Министерство образования и науки Российской Федерации

Федеральное государственное автономное образовательное учреждение высшего профессионального образования

«Московский физико-технический институт

(государственный университет)»

Факультет радиотехники и кибернетики

Кафедра радиолокации, управления и информатики

Алгоритм и программное обеспечение нахождения весовых коэффициентов, определяющих степень значимости априорной информации о движущихся радиолокационных объектах

Выпускная квалификационная работа

(магистерская работа)

Направление подготовки: 03.04.01 Прикладные математика и физика

Выполнил:

студент 114а группы \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Сорокин Александр Игоревич

Научный руководитель:

к.т.н., \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Фащев Виталий Владимирович

Москва, 2017

Содержание

[ВВЕДЕНИЕ 3](#_Toc484098909)

[**1 Весовые коэффициенты** 6](#_Toc484098910)

[***1.1 Метод ранжирования*** 6](#_Toc484098911)

[***1.2 Метод приписывания баллов*** 8](#_Toc484098912)

[***1.3 Метод рандомизированных сводных показателей*** 9](#_Toc484098913)

[***1.5 Система весовых коэффициентов Фишберна*** 10](#_Toc484098914)

[***1.6 Статистические методы*** 10](#_Toc484098915)

[***1.6 Выводы главы*** 17](#_Toc484098916)

[**4 Результаты** 19](#_Toc484098917)

[**5 Заключение** 20](#_Toc484098918)

[**6 Приложение** 22](#_Toc484098919)

[**7 Список литературы** 23](#_Toc484098920)

ВВЕДЕНИЕ

Распознавание образов как научная дисциплина начала формироваться одновременно в СССР и США примерно со второй половины 50-х годов. Математическим аппаратом постановки и решения задач распознавания образов с момента их возникновения явилась теория статистических решений. Основы этой теории разработаны Дж. Нейманом и К. Пирсоном. Классические результаты теории статистических решений послужили базой для алгоритмов распознавания, обеспечивающих определение класса, к которому может быть отнесен неизвестный объект, основываясь на экспериментальных измерениях некоторого набора параметров (признаков), характеризующих этот объект, и определенных данных, описывающих классы рассматриваемых целей. Данные алгоритмы активно применяются в радиолокации [1].

Разработка систем распознавания, и в частности радиолокационного распознавания (РЛР), сопряжена с решением целого ряда задач [2]. В частности, существует актуальная задача - селекция (выделение) боевых блоков баллистических ракет на фоне ложных целей и в условиях помех и ядерных взрывов [3]. Опыт работ, провеленных в ГП «НИИРП» в 70-80-х годах 20 века, показал, что для решаения задачи селекции радиолокационных целей актуальным является наличие качественно составленного априорного описания всех движущихся радиолокационных целей – совокупность боевых блоков, тяжёлых ложных целей, ЛЛЦ и других средств преодоления системы ПРО [3].

В свою очередь задача распознавания в рамках радиолокации состоит, в разделении некой совокупности классифицируемых объектов на ряд классов в соответствии с выбранным принципом классификации, т.е. с некоторой вероятностью классифицировать объект-цель, снизив при этом вероятность ошибочного определения (ошибки 1-го, 2-го рода). В качестве принципа классификации – разбиения по классам используются различные признаки: качественные (тип двигателя), количественные (скорость, ускорение, ЭПР, размер объекта и т.д). Поэтому в описании классов должны содержаться сведения о качественных признаках объектов, о законах распределения имеющих количественные выражения, а также априорное описание – например количество объектов заданного класса.

Рассмотрим далее характер априорной информации, которая имеется при решении задач селекции и классификации. Как уже отмечалось, при классификации мы обычно располагаем сведениями о составе генеральной совокупности объектов, которые предстоит распознавать. Эти сведения содержат априорных вероятностей появление объектов различных классов. При селекции мы располагаем сведениями о составе конкретной выборки наблюдаемых объектов. Например, в наблюдаемой группе содержится ровно один объекта первого класса, в наблюдаемой группе содержится не менее одного и не более трех объектов первого класса.

Характер располагаемых априорных сведений определяет не только результат решения задачи селекции, но и саму структуру алгоритмов (решающих правил), применяемых при ее решении [3].

Как уже отмечалось ранее для решения задачи селекции актуальным является наличие качественно составленного априорного описания. С этой целью были разработаны таблицы априорных описаний, включающие в себя не только эталон для конкретного признака селекции, но и веса (весовые коэффициенты), отнесенные к этим признакам и определяющие степень доверия к заданному виду априорной информации.

При проведении работ использовался экспертный метод определения весовых коэффициентов. Его низкая эффективность заключалась в частых корректировках после каждой натурной работы, но компенсировалась достаточным количеством испытаний.

В современных реалиях нет возможности проводить натурные испытания достаточном объеме и как следствие необходимо автоматизировать расчет весовых коэффициентов.

В данной работы предложена методика автоматического расчета весовых коэффициентов. Целью данной работы является:

* Исследование методов расчета весовых коэффициентов
* Разработка алгоритма расчета весовых коэффициентов заданных классов радиолокационных объектов
* Разработка программного обеспечения для расчета весовых коэффициентов заданных классов радиолокационных объектов.

Также будет продемонстрирована эффективность разработанного метода по сравнению с другими алгоритмами.

**1 Весовые коэффициенты**

Для описания какого-либо процесса или явления, работы сложной системы или объекта используют, как правило, некоторый набор показателей, характеризующих эти процессы или объекты с разных сторон. С течением времени и под действием различных объективных и субъективных факторов эти показатели меняются, причем по-разному [1]. Поскольку роль показателей неодинакова, для оценки влияния часто используются весовые коэффициенты. В большинстве случаев весовые коэффициенты определяются на основании экспертных суждений, по определенным алгоритмам, сравнивающих факторы между собой [], но также существуют математические методы расчета, о которых будет подробно рассказано в данной главе.

После проведения сравнительного анализа будут рассмотрены статистические методы для выявления наиболее подходящего для решения поставленной задачи.

***1.1 Метод ранжирования***

Данный метод позволяет упорядочить компоненты по степени возрастания или убывания их влияния в зависимости от особенностей рассматриваемого события. Группа из 𝑛 экспертов, специалистов в исследуемой области, высказывается относительно важности 𝑚 частных показателей. Самому важному показателю соответствует ранг 𝑚,

следующему – (𝑚 − 1) и т.д., ранг, равный 1, имеет наименее важный показатель. Результаты опроса экспертов сводят в Таблицу 2.1, в последней строке которой записывают сумму рангов, выставленных экспертами.

*Таблица 2.1 ― Результаты опроса экспертов по рассматриваемым компонентам.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Номер эксперта | Рассматриваемые компоненты | | | |
| *X*1 | *X*2 | *…* | *Xn* |
| 1 | *r*11 | *r*12 | *…* | *r*1*n* |
| 2 | *r*21 | *r*22 | *…* | *r*2*n* |
| … | *…* | *…* | *…* | *…* |
| *M* | *rm*1 | *rm*2 | *…* | *rmn* |

Оценку важности той или иной компоненты проводит группа специализированных экспертов, и каждый из них представляет свой вектор оценок по данной группе компонент, основываясь на знаниях в области слабо формализованных задач. Компоненты расставляются в порядке их важности по следующему правилу:

1. Эксперт располагает компоненты по убыванию их важности слева направо.
2. Каждой компоненте присваивается оценка от *n* до 1 (самой важной – *n* и далее по убыванию до 1).
3. Для каждой компоненты высчитывается сумма оценок, и далее высчитывается доля от всех полученных сумм. В виде формулы это можно представить так:

|  |  |  |
| --- | --- | --- |
|  |  |  |

где – весовой коэффициент *j*-й компоненты; – оценка, поставленная *j*-й компоненте *i*-м экспертом [16].

Достоинством метода является его вычислительная простота, а недостатком – необходимость опроса экспертов, определение их необходимого числа, квалификации и т.д.

***1.2 Метод приписывания баллов***

Часто бывает желательным определить, насколько одна компонента более значима, чем другие. Поэтому этот метод основан на том, что эксперты оценивают важность частной компоненты по определённой шкале, например, от 0 до 10. Именно поэтому метод непосредственной оценки иногда именуют также балльным методом или методом прямой расстановки. При этом разрешается оценивать важность дробными величинами или приписывать одну и ту же величину из выбранной шкалы нескольким компонентам. Таблица оценок представлена так же, как и в предыдущем методе [14].

Алгоритм расчета весовых коэффициентов следующий:

1. Каждый эксперт проставляет оценки по всем компонентам в рамках заданной шкалы.
2. Происходит пересчет оценок по формуле:

|  |  |  |
| --- | --- | --- |
|  |  |  |

1. Далее полученные оценки для каждой компоненты суммируются и нормируются, также как и в предыдущем методе.

|  |  |  |
| --- | --- | --- |
|  |  |  |

***1.3 Метод рандомизированных сводных показателей***

Суть данного метода состоит в следующем. Строится дискретная модель неопределенности задания весовых коэффициентов [9, 10, 11], в которой предполагается, что каждый из этих коэффициентов измеряется с точностью до конечного шага , определяемого натуральным числом 𝑛 > 1. Таким образом, весовые коэффициенты могут принимать только дискретные значения:

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

Тогда множество всех возможных векторов весовых коэффициентов

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

Где есть множество возможных значений индекса , является конечным множеством, содержащих число элементов , равное:

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

Неопределенность выбора конкретного вектора весовых коэффициентов 𝑤(𝑡) из множества всех возможных векторов весовых коэффициентов 𝑊(𝑚, 𝑛) получаются случайным образом при помощи случайного индекса 𝑡̃, равномерно распределенного на множестве :

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

В результат получается случайный вектор весовых коэффициентов ( и равномерно распределенный на множестве 𝑊(𝑚, 𝑛).

При этом математическое ожидание и стандартное отклонение ( – дисперсия случайной величины ) – го весового коэффициента:

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

являются оценками весовых коэффициентов и метой их точности, соответственно.

***1.5 Формулы Фишберна***

Данный формулы позволяет позволяют определить весовые коэффициенты, если относительно показателей известна некоторая информация [11]. Они могут быть упорядочены по мере убывания их важности: . В этом случае весовые коэффициенты образуют убывающую арифметическую прогрессию и могут быть определены по формуле (первая формула Фишберна):

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

Также можно усилить простое линейное упорядочение:

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

Весовые коэффициенты образуют убывающую геометрическую прогрессию, а их значения определяются по формуле (вторая формула Фишберна):

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

Относительно весовых коэффициентов могут быть известны интервалы их возможных значений (интервальные соотношения упорядочения):  
 . В этом случае используется, так называемая, третья формула Фишберна:

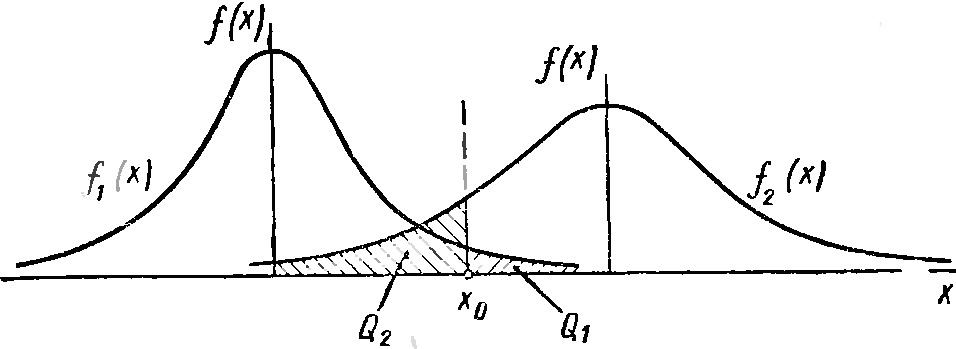
|  |  |  |
| --- | --- | --- |
|  |  | (3) |

***1.6 Статистические методы***

Говоря о статистических методах распознавания, мы предполагаем установление связи между отнесением объекта к тому или иному классу (образу) и вероятностью ошибки при решении этой задачи. В ряде случаев это сводится к определению апостериорной вероятности принадлежности объекта к определенному классу при условии, что признаки этого объекта приняли значения .

Основное преимущество статистических методов распознавания состоит в возможности одновременного учета признаков различной физической природы, так как они характеризуются безразмерными величинами – вероятностью их появления.

Широкий класс задач математической статистики связан с минимизацией функционала, называемого средним риском. Для решения задачи распознавания необходимо разделить исследуемое признаковое пространство на два полупространства, таким образом, чтобы в среднем по значению среднего риска было бы равно сумме потерь, связанным с правильными и неправильными решениями (рисунок 1).



*Рис. 1. Плотность распределения признаков исследуемых классов объектов*

Необходимо осуществлять выбор порогового значения , а как следствие и выбор стратегии, чтобы значение принимало минимальное значение. Далее рассмотрим основные стратегии, применяемые при решении задачи распознавания.

***1.6.1 Стратегия Байеса***

Стратегия Байеса обеспечивает самую низкую вероятность ошибки [5]. Данная стратегия выбора решения направлена на снижение значения среднего риска, т.е. решается задача минимизации среднего риска, значение которого находится через платежную матрицу, элементы которой соответствуют правильным и ошибочным решениям системы распознавания.

Предполагается, что при классификации имеется полное описание классов: плотность распределения вероятности , , априорные вероятности и матрицы потерь (1):

|  |  |
| --- | --- |
|  | (1) |

Суть Байесовского подхода в разработке такого классификатора, который будет правильно определять наиболее вероятный класс для пробного образа. Следовательно, задача состоит в определении наиболее вероятного класса.

Пусть задано классов , имеется ,– вероятность того, что неизвестный объект, представляемый вектором признаков , принадлежит классу .

Вероятностьназывается апостериорной, поскольку задает распределение индекса класса после эксперимента, т.е. после того как значение вектора признаков , было получено.

Следовательно, для решения поставленной задачи необходимо выбрать решающее правило таким образом: распознаваемый объект относиться к тому классу, для которого апостериорная вероятность выше [2]. Данное привило классификации называется Байесовским: если , то объект следует отнести к классу , в противном случае объект относят к классу. Следовательно, для Байесовского решающего правила необходимо получить апостериорные вероятности , которые можно найти из формулы Байеса.

|  |  |
| --- | --- |
|  | (2) |

где  априорные вероятности, показывающие принадлежность объекта к каждому из классов, а  функция правдоподобия  по отношению к .

В случае, когда неизвестны ни априорные вероятности, ни функции правдоподобия, их можно оценить методами математической статистики на множестве прецедентов. Например, , где  – число прецедентов из , а  – общее число прецедентов.

Выбор любой другой стратегии, отличной от байесовской, влечет за собой увеличения числа ошибок решений о принадлежности классифицируемых объектов к соответствующим классам, а, следовательно, к увеличению среднего риска [1].

Применительно к нашей задаче нельзя использовать Байесовский подход, так как для определения значения среднего риска необходимо знать матрицу потерь, однако формулу Байеса можно использовать для расчета апостериорной информации. Зная количество объектов каждого класса, можно вычислить априорную вероятность для каждого из классов. Расчет функции правдоподобия будет показан позже.

**Гарантирующая (минимаксная) стратегия**

При построении систем распознавания возможны такие ситуации, когда априорные вероятности появления объектов соответствующих классов неизвестны. Минимизировать значение среднего риска принятия решений на основе байесовской стратегии в этом случае не представляется возможным. Применительно к этой ситуации рационально использовать минимаксный критерий – критерий, который минимизирует максимально возможное значение среднего риска. Забегая вперед, можно также с уверенностью сказать, что данный метод также не подходит, так как мы не знаем матрицу потерь. Но для полноты анализа необходимо рассказать и про данный метод.

Его принцип заключается в том, что выбирается такая априорная вероятность, при которой значение байесовского риска будет максимальным.

При наличии классов  и  байесовский риск с учетом того, что  
, , a  и , равен:

|  |  |
| --- | --- |
|  | (3) |

Для определения алгоритма принятия решения, соответствующего минимаксной стратегии, продифференцируем (3) по ) и приравняем производную нулю. В результате получим:

|  |  |
| --- | --- |
|  | (4) |

где  и  ошибки первого и второго рода соответственно.

Данные значения определяют вероятность ошибочных решений и определяются , . Также ошибки первого рода можно выразить через ошибки второго рода:

|  |  |
| --- | --- |
|  | (5) |

Запишем выражение (4) как отношение:

|  |  |
| --- | --- |
|  | (6) |

Определение  позволяет записать алгоритм классификации так: , если и , если .

Если , то, как следует из (4), минимаксная стратегия приводит к равенству условных вероятностей ошибок первого и второго рода.

**Стратегия Неймана-Пирсона**

Во многих физических ситуациях бывает затруднительно предсказать достаточно реалистичные априорные вероятности. В тех случаях, когда нам известна лишь функции плотности  рассматривают стратегию Неймана-Пирсона.

Основной критерий – ограничить одну из вероятностей неправильного принятия решения и максимизировать (или минимизировать) другую. Однако при этом некоторые значения ошибок жестко задаются наблюдателем и строится наилучшее при таких условиях разбиение, оптимизирующее совокупность оставшихся свободных элементов.

В зависимости от принимаемого решения при распознавании движущегося радиолокационного объекта находится допустимое значение ошибки первого рода, которое не должно превышать какого-то значения . При данном ограничении  требуется найти решение  задачи .

Этим условиям удовлетворяет уравнение , из которого находиться . Если выбрать ,  будет возрастать. При , возрастет , что видно на рисунке 1.

Данная стратегия подходит для решения поставленной задачи.

**Выбор стратегии**

После проведенного анализа становится понятным, что для решения поставленной задачи необходимо использовать стратегию Неймана-Пирсона. Для каждого из исследуемых параметров объекта необходимо задать максимальные значения ошибок первого и второго рода. Далее, так как нам известны плотности распределения вероятностей, для каждого из параметров необходимо решить уравнение , после чего будет найдено пороговое значение  для каждого из параметров объекта.

Будем использовать формулу Байеса (2) для нахождения апостериорной вероятности принадлежности объекта к одному из классов. Покажем расчет функции правдоподобия.

Закон распределения вероятностей наблюдаемой p-мерной случайной величины  описывается функцией , зависящей от неизвестного параметра , причем будем понимать под  вероятность , если  дискретна, и значение плотности вероятности в точке , если  непрерывна. Поскольку при гипотетическом варианте понимание случайной выборки  суть независимые и одинаково распределенные случайные величины, то для любого заданного набора значений  их совместная плотность (вероятность) будет

|  |  |
| --- | --- |
|  | (7) |

Таким образом, функция , определенная равенством (7), задает вероятность получения, при извлечении выборки объема  именно наблюдений . Поэтому, чем больше значение, тем правдоподобнее (или более вероятна) система наблюдений  при заданном значении параметра  [4]. Параметр  может быть представлен, как набор величин .

Так как функция правдоподобия зависит от плотности вероятностей, значения  должны быть параметрами уравнения плотности распределения.

Например, для нормального распределения  параметрами будут являться , а значениями выборки, в нашем случае, параметры объекта .

Так как плотности распределения по условию задачи нам известны, функция правдоподобия примет вид:

|  |  |
| --- | --- |
|  | (8) |

Подставив выражение (8) в формулу Байеса можно найти апостериорную вероятность и, как следствие, решить задачу отнесением движущихся радиолокационных объектов к тому или иному классу (образу).

***1.6 Выводы главы***

были рассмотрены два способа расчета весовых коэффициентов. Анализируя вычислительную сложность обоих методов, можно сделать вывод, что для метода ранжирования требуется меньшее количество элементарных арифметических операций, чем для метода непосредственной оценки.

Показатель эффективности для обоих случаев одинаков, так как зависит от количества экспертов и является функцией от параметра *m.* Следует также заметить, что оба способа по показателю качества дают невысокую точность измерения, поскольку возможна низкая согласованность экспертов.

Метод непосредственной оценки лучше применять в том случае, когда степень важности компонент сильно варьируется (в некоторых случаях компонента может вообще не учитываться). Метод ранжирования предназначен для тех случаев, когда необходима «строгая» расстановка приоритетов компонент, а также, когда нет возможности дать точную оценку важности между рассматриваемыми компонентами.

В некотором роде метод ранжирования можно рассматривать как частный случай метода непосредственной оценки.

Поскольку оба метода основаны на массовую теоретическую оценку, их применение в данной работе будет невозможна. Причиной отказа от этих способов расчета может также послужить и то, что в алгоритмах самообучения системы необходимо использование результатов работы предыдущих тактов системы. Поэтому возникает необходимость в разработке правильного алгоритма перерасчета весовых коэффициентов.

**4 Результаты**

**5 Заключение**

В данной работе была предложена методика автоматизированного расчета весовых коэффициентов.

В ходе обзора статистических методов были рассмотрены три основные стратегии и показан наилучший метод классификации – стратегия Байеса. Проведя анализа известных методов расчетов весовых коэффициентов, была поставлена задача реализации нового подхода сокращающего время расчетов весов.

В ходе разработки проведен вывод формулы для пересчета весовых коэффициентов с использованием формулы Байеса. На ее основе был предложен алгоритм обучения, а также техническая реализация алгоритма классификации радиолокационных целей.

В третьей части работы рассмотрены технические и программные особенности системы. Проведен анализ по выбору БД, наиболее соответствующей требованиям системы. Также показаны преимущества SQLite и её подключение к проекту. Рассмотрен интерфейс системы с кратким пояснением его функционирования. Описана программная реализация вышеизложенных алгоритмов.

Результатом данной работы является методика, а также программное обеспечение нахождения весовых коэффициентов, определяющих степень значимости априорной информации. Было успешно выполнена задача сокращения времени расчета весов, также показан недостаток, такой как ресурсозатратность.

Таким образом, результатом выпускной квалификационной работы является созданная автоматизированной система классификации радиолокационных целей, в основе которой лежит байесовский классификатор. К дополнительным возможностям данной системы можно отнести самообучение, а также хранение всех результатов селекции.

**6 Приложение**

**7 Список литературы**

1. Ширман Я.Д., Манжос В.Н. Теория и техника обработки радиолокационной информации на фоне помех. М.: Радио и связь, 1981. 416 с.
2. Селекция и распознавание на основе локационной информации./ А.Л. Горелик, Ю.Л. Барабаш, О.В. Кривошеев, С.С. Эпштейн; Под. ред.  А.Л. Горелика. – М.: Радио и связь, 1990. – 240 с.
3. Щит России: системы противоракетной обороны. — М.: Издательство МГТУ им. Н.Э. Баумана, 2009. — 504 с.
4. Селекция и распознавание на основе локационной информации /А.Л. Горелик, Ю.Л. Барабаш, О.В. Кривошеев, С. Эпштейн / Под ред. А.Л. Горелика. - М.: Радио и связь, 1990. - 240 с.
5. Френкс, Л. Теория сигналов : перевод с англ. / Л. Френкс ; под ред. Д. Е. Вакмана. – М. : Советское радио, 1974. – 344 с.
6. Горелик А.Л., Скрипкин В.А. Методы распознавания. М.: Высшая школа, 1977. — 262 с.
7. Фукунага К. Введение в статистическую теорию распознавания образов.: Пер. с англ. М.: Наука, главная редакция физико-математической литературы, 1979. - 368 с.
8. Чернова Н. И. Математическая статистика: Учеб. пособие / Новосиб. гос. ун-т. Новосибирск, 2007. 148 с.
9. Блекуэлл Д., Гиршик М.А. Теория игр и статистических решений. М.: ИЛ, 1958.-374 с.
10. Татузов А.Л. Нейронные сети в задачах радиолокации. М.: Радиотехника, 2009. 432 с.
11. Сенин А.Г. Распознавание случайных сигналов. Новосибирск: Наука. Сибирское отделение. 1974. 76 с.
12. Фомин В.Н. Рекуррентное оценивание и адаптивная фильтрация. — М.: Наука. 1984. 288 с.
13. Цифровая обработка изображений в информационных системах: Учебное пособие / И.С. Грузман, В.С. Киричук и др. - Новосибирск, НГТУ, 2002. - 352 с.