

Методы вычислительной диагностики

Задачи вычислительной диагностики

1. Постановка объективного диагноза на базе обобщения накопленного медицинского опыта
2. Выяснение достаточности медицинской информации для постановки диагноза
3. Выработка рекомендаций по наиболее целесообразному маршруту обследования
4. Оценка эффективности лечения и прогнозирования результатов

Методы вычислительной диагностики

Терминология

Признак – любой заранее сформулированный вопрос, касающийся:

- 1) определенных характеристик, полученных клиническими, лабораторными и инструментальными методами исследования, применяемыми при конкретной болезни;
- 2) типологических, психоэмоциональных и конституциональных особенностей обследованного больного;
- 3) воздействия факторов внешней среды, которые могут быть значимыми при рассматриваемой болезни.

Симптом – заранее сформулированный ответ на любой из вопросов (признаков).

Например, признаком является состояние тяжести больного. К нему относятся симптомы: легкое состояние, состояние средней тяжести, тяжелое состояние и т. п.

Симптокомплекс – любое сочетание симптомов, относящихся к разным признакам.

Устойчиво наблюдаемая совокупность симптомов называется синдромом.

Методы вычислительной диагностики

Терминология

Частота симптома – количество случаев (наблюдений), в которых встречается рассматриваемый симптом при данной болезни в данной группе наблюдения.

Вероятность симптома – отношение частоты симптома при данной болезни в данной группе наблюдений к объему этой группы.

Алгоритм – определенная последовательность элементарных предписаний или правил, выполнение которых приводит к решению поставленной задачи.

Методы вычислительной диагностики

- 1) Логические методы
- 2) Вероятностные методы
 - а) Балльные – диагноз ставится на основе сложения баллов. Балл (вес) симптома для соответствующего диагноза определяется заранее и, как правило, является мнением врача о важности того или иного симптома для данного диагноза.
 - б) Явные – диагностика проводится на базе вычислений апостериорных вероятностей болезней из группы рассматриваемых. В основе явных вероятностных методов лежит использование формулы Байеса.
 - в) Неявные – наиболее вероятный диагноз определяется без вычисления апостериорных вероятностей в явном виде

Явные вероятностные методы

Условные вероятности. Формула Байеса

Условной вероятностью некоторого события относительно другого события называют отношение вероятности их совместного появления к вероятности второго события. Таким образом, условная вероятность появления симптома S при данном заболевании D равна:

$$P(S/D) = \frac{P(S \cap D)}{P(D)}, \quad (1)$$

где $P(D) \neq 0$.

Символ S обозначает любые сведения о пациенте в терминах признаков, симптомов, симптомокомплексов, диагностических тестов, оценок параметров физиологических сигналов и т. п.

Символ D обозначает заболевание, комплекс из нескольких заболеваний или отсутствие проявления заболеваний.

Явные вероятностные методы

Условные вероятности. Формула Байеса

$$P(S/D) = \frac{P(S \cap D)}{P(D)}, \quad P(S \cap D) = P(S/D)P(D),$$

$$P(D/S) = \frac{P(D \cap S)}{P(S)},$$

$$P(D/S) = \frac{P(S/D)P(D)}{P(s)}, (2)$$

Формула Байеса

$$\frac{P(D/S)}{P(D)} = \frac{P(S/D)}{P(s)}, \quad (3)$$

Явные вероятностные методы

Условные вероятности. Формула Байеса

[Ссылка на методическое пособие по вычислительной диагностике](#)

Явные вероятностные методы

Правдоподобие. Отношение правдоподобия

Правдоподобие некоторой гипотезы (диагноза) представляет собой вероятность действительного наблюдения результатов эксперимента (симптомов) при данной гипотезе (данном диагнозе).

Величина $P(S/D)$ и есть правдоподобие.

Таким образом, теорема Байеса говорит, что конечная вероятность гипотезы $P(D/S)$ пропорциональна ее начальной вероятности $P(D)$, умноженной на ее правдоподобие $P(S/D)$.

В зависимости от того, что является переменной в вероятностной функции, используется термин «условная вероятность симптома S_i при диагнозе D_1 $P(S_i/D_1)$ » или «правдоподобие диагноза D_j при наличии симптома S_1 $P(S_i/D_j)$ ».

Явные вероятностные методы

Правдоподобие. Отношение правдоподобия

Для диагностики двух заболеваний наиболее удобно использовать понятие **отношения правдоподобия**:

$$L(D_1, S) = P(S/D_1)P(S/D_2) \quad (8)$$

Для двух диагнозов D_1 и D_2 и одного и того же симптома (комплекса симптомов) S из формулы Байеса (2) следует, что отношение апостериорных вероятностей диагнозов равно

$$\frac{P(D_1/S)}{P(D_2/S)} = \frac{P(S/D_1) P(D_1)}{P(S/D_2) P(D_2)} = L(D_1, S) \frac{P(D_1)}{P(D_2)} \quad (9)$$

Для условно независимых симптомов

$$\frac{P(D_1/S_1, S_2, \dots, S_N)}{P(D_2/S_1, S_2, \dots, S_N)} = \prod_{i=1}^N \frac{P(S_i/D_1) P(D_1)}{P(S_i/D_2) P(D_2)} = \prod_{i=1}^N L(D_1, S_i) \frac{P(D_1)}{P(D_2)} \quad (10)$$

Соотношение позволяет по отношениям правдоподобия диагнозов для разных симптомов и по априорным вероятностям заболеваний определить во сколько раз одно из них вероятнее другого.

Явные вероятностные методы

Диагностические коэффициенты

Для упрощения расчетов умножение вероятностей симптомов и заболеваний заменяют сложением их логарифмов.

При этом формула принимает вид

$$\lg \frac{P(D_1/S_1, S_2, \dots, S_N)}{P(D_2/S_1, S_2, \dots, S_N)} = \sum_{i=1}^N \lg \frac{P(S_i/D_1)}{P(S_i/D_2)} + \lg \frac{P(D_1)}{P(D_2)} \quad (11)$$

Для удобства при диагностике складывают не просто логарифмы, а диагностические коэффициенты (ДК), определяемые как

$$\text{ДК}_i = 10 \lg \frac{P(S_i/D_1)}{P(S_i/D_2)} = 10 \lg L(D_1/S_i) \quad (12)$$

В этом случае для ВД пользуются формулой

$$10 \lg \frac{P(D_1/S_1, S_2, \dots, S_N)}{P(D_2/S_1, S_2, \dots, S_N)} = \sum_{i=1}^N \text{ДК}_i + 10 \lg \frac{P(D_1)}{P(D_2)} \quad (13)$$

Явные вероятностные методы

Последовательная статистическая процедура

Пусть проводится ВД двух заболеваний D_1 и D_2 по формулам (10) и (13). Введем величины: α – вероятность ошибки первого рода (вероятность ошибочного предположения у больного заболевания D_2 , в то время как у него заболевание D_1); β – вероятность ошибки второго рода (вероятность ошибочного предположения у больного заболевания D_1 , в то время как у него заболевание D_2). Величины α и β выбираются на основании клинического опыта и здравого смысла. Ошибка первого рода считается более серьезной, чем ошибка второго рода, так как D_1 – всегда более серьезное заболевание. Задавшись значениями α и β , находят пороги диагностики заболеваний D_1 и D_2 .

При использовании формулы (10) порог диагностики (ПД) заболевания D_1 равен

$$\text{ПД}_1 = \frac{1 - \alpha}{\beta}$$

порог диагностики для заболевания D_2 равен

$$\text{ПД}_2 = \frac{\alpha}{1 - \beta}$$

При применении формулы (13) эти пороги равны соответственно:

$$\text{ПД}_1^* = 10 \lg \frac{1 - \alpha}{\beta} \quad \text{ПД}_2^* = 10 \lg \frac{\alpha}{1 - \beta}$$

Диагностический алгоритм с использованием последовательной статистической процедуры (ПСП) Вальда заключается в следующем. У пациента рассматривается произвольный симптом. Для него вычисляется отношение правдоподобия или ДК. Если это отношение больше, чем ПД_1 , ставится диагноз заболевание D_1 . Если меньше, чем ПД_2 – ставится диагноз заболевание D_2 . В том же случае, когда

$$\text{ПД}_2 < \frac{P(S/D_1)}{P(S/D_2)} < \text{ПД}_1 \quad \text{или} \quad \text{ПД}_2^* < \text{ДК} < \text{ПД}_1^*$$

учитывается следующий симптом, имеющийся у больного. Вычисляется отношение правдоподобия уже по обоим этим симптомам (складываются ДК обоих этих симптомов) и опять сравниваются с диагностическими порогами. Операция повторяется до тех пор, пока значение произведения отношений правдоподобия (суммы ДК) не выйдет за один из порогов. Если все симптомы у больного исчерпаны, а значение ни одного из порогов не достигнуто, то диагноз считается неопределенным. Преимуществом рассмотренного алгоритма является возможность постановки диагноза по небольшому числу симптомов. При этом симптомы учитываются последовательно по степени их важности (информативности). Если диагноз поставлен, нет необходимости учитывать менее важные симптомы. Применение ПСП создает предпосылки к использованию более трудоемких или более дорогостоящих диагностических методик на более поздних этапах диагностического процесса и даже позволяет в целом ряде случаев обходиться без них.

Явные вероятностные методы

Алгоритм, основанный на сложении коэффициентов корреляции симптомов

Корреляция симптомов с конкретным заболеванием подтверждается в результате разнообразных верификаций. Пусть проводится ВД между заболеваниями D_1 и D_2 , причем они являются единственно возможными и взаимоисключающими случаями. Для вычисления коэффициента корреляции определим по группе обучения следующие показатели:

a_i – частота наличия симптома S_i при заболевании D_1 ;

b_i – частота наличия симптома S_i при заболевании D_2 ;

c_i – частота отсутствия симптома S_i при заболевании D_1 ;

d_i – частота отсутствия симптома S_i при заболевании D_2 .

Коэффициент корреляции симптома S_i с подтвержденным диагнозом «заболевание D_1 » определяется по формуле

$$r_i = \frac{a_i d_i - b_i c_i}{\sqrt{(a_i + c_i)(b_i + d_i)(a_i + b_i)(c_i + d_i)}} \quad (14)$$

Для интерпретации r_i применяют следующую шкалу:

S_i слабо коррелирует с D_1 , если $0 < r_i \leq 0,3$;

S_i умеренно коррелирует с D_1 , если $0,3 < r_i \leq 0,5$;

S_i заметно коррелирует с D_1 , если $0,5 < r_i \leq 0,7$;

S_i сильно коррелирует с D_1 , если $0,7 < r_i \leq 1$;

S_i слабо коррелирует с D_2 , если $-0,3 \geq r_i > 0$;

S_i умеренно коррелирует с D_2 , если $-0,5 \geq r_i > -0,3$;

S_i заметно коррелирует с D_2 , если $-0,7 \geq r_i > -0,5$;

S_i сильно коррелирует с D_2 , если $-1 \geq r_i > -0,7$.

Алгоритм постановки диагноза с использованием коэффициентов корреляции симптомов заключается в следующем. Складываются r_i для всех симптомов, учитываемых у больного. Если их сумма превышает порог 0,5, ставится диагноз «заболевание D_1 ». Если она меньше чем $-0,5$, то ставится диагноз «заболевание D_2 ». В случае если $-0,5 < \sum_i r_i \leq 0,5$, диагноз считается неопределенным. Для улучшения качества ВД значение порога можно изменить, подобрав его по группе обучения.

Структура процесса вычислительной диагностики

Этап	Задачи
I. Сбор информации	Получение клинического материала, адекватного диагностической задаче
II. Отбор информации	1. Определение информативности симптомов. 2. Отсев малоинформативных симптомов
III. Классификация	1. Диагностика по группе контроля по отобранным симптомам. 2. Корректировка построенной системы ВД. 3. Применение системы ВД в клинике на основе выбранного диагностического алгоритма

Сбор информации

В общем случае этап сбора информации состоит из нескольких процедур:

- 1) постановка диагностической задачи;
- 2) поиск верифицированного материала (т. е. ИБ), адекватного диагностической задаче, и разбиение этого материала на группы «заболевание D1» и «заболевание D2»;
- 3) разбиение верифицированного материала на группу обучения и группу контроля;
- 4) составление карты признаков и симптомов, свойственных заболеванию D1 и заболеванию D2;
- 5) кодировка верифицированного материала группы обучения по карте признаков и симптомов.

Сбор информации

Постановка диагностической задачи

Постановка диагностической задачи состоит в разбиении рассматриваемых случаев на диагнозы, которые должны составлять полную группу событий.

Сбор информации

Поиск верифицированного материала

Поиск верифицированного материала заключается в отборе определенного количества ИБ с верифицированными диагнозами. Верифицированность обеспечивается данными вскрытий, биопсий, а также в ходе хирургических операций, в результате длительных наблюдений.

Отбор ИБ с верифицированным диагнозом должен проводиться в соответствии со следующими требованиями:

- а) однородность. Следует отбирать ИБ за несколько последних лет и только из тех лечебных учреждений, в которых существуют сходные взгляды на симптоматику, присущую выбранным заболеваниям;
- б) качественность. Не следует отбирать ИБ, в которых содержится совокупность симптомов, явно недостаточная даже для приблизительной постановки диагноза;
- в) диагностическая нетривиальность. Среди отобранных ИБ должны преобладать диагностически трудные, т. е. нетипичные случаи. Типичные случаи, преобладающие в клинике любого заболевания, для ВД интереса не представляют, зато в силу своего большого удельного веса они могут поглотить интересные случаи, ради которых и создается система ВД.

Сбор информации

Разбиение верифицированного материала на группу обучения и группу контроля

Разбиение верифицированного материала на группу обучения и группу контроля является необходимой процедурой, без проведения которой невозможно будет провести проверку качества работы системы ВД. В зависимости от объема материала процедуру разбиения проводят следующим образом. При недостаточном количестве ИБ в группу обучения включают весь имеющийся в наличии материал, а группу контроля формируют из поступающих в дальнейшем ИБ, в которых диагнозы уточняются со временем. При наличии большого количества ИБ группу обучения и группу контроля формируют одновременно. В этом случае к группе контроля предъявляется следующее основное требование: она должна быть отобрана из имеющегося материала случайным образом. Минимальный объем группы контроля, соответствующий предполагаемому уровню диагностических ошибок, легко определяется по известным из курса статистики соотношениям. Следует подчеркнуть, что при использовании для проверки системы ВД группы обучения вместо группы контроля оценки диагностических ошибок могут оказаться заниженными.

Сбор информации

Карта признаков и симптомов

Карта признаков и симптомов представляет собой вопросник, который включает в себя заранее составленные ответы на каждый вопрос. В ходе составления карты признаков и симптомов решаются две задачи: отбор признаков в карту и их градуировка, т. е. разбиение на симптомы. В карту включаются все признаки и симптомы, которые, по мнению врача, представляют определенную диагностическую ценность и традиционно проверяются при обследовании пациента. При их градуировке учитываются следующие основные правила.

1. Выделяются такие симптомы, условные вероятности (относительные частоты) которых при заболеваниях D1 и D2 различаются.
2. Следует избегать, насколько это возможно, слишком резких различий между условными вероятностями соседних симптомов.
3. Градуировку необходимо проводить таким образом, чтобы каждый симптом встречался не менее чем в 5–7 наблюдениях, относящихся к каждому заболеванию D1 и D2. Если какой-либо симптом встречался только в наблюдениях, относящихся к D1 (D2), и ни разу не встречался в наблюдениях, относящихся к D2 (D1), то необходимо условно положить, что он встречался в одном наблюдении, относящемся к D2 (D1). Карта составляется таким образом, чтобы у пациента обязательно был хотя бы один из симптомов каждого признака, что приводит, например, к включению в карту таких симптомов, как «норма по данному признаку». В то же время недопустимо наличие у пациента более одного симптома каждого признака.

Сбор информации

Кодировка материала группы обучения

Кодировка материала группы обучения должна позволить установить из закодированной ИБ:

- а) какие признаки проверялись для данного больного – наличие признаков;
- б) какие не проверялись – пропуск признака;
- в) какие симптомы имеются у данного больного – наличие симптома;
- г) какие симптомы отсутствовали у данного больного – отсутствие симптома;
- д) про какие симптомы, содержащиеся в карте признаков и симптомов, мы не можем сказать есть они или нет – пропуск симптома.

Отбор информации

Определение информативности симптомов

Задача определения информативности симптомов складывается из следующих процедур:

- 1) подсчет исходных данных;
- 2) расчет информативности симптомов по результатам подсчета исходных данных.

Отбор информации

Отсев малоинформативных симптомов

Как следует из теории распознавания образов, чрезмерное увеличение числа симптомов, по которым проводится ВД, может привести к снижению достоверности результатов ВД. Поэтому среди всех имеющихся симптомов отбирают только информативные, которые и будут использованы в выбранном алгоритме ВД.

Классификация

Этап классификации заключается в последовательном решении следующих задач.

1. Диагностика по группе контроля по отобранным симптомам.
2. Корректировка построенной системы ВД.
3. Применение системы ВД в клинике на основе выбранного диагностического алгоритма.

Материалы

Спиридонов И.Н., А. В. Самородов А.В. Методы и алгоритмы вычислительной диагностики
<https://press.bmstu.ru/catalog/item/2627/>