, проведённых с лживыми лицами и 102 теста с правдивыми. Среди тестов с лживыми ответами присутствует 11 тестов с одним повтором, 55 с двумя, 10 с тремя и один с пятью повторами. Среди тестов с отсутствием лживых ответов имеется 56 тестов с одним повтором, 40 с двумя, 5 с тремя и один с четырьмя. Все полиграммы тестов были записаны в ПО полиграфа «Диана - 4» с использованием каналов: Дыхание, КГР, ФПГ. Результаты OL на этих выборках сравнивались с результатами алгоритма ChanceCalc (CC). Алгоритм OL вычисляет оценку вероятности принадлежности наблюдаемого значения, нормализованной CP к множеству значений Tr, что на предметном уровне можно трактовать как вероятность лживости (ВЛ) ответа на ПВ/значимости ПВ. Алгоритм CC вычисляет для каждого стимула в ТЗВ число от 0 до 1, которое имеет один из вариантов трактовки – «вероятность неслучайности реакции», далее - ВНР. Трактовать ВНР как ВЛ затруднительно. Так как заранее известно, что опрашиваемый либо отвечает лживо на ПВ, либо правдиво. А на остальные стимулы теста тот отвечает правдиво. CC может вычислить значение ВНР для ПВ равную, например, 0,9. И вычислит значения ВНР для других стимулов. В сумме ВНР для всех стимулов будет равна больше единицы. Для полученных значений ВЛ и ВНР, отдельно для каждой выборки вычислялись выборочные коэффициенты корреляции и строились диаграммы рассеяния. На рисунке 22 размещена диаграмм рассеяния для выборки с лживыми ответами.

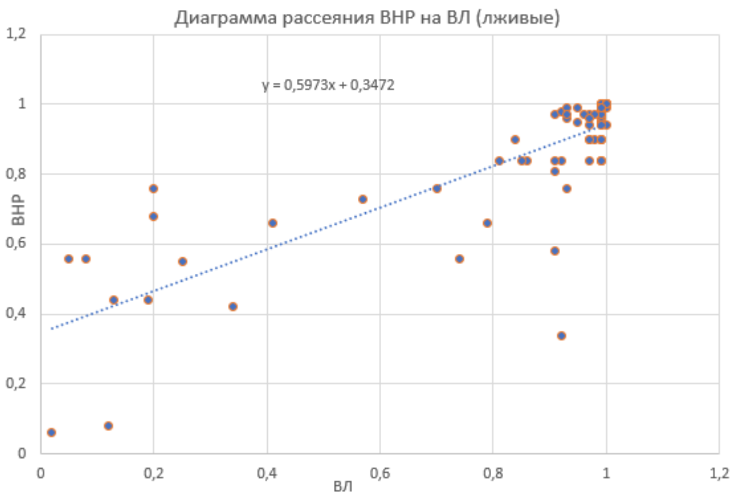

Рисунок 22. Диаграмма рассеяния ВНР на ВЛ (лживые)

Выборочный коэффициент корреляции для выборки с лживыми ответами равен 0,816. Такое его значение говорит о сильной положительной связи между ВНР и ВЛ. На рисунке 23 диаграмма рассеяния для выборки с отсутствием лживых ответов. Выборочный коэффициент корреляции для выборки с отсутствием лживых ответов равен 0,710. И в этом случае связь между ВНР и ВЛ по шкале Чеддока носить сильный положительный характер.

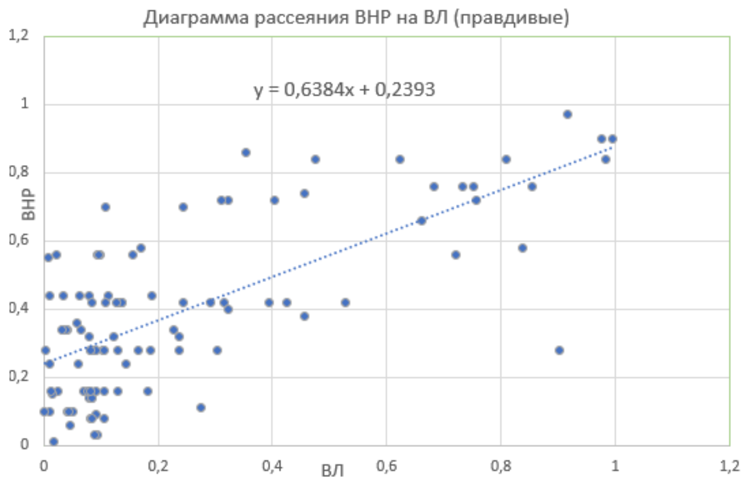

Рисунок 23. Диаграмма рассеяния ВНР на ВЛ (правдивые)

Так как в отношении СС отсутствует информация о правилах и порогах принятия решения, кроме рекомендации принимать решение о «неслучайности реакции» при превышении ВНР значения в 0,75, то правила принятия решения определялись одновременно для OL и СС. Если следовать, рекомендации создателей СС и принимать решение о значимости ПВ при значении ВЛ (ВНР) равном и более 0,75, то будут отсутствовать неопределённые решения, что увеличит количество ошибочных решений. В таблице 18 можно увидеть результаты алгоритмов на тестовых выборках полученные при указанном правиле.

Таблица 18. Результаты алгоритмов, правило ‘> 0,75’

Тесты и алгоритмы / Решение и метрики	Тесты с ложным ответом		Тесты без ложного ответа	
	OL	CC	OL	CC
Правильное решение	63	62	92	90
Ошибочное решение	14	15	10	12
Параметры	Чувствительность		Специфичность	
Значения	0,82	0,81	0,90	0,88
Общая точность	OL		CC	
	0,87		0,85	

Результаты алгоритмов OLи CC при таком правиле решения не имеют статистически значимого различия. Для выборки с ложными ответами тест Xi2 на однородность даёт значение p-value = 1, а для выборки без ложных ответов p-value = 0,821. Данное правило даёт большой процент ошибочных решения 18% для OL и 19% CC для «лживых», соответственно 10% и 12% для правдивых. В силу статистической незначимости различия результатов алгоритмов процент ошибки можно усреднить – 18,5% и 11%. Результаты правила ‘<0,5 и >0,9’ помещены в таблицу 19. Согласно этому правилу решение о значимости ПВ принимается при ВЛ (ВНР) равно и больше 0,9. Решение о незначимости при равно и меньше 0,5.

Таблица 19. Результаты алгоритмов, правило ‘<0,5 и >0,9’

Тесты и алгоритмы / Решение	Тесты с ложным ответом		Тесты без ложного ответа	
	OL	CC	OL	CC
Правильное решение	58	49	86	73
Ошибочное решение	11	6	5	3
Неопределённый результат	8	22	11	26
Параметры	Чувствительность		Специфичность	
Значения	0,84	0,89	0,95	0,96
Процент ошибок	14%	8%	6%	4%
Процент не опред. решений	10,4%	28,6%	10,8%	25,5%
Общая точность	OL		CC	
	0,90		0,93	
Общий процент не определённых результатов	10,6%		26,8%	

При введении правила ‘<0,5 и >0,9’ результаты классификации полиграмм ТЗВ алгоритмами OLи CC имеют статистически значимое различие. В тесте Xi2 на однородность для «лживых» p = 0,013, для «правдивых» – 0,022. Но это не значит, что чувствительность, специфичность и общая точность при использовании алгоритма CC и этого правила будут обязательно превышать эти же показатели при использовании алгоритма OL, так как статистически значимая разница между алгоритмами возникает за счёт существенного различия в процентах не определённых исходов, которое имеют различие почти в 2,5 раза. При исключении неопределённых результатов статистически значимой разницы между результатами алгоритмов не обнаруживается. Тест Xi2 даёт достигаемый уровень значимости р = 0,585, точный тест Фишера – р = 0,729. Для OL правило ‘ <0,5 и >0,9’ по сравнению с правилом ‘> 0,75’ несколько уменьшает процент ложноотрицательных ошибок (в 1,3 раза) и более существенно уменьшает процент ложноположительных ошибок (в 1,8 раза). При этом второе правило даёт приемлемый процент неопределённых результатов. Учитывая, что общее решение по результатам опроса, как правило, не принимается по одному ТЗВ, можно отдать предпочтение правилу ‘<0,5 и >0,9’. Так как при исключении неопределённых результатов между результатами OL и CC отсутствует статистически значимая разница, то можно получить оценку параметров алгоритмов путём усреднения их результатов. Окончательно для OL будем иметь:

1. Чувствительность – 0,86;
2. Специфичность – 0,95;
3. Общая точность – 0,91;
4. Не определённые решения – 0,1.

Не без интересно отметить, что этот результат совпадает с показателями чувствительности и специфичности ТЗВ, которые были получены в исследованиях японских специалистов Огава, Мицуда и Цунеока (2013) [4,14]. Отличие имеется только в количестве неопределённых результатов. В японских исследованиях их 20%. Японские исследования проведены в лабораторных условиях, а не на полевых данных, и при единообразном количестве повторов теста, равном количеству стимулов в ряду.

Тест вопросов сравнения

Тест вопросов сравнения (ТВС) отличается от ТЗВ. В данном типе теста вводятся так называемые вопросы сравнения (ВС), которые предназначены для того, чтобы сравнивать с ними реакции на проверочные вопросы (ПВ). Если реакция на ВС сильнее чем на ПВ, то принимается решение о правдивости проверяемого, если реакции на ВС слабее, чем на ПВ, то о лживости. Правила составления и ввода ВС можно посмотреть в специализированной литературе [5, 12].

Однотемный ТВС в алгоритме Out_LIEr можно рассматривать как совокупность тестов с одним проверочным вопросом (ПВ). Количество таких тестов равно количеству включённых ПВ. Каждый такой тест отличается от другого показателями признаков реакций на ПВ. Вопросы сравнения в таких тестах общие. Входным параметром алгоритма для расчёта по условно первому тесту является априорная вероятность. На его выходе апостериорная вероятность, которая подаётся на вход следующей итерации алгоритма. Для классификации полиграмм таким образом необходимо построить распределения вероятностей СР в каждом из классов для одного проверочного вопроса. Для этого использовалась российская выборка полиграмм в которую входит 200 тестов с известным решением. Так как в эту выборку входят тесты проведённые с 140 лицами (в среднем более одного теста на опрашиваемого) из неё случайным образом отбиралось по одному тесту на опрашиваемого. Таким образом отбиралось 140 полиграмм тестов из них 81 тест с правдивыми ответами на ПВ и 59 со лживыми. В случае выборки с правдивыми ответами отбиралось 20 подвыборок. На рисунке 24 изображена типичная оценка плотности вероятностей распределения ‘силы реакции’ правдивых (Tr) на подвыборке и оценка плотности вероятностей на выборке (100000 элементов) взятой из нормального распределения с параметрами оценёнными на подвыборках.

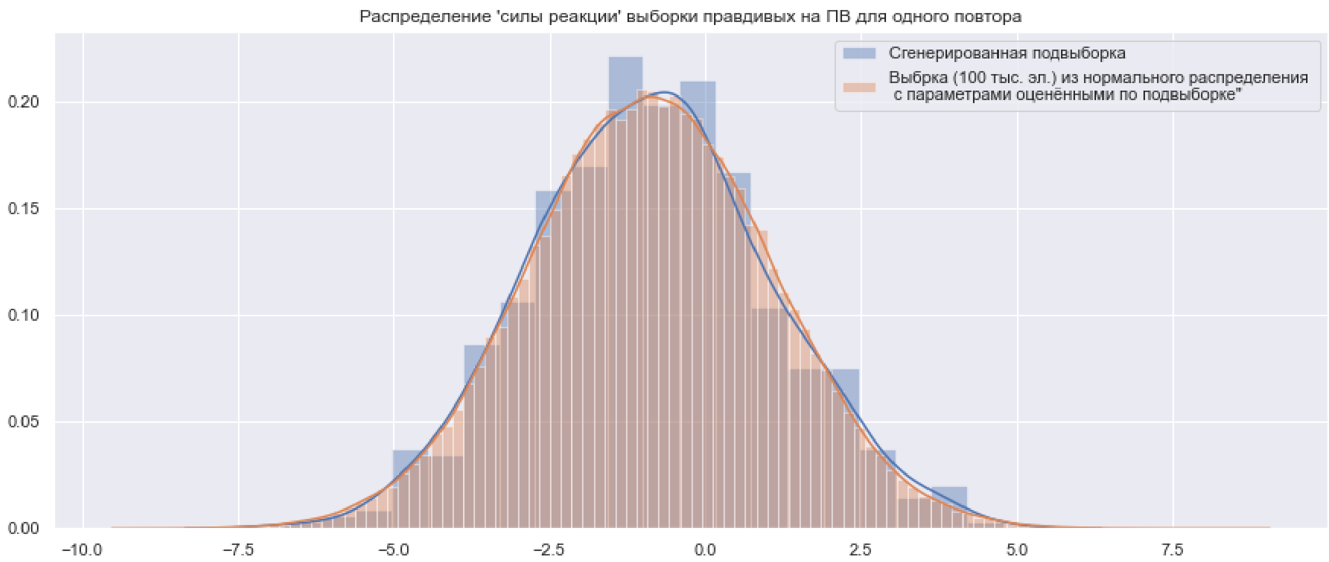

Рисунок 24. Оценка плотности вероятностей Tr

Результат теста Шапиро-Уилка на нормальность – p = 0,894. Результат теста Колмогорова-Смирнова на однородность свыборкой (80 элементов) из нормального распределения, с параметрами оценёнными, путём усреднения, на 20 подвыборках – p = 0,996. На рисунке 25 график квантиль-квантиль для нормального распределения и выборки из Tr.

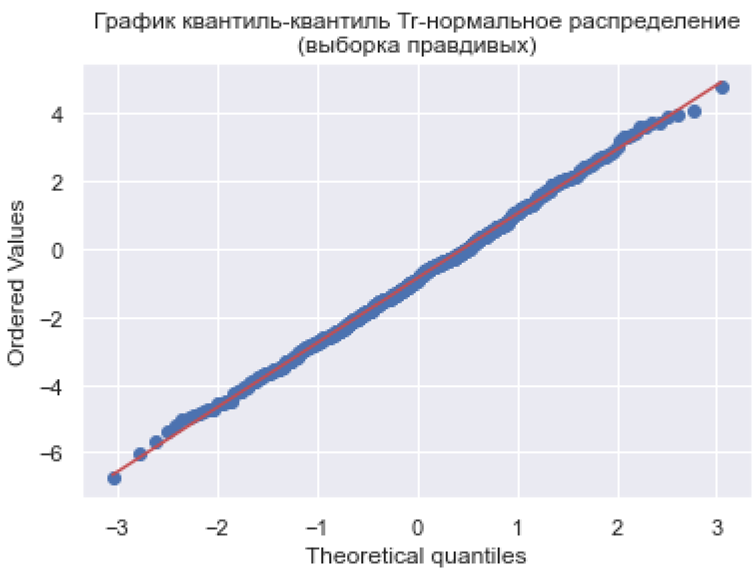

Рисунок 25. График квантиль-квантиль нормальное распределение –Tr

Во всех 20 случаях статистические тесты уверено показывали, что подвыборки взяты из нормального распределения. Таким образом получена плотность вероятностей распределения Tr, которая используется в алгоритме для классификации полиграмм.

В случае тестов с ложными ответами так же отбиралось 20 подвыборок. Но в этом случае в тесте Шапиро-Уилка достигаемый уровень значимости в 8-ми случаях был меньше 0,05. В 12 случаях больше. Если подвыборки извлекались из нормального распределения, то вероятность такого события равна 2,7e-06, что заставляет отказаться от предположения о нормальности ‘силы реакции’ (Fl) в тестах со лживыми ответами. Вместе с тем тест Колмогорова-Смирнова на однородность с выборкой объёмом в 59 элементов, извлечённой из нормального распределения, во всех случаях давал p-значение большее 0,05. Вероятность такого исхода в случае однородности выборок равна 0,36 – является высокой. Из результатов статистических тестов можно сделать вывод, что Fl не является нормально распределённой случайной величиной, но её закон распределения отличается от нормального достаточно мало, так, что тест Колмогорова-Смирнова этого не обнаруживает. Но в таком случаем отклонением от нормальности в задаче классификации можно пренебречь. На рисунке 26 изображена оценка плотности вероятностей распределения ‘силы реакции’ (Fl) и выборка (100000 элементов) взятая из нормального распределения с параметрами оценёнными на подвыборках

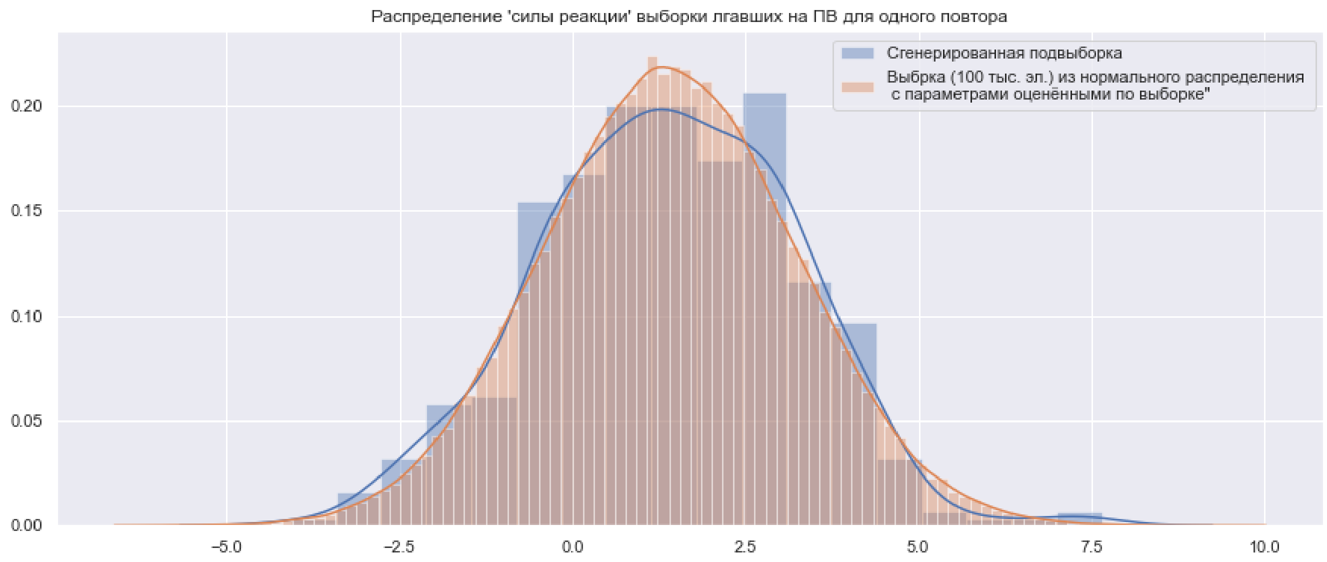

Рисунок 26. Оценка плотности вероятностей Fl

На рисунке 27 график квантиль-квантиль для Fl.

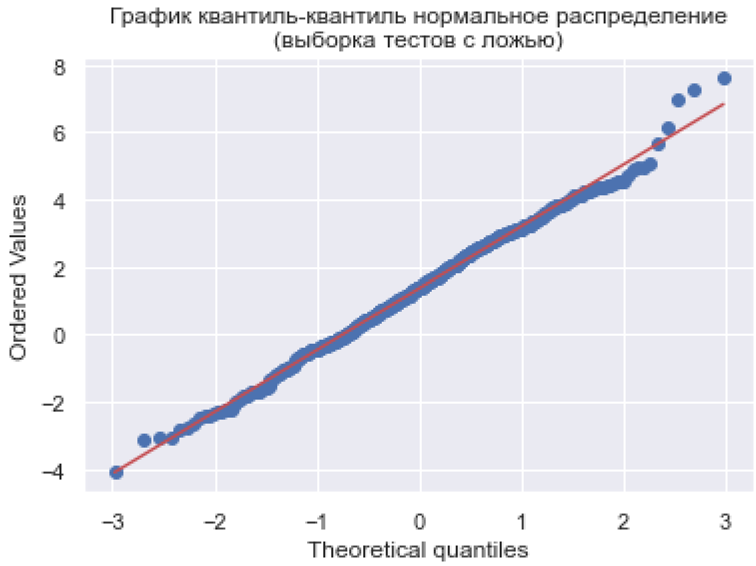

Рисунок 27. График квантиль-квантиль нормальное распределение –Fl

Р-значение в тесте Шапиро-Уилка для выборки с рисунка 20 равно 0,218. А р-значение в тесте Колмогорова-Смирнова – 0,442. В алгоритме для классификации полиграмм для Fl используется нормальное распределение с параметрами, которые получены путём усреднения на сгенерированных подвыборках.

Испытание алгоритма на ТВС

На рисунке 28 помещены плотности распределений, которые используются в алгоритме для классификации полиграмм. Плотности значительно перекрывают друг друга, поэтому для успешной классификации полиграмм требуется не однократное повторение ПВ в ходе тестов. И те полиграммы не будут приводить к неопределённому результату на которых реагирование будет стабильнее и с большей СР.

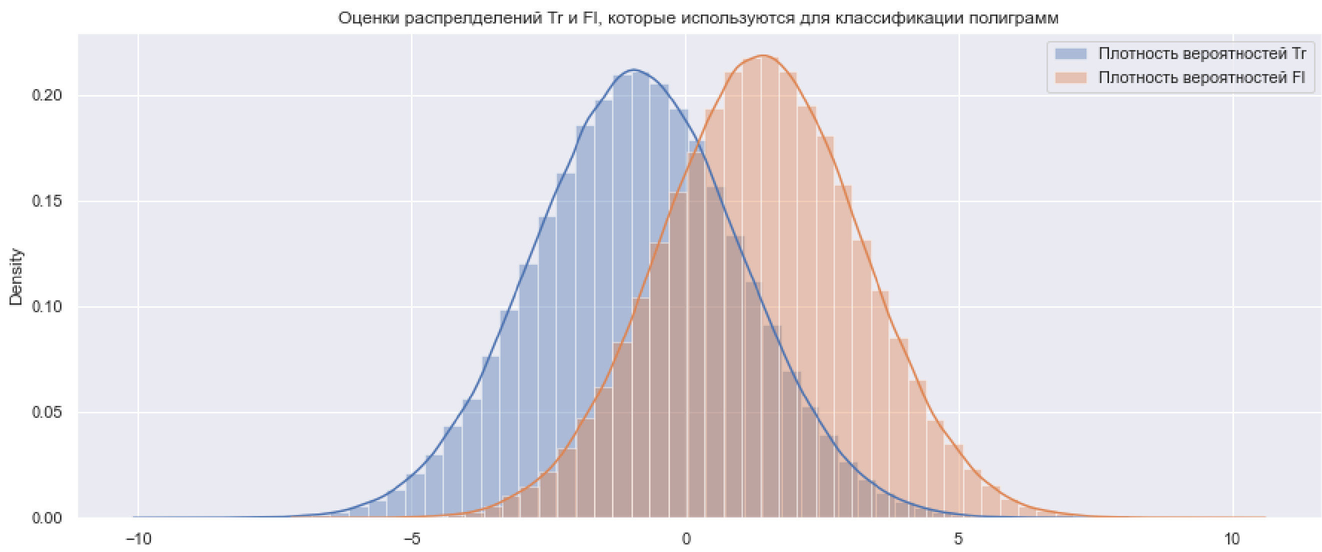

Рисунок 28. Оценки распределений Tr и Fl

Алгоритм испытывался на двух выборках полиграмм. Первая – рандомизированная выборка из архива подтверждённых дел Национального центра оценки достоверности информации (США). Объём выборки – сто подтверждённых случаев, из них пятьдесят тестов с правдивыми опрашиваемыми, пятьдесят со лживыми. Выборка, описанная в исследовании Krapohl, Cushman [7]. В настоящее время изображения этих полиграмм находятся в сетевом сервисе «Яндекс-диск» [6]. Вторая выборка российского происхождения объёмом 200-ти случаев с известным решением. В неё включены 100 случаев с правдивыми ответами и 100 случаев с лживыми ответами. Полиграммы были записаны на полиграфах «Диана», «Триумф», «Эпос», «Лафайет». Испытание алгоритма «Out_LIEr» производилось в сравнении с результатами алгоритма «Сокол» [5]. В таблице 20 показаны результаты алгоритмов OL и «Сокол» на американской выборке, так же на рисунке имеются результаты американских алгоритмов и результаты обсчёта полиграмм экспертами из США и России. В обсчёте полиграмм участвовало 5 российских полиграфологов и 10 американских. В таблице приведены усреднённые результаты экспертов, округлённые до целых чисел. В таблицу 21 помещены параметры алгоритмов, оценённые на американской выборке.

Таблица 20. Результаты испытания алгоритмов на американской выборке в сравнении с другими алгоритмами

Алгоритм	**TN**	**TP**	**Corr.**	**FN**	**FP**	**Err.**	**IN**	**IP**	**INC**
ESS-M	46	40	86	2	4	6	2	6	8
PA	39	38	77	4	4	8	7	8	15
OSS2	38	36	74	2	4	6	10	10	20
OSS3	43	36	79	2	5	7	5	9	14
HSS	41	42	83	9	4	13	4	0	4
Сокол	47	39	86	1	2	3	1	10	11
OL	48	40	88	3	1	4	1	7	8
Эксп. (RU)	35	41	76	4	3	7	11	6	17
Эксп. (USA)	41	40	81	6	6	12	4	3	7

Таблица 21. Параметры алгоритмов оценённые на американской выборке

Алгоритм	Sen	Spe	Acc	Inc
ESS-M	0,95	0,92	0,93	0,08
PA	0,90	0,91	0,91	0,15
OSS2	0,90	0,95	0,93	0,20
OSS3	0,88	0,96	0,92	0,14
HSS	0,91	0,82	0,86	0,04
Сокол	0,98	0,96	0,97	0,11
OL	0,93	0,98	0,96	0,08
Эксп. (RU)	0,91	0,92	0,92	0,17
Эксп. (USA)	0,87	0,87	0,87	0,07

Результаты алгоритмов OL и «Сокол» на американской выборке близки между собой и превосходят результаты американских алгоритмов и экспертов обеих стран. Работа алгоритма проверялась на российской выборке полиграмм с известным решением из 200 полиграмм. В таблице 22 находятся результаты проверки алгоритмов. Полиграммы получены при тестировании опрашиваемых, когда на одного опрашиваемого приходится по несколько тестов, что искажает в худшую сторону параметры алгоритмов. Происходит это по причине того, что плохая физиология опрашиваемого присутствует во всех тестах, которые с ним проводились, а при плохой физиологии полиграфологи увеличивали количество тестов, проведённых с ним.

Таблица 22. Результаты алгоритмов на 200 российских полиграммах (с повторами)

Алгоритм	**TN**	**TP**	**Corr.**	**FN**	**FP**	**Err.**	**IN**	**IP**	**INC**
Сокол	86	78	164	7	7	14	7	15	22
OL	80	72	152	5	7	12	15	21	36

В таблице 23 находятся оценки параметров алгоритмов на 200-х российских полиграммах.

Таблица 23. Параметры алгоритмов на 200 российских полиграммах (с повторами)

Алгоритм	Sen	Spe	Acc	Inc
Сокол	0,92	0,92	0,92	0,11
OL	0,91	0,94	0,93	0,18

С целью получения корректных оценок из выборки российских тестов генерировалась 1000 подвыборок объёмом 100 полиграмм так что бы в неё входило 50 полиграмм с правдивыми ответами и 50 с лживыми. При этом в подвыборки входило только по 1 тесту от опрашиваемого. Результаты вычисления параметров алгоритмов были усреднены по всем подвыборкам и округлялись до целых чисел. Результаты оценки параметров алгоритмов представлены в таблице 24.

Таблица 24. Параметры алгоритмов на сгенерированных подвыборках (без повторов)

Алгоритм	Sen	Spe	Acc	Inc
Сокол	0,95	0,94	0,95	0,09
OL	0,93	0,95	0,94	0,14

Показатели алгоритмов на российской выборке ниже, чем на американской, но всё же выше, чем показатели американских алгоритмов на американской выборке. В целом можно сказать, что OL несколько уступает по своим параметрам «Соколу».

На объединённой выборке из 300 полиграмм (американских и российских) изучалась направление и теснота связи между результатами алгоритмов на полиграммах. На рисунке 29 изображена карта сопряжения результатов алгоритмов на объединённой выборке. Обозначение (Л Л) – первая буква означает истинное положение (в данном случае лживые ответы), вторая буква результат алгоритма. INC – обозначение ситуации когда решение не принято из-за не достаточно убедительного результата алгоритма на полиграмме.

	Сокол					INC	INC	INC
OL		Л Л	Л П	П П	П Л	Л Л	Л П	П Л
	Л Л	105	0			7	0	
	Л П	1	6			3	0	
	П П			125	1			1
	П Л			0	6			0
INC	Л Л	10	0			6	0	
INC	Л П	1	2			8	1	
INC	П П			5	0			1
INC	П Л			3	2			5

Рисунок 29. Карта сопряжения результатов алгоритмов на объединённой выборке

Не учитывая наличие не определённых исходов, совпадений правильных решений алгоритмов – 259, совпадений решений включая ошибочные – 281, противоположных решений – 19. При введении не определённых исходов количество противоположных решений резко сокращается до 2 случаев. Очевидно, что при этом исключаются полиграммы с плохой физиологией и не стабильным реагированием. Но также уменьшается и совпадение правильных решений до 230, и просто совпадений до 242 случаев. В таблице 25 размещены эти же числа отдельно для случаев правдивых и лживых.

Таблица 25. Совпадения решений алгоритмов

	Правдивые	Лживые
Без неопределённого результата		
Совпадение правильных решений	131	128
Совпадение решений	144	137
Противоположные решения	6	13
С неопределённым результатом		
Совпадение правильных решений	125	105
Совпадение решений	131	111
Противоположные решения	1	1

В каждой категории из таблицы 25 уменьшение числа совпадений и противоположных решений произошло существенно за счёт перевода полиграмм лживых в статус неопределённых решений. На рисунке 30 изображена таблица сопряжённости результатов алгоритмов.

OL \ Сокол	Л	П	INC
Л	111	0	7
П	2	131	5
INC	13	10	21

Рисунок 30. Таблица сопряжённости результатов алгоритмов

Результат точного теста Фишера для этой таблицы – достигаемый уровень значимости равен 2,2е-16, что однозначно свидетельствует, что о высокой взаимосвязи результатов алгоритмов и о близости их параметров. Это означает, что на одной полиграмме алгоритмы с большой вероятностью дадут одинаковый результат. На рисунке 31 таблица сопряжённости 2 на 2 – результатов алгоритмов без учёта не определённых результатов (удалены из анализа).

	Сокол		
OL		Л	П
	Л	111	0
	П	2	131

Рисунок 31. Таблица сопряжённости 2 на 2 результатов алгоритмов

Точный тест Фишера для этой таблицы даёт значение достигаемого уровня значимости меньшее 2,2e-16. Коэффициент контингенции для принятых решений алгоритма равен 0,98. Такое его значение говорит о высокой положительной связи результатов алгоритмов.

Кроме указанных выборок с известным решением сравнительная проверка алгоритмов проводилась на независимой выборке из двухсот полиграмм ТВС с неизвестным решением, которые были записаны на полиграфе «Диана». На рисунке 32 размещена таблица сопряжения 2 на 2 принятых решений алгоритмов.

OL / Скл	правда	ложь
правда	102	0
ложь	1	35

Рисунок 32. Таблица сопряжённости 2 на 2 результатов алгоритмов на выборке из 200 полиграмм

Результат точного критерия Фишера для этой таблицы – p <2,2e-16. И коэффициент контингенции такой же, как и на выборке из 300 полиграмм – 0,98.

Все тесты из вышеуказанных выборок с вопросами вероятной лжи (ВВЛ). Сравнительное испытание алгоритмов производилось и на выборке тестов (99 случаев) с вопросами управляемой лжи (ВУЛ) – на выборке тестов DLST. Дыхание было исключено из анализа. На рисунке 33 таблица с сопряжёнными результатами алгоритмов на этой выборке. З – решение алгоритма «стимул значим», НЗ – решение алгоритма «стимул не значим», НВ –нет вывода.

(Без дыхания)

С \ OL	З	НЗ	НВ	Сумма	
З	10	1	6	17	
НЗ	1	63	0	64	
НВ	1	6	11	18	84
Сумма	12	70	17	99	2

Рисунок 33. Таблица с сопряжёнными результатами алгоритмов на тестах однотемного скрининга.

.

Результат точного теста Фишера –p <2,2e-16. Коэффициент контингенции равен 0,89. Такое значение по шкале Чеддока означает высокую положительную связь между результатами алгоритмов.

Вывод:

Используя для классификации полиграмм оптимальное байесовское решающее правило, алгоритм Out_LIEr обладает хорошей точностью как в случае МВСИ, так и в случае ТВС. Общая точность алгоритма, если и не превосходит, то не уступает американским алгоритмам и системам экспертного обсчёта полиграмм. OL CQT незначительно уступает в характеристиках алгоритму «Сокол», при высокой степени согласованности их результатов. Вероятность того, что алгоритмы на одной и той же полиграмме (на одних и тех же метрических данных) примут одинаковое решение оценивается в 0,99. Если исключить неопределённые результаты, то результаты OL CIT не имеют статистически значимого отличия от результатов алгоритма ChanceCalc. А при использовании рекомендованного правила принятия решения ‘<0,5 и >0,9’ OL CIT имеет существенно меньшее число неопределённых результатов чем ChanceCalc. Значения чувствительности и специфичности ТЗВ полученные с помощью OL на полевых данных близки (совпадают, если не говорить о неопределённых результатах) к соответствующим значениям, которые были получены в ходе исследований японских специалистов.

Список литературы

1. Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д.; под ред. Айвазяна С. А. Прикладная статистика: Классификации и снижение размерности: Справ. изд. / – М.: Финансы и статистика, 1989, – 607 с.
2. Ллойд Э., Ледерман У. (ред.). Справочник по прикладной статистике. Том 2. – М.: Финансы и статистика, 1990. - 526 с.
3. Молчанов А.Ю., Оглоблин С.И. Инструментальная детекция лжи: академический курс. – Ярославль: Нюанс, 2004. – 335 с.
4. Осуги Акеми. Проверка на полиграфе как метод обнаружения скрываемой информации: методика и практика использования полиграфа в Японии. //Детекция лжи. 2020. С.98-115
5. Поповичев С.В. Ложь как психофизиологический феномен. //Детекция лжи. 2020. С.116-147
6. Поповичев С.В. Тестирование с применением полиграфа: экспериментальное исследование или диагностическое измерение. //Детекция лжи. 2020. С.129-155
7. Поповичев С.В. Легко солгать тяжело: Инструментальная детекция лжи: от идеологии к технологии. – М.: ЗАО «Группа ЭПОС», 2011, –400 с.
8. Летков Ю. В., Калафати А. Ю. 2020. Алгоритм классификации полиграмм «Сокол». PREPRINTS.RU. https://doi.org/10.24108/preprints-3112144.
9. Размеченные полиграммы: [Электронный каталог] /2020. – URL: https://yadi.sk/d/6gxZj3keSPD_5g?w=1. (Дата обращения 01.10.2021).
10. Krapohl D.J. and Cushman, B., 2006. Comparison of evidentiary and investigative. decision rules: A replication. Polygraph, 35(1), 55-63.
11. Kircher J.C. and Raskin D.C., 1988. Human Versus Computerized Evaluations of Polygraph Data in a Laboratory Setting. Journal of applied psychology. 73(2) 291-302.

12. Kircher J.C., Kristjansson S.D., Gardner M.K., and Webb A. 2005 Human and Computer Decision-Making in the Psychophysiological Detection of Deception Polygraph, 2012 (reprint), 41(2) 77-126
13. American Polygraph Association, 2011. Meta-analytic survey of criterion accuracy of validated polygraph techniques. Polygraph 40 (4), 196–305. http://www.polygraph.org/section/research-standards-apa-publications.
14. Accuracy of Concealed Information Test as a Memory Detection Technique: A Laboratory Study. January 2013. Japanese Journal of Forensic Science and Technology 18(1):35-44

Приложение А

Для проверки наличия эффекта «лесенки», который возникает из-за размещения проверочного вопроса ближе к центру ряда, случайным образом было отобрано 50 первых повторов ТЗВ с 5-ю стимулами в тесте. В каждом таком тесте первый «бросовый» стимул, который опрашиваемый должен воспринимать как равноценный с остальными стимулами ряда, но на самом деле оценивается полиграфологом как не релевантный, обозначался как в ПО полиграфа Диана буквой «О» и не включался в анализ. Остальные стимулы задавались в тесте в порядке соответствующему числовому обозначению стимула. В полученных данных производился подсчёт случаев, когда каждому из стимулов теста присваивался наивысший ранг (1 ранг–наибольшая СР). В таблице А1 результаты подсчёта.

Таблица А1. Частоты получения 1-го ранга стимулами ТЗВ

Стимул	Нв1	Нв2	Нв3	Нв4	Пв1	Всего
Частота	23	9	7	7	4	50

На рисунке А1 диаграмма с частотами 1-го ранга.

Рисунок А1. Частоты получения 1-го ранга стимулами ТЗВ

Результаты Xi2 (р <0,001) отвергают предположение о равномерном распределении 1-го ранга среди стимулов первого повтора ТЗВ.