УНИВЕРСАЛЬНЫЕ РЕШАЮЩИЕ ПРАВИЛА: ПОДДЕРЖКА И ПРИНЯТИЕ РЕШЕНИЙ В ЗАДАЧАХ РАСПОЗНАВАНИЯ, ДИАГНОСТИКИ, ВЫБОРА, КЛАССИФИКАЦИИ

Крисилов А.Д., Крисилов В.А., Чумаченко В.В.

ПОСТАНОВКА ЗАДАЧИ И ИСХОДНЫЕ ПОЛОЖЕНИЯ

Где ни просвищет грозный меч,

Где конь горячий ни промчится, …

(А. Пушкин)

Принятие и обоснование решений является неотъемлемой частью целого ряда задач, относимых к области искусственного интеллекта. В задачах классификации и распознавания образов, в задачах выбора и оценки оказывается необходимым принимать решения, обосновывать выбор и т. д. (см., напр., [1, 2] и др.) В подавляющем числе случаев решающие правила строятся в предположении о независимости признаков, составляющих описание распознаваемых или анализируемых объектов. Желание учесть зависимости между элементами описания требует весьма больших затрат времени и памяти и в итоге приводит, как правило, к учету лишь парных зависимостей. Мало того, достаточно редкими являются случаи, когда разработчики или пользователи обладают информацией о реальных величинах условных вероятностей появления признаков (друг от друга) в различных процессах, в распознаваемых классах etc.

В то же время довольно трудно представить задачу, в которой описание распознаваемых объектов складывалось бы из признаков, независимых в совокупности. В медицинской или технической диагностике, при машинном чтении текста или в природо-охранных задачах - всюду признаки, составляющие описание анализируемых объектов, связаны друг с другом довольно большими условными вероятностями появления-непоявления (см. напр., [3, 4] и др.). Применение в этих случаях решающих правил, работающих с признаками как с независимыми, приводит к ошибкам, тем большим, естественно, чем больше признаки зависят друг от друга.

Таким образом, существует задача: построить такие решающие правила, кото-рые в возможно большей мере учитывали бы связи и зависимости между признаками - без привлечения таких громоздких методов, как полный перебор или сходные с ним.

Построение требуемой решающей функции (или набора функций) удобно провести на примере задачи распознавания/классификации [5].

Рассмотрим множество подлежащих распознаванию ситуаций, состоящее из S подмножеств или классов:

$$M = \{M_j\};\ j = 1,\ldots, S \tag{1}$$

В нашем случае понятие "класс" объединяет некоторую совокупность объектов (ситуаций), для которых определено понятие близости или сходства.

Каждый класс описывается совокупностью параметров или признаков общим числом n. Это означает, что в общем случае описание данного класса (и данной ситуации) включает в себя n компонент:

$$Q = \{Q_i\};\ i = 1,\ldots, n \tag{2}$$

Каждый признак Q_i в том или ином объекте характеризуется определенной степенью выраженности v_i; для простоты будем считать признаки двоичными, на общность наших построений это не повлияет.

Обозначим каждую анализируемую ситуацию символом f_k, где k - текущий номер данной рассматриваемой ситуации или объекта. Данная ситуация в терминах нашего описания (2) выражена следующим образом:

$$f_k = \{v_1,\ldots, v_i,\ldots, v_n\}. \tag{3}$$

Для принятого упрощения объект f_k представлен в пространстве признаков двоичным n-мерным вектором.

Задача состоит в том, чтобы построить распознающую систему W (в виде автономного устройства или программы), могущую принимать решения об отнесении очередной анализируемой ситуации к определенному подмножеству M_j из выражения (1).

Это означает, что нужно найти, сформировать такое решающее правило, такое количественное обоснование, которое бы наилучшим образом позволяло отнести каждую из рассматриваемых ситуаций к тому или иному классу. Слова «наилучшим образом отнести» означают, в частности, что оптимальность решающих правил является понятием относительным, и нужно признать определенную зависимость выбираемых (конструируемых) решающих правил от целого ряда внешних условий. Среди таких условий могут быть требование минимального риска или максимальной близости к эталону и т. д. Наиболее распространенным является требование сохранения минимальной ошибки, то есть максимальной достоверности даваемых системой ответов. В настоящей работе уделяется внимание именно таким решающим правилам.

СОСТАВЛЕНИЕ ОПИСАНИЯ ОБЪЕКТА И ПРИНЯТИЕ РЕШЕНИЙ

Сперва производится обучение системы W различению предполагаемых классов, например, путем показа реальных известных объектов или ситуаций и запоминания их характеристик, т. е. формирования эталонов.

В результате, описание внешнего мира в достаточно простом случае может быть представлено в нашей распознающей системе набором или одной из матриц вида

$$C_n = \left\| p_{ji} \right\|, \tag{4}$$

где строки пронумерованы по j (классы), а столбцы – по i (признаки).

Здесь p_{ji} представляет собой число, могущее отражать различные величины: информационный, частотный, диагностический (экспертный) или другой вес i-того признака для j-того класса, его нечеткую оценку и т. д. Каждая строка этой матрицы представляет собой эталонное описание класса M_j, которое и нужно хранить в памяти устройства и с которым будет сравниваться неизвестный распознаваемый объект.

Этап принятия решения об отнесении неизвестной ситуации к тому или иному классу, состоит в следующем.

Входной вектор $f_к$, (состоящий из единиц и нулей, (выражение (3)), оценивается по степени сходства с каждой из строк матрицы C_n, выражение (4).

Эта оценка (обозначим ее X_{kj} – оценка $к$-того входного объекта на j-том эталоне) производится по тому или иному критерию и имеет смысл меры близости к данному эталону, степени сходства с ним, вероятности порождения данного вектора данным классом и т. д., - в зависимости от выбранного характера описания и других обстоятельств. Назовем это число интегральной характеристикой данного входного объекта относительно данного эталона. В упоминавшемся ранее пространстве признаков эта оценка есть расстояние между концом данного входного вектора и точкой (или областью), соответствующей описанию данного класса (см., например, [5, 6, 7] и др.).

После вычисления X_{kj} для всех j оказывается возможным реализовать следующее решающее правило:

$$f_k \in M_j \,/\, R_f = \operatorname*{extr}_j X_{kj}, \tag{5}$$

то есть «данный объект принадлежит подмножеству M_j, если оценка принимает экстремальное значение (R_f) на эталоне с номером j». Если мы оцениваем в указанном пространстве расстояние до «идеального» представителя данного класса, экстремум должен иметь смысл минимума; если ищем вероятность порождения данного вектора данным классом – смысл максимума и т. д.

Таким образом, решающее правило (5) требует непосредственного численного определения степени сходства данного неизвестного объекта или ситуации с каждым из эталонов, запасенных в памяти системы, и указывает в качестве ответа тот из классов, с которым это количественно измеренное сходство оказывается наибольшим.

РАЗНОВИДНОСТИ РЕШАЮЩИХ ПРАВИЛ И УЧЕТ ЗАВИСИМОСТЕЙ МЕЖДУ ПРИЗНАКАМИ

Принятие решения по правилу (5) для случая, когда параметры рассматриваемых ситуаций имеют вероятностную природу, целесообразно производить следующим образом [3, 5]:

$$f_k \in M_j / R_f = \max_j P(M_j / f_k) = \max_j \prod_i^n p_{ji}^{(\nu_i)}, \qquad (6)$$

где: $P(M_j / f_k)$ – вероятность принадлежности данного вектора f_k к классу M_j;

$p_{ji}^{(\nu_i)}$ - равняется p_{ji} , если $\nu_i = 1$ и равняется $1 - p_{ji}$, если $\nu_i = 0$ здесь подразумевается, что символ p_{ji} из выражения (4) обозначает вероятность появления данного признака в представительной совокупности объектов данного класса.

Отметим, что в решающем правиле (6) величина $P(M_j / f_k)$, подсчитываемая как произведение для данного f_k всех $p_{ji}^{(\nu_i)}$ при фиксированном j, собственно, и представляет собой оценку X_{kj} из выражения (5). В нашем случае эта величина есть скалярное произведение вектора f_k (то есть конкретной входной реализации) на j-тую строку матрицы C_n. Отметим также, что при вероятностной постановке задачи нет более весомых критериев для принятия правильного решения, чем решающее правило (6).

Итак, в выражении (6) величина R_f указывает класс, который мог породить данный вектор f_k с наибольшей вероятностью, - если признаки были независимы.

Для учета зависимостей между признаками введем вторую стадию обучения [5, 6]. Будем снова предъявлять системе W множество известных объектов класса M_j и для каждого объекта вычислять X_{kj}, то есть его оценку на своем эталоне. Далее будем фиксировать частоту (или вероятность) появления различных величин $X_{kj.}$

Результатом второй стадии обучения для данного M_j будет кривая распределения вероятностей (КРВ) $P(X_{kj}/M_j)$ различных оценок X_{kj} всех предъявленных f_k є M_j на эталоне своего класса (здесь є – символ принадлежности).. Поступив так же с другими классами, получим набор КРВ для всех подлежащих анализу классов решений.

Теперь для принятия решения может быть предложено следующее правило:

$$f_k \in M_j / R_\Sigma = \max_j P(X_{kj} / M_j). \qquad (7)$$

Здесь решение принимается не по максимуму оценки данного f_k на данном эталоне, а по максимуму вероятности данной оценки для соответствующего класса решений [5].

Анализируя процедуру обучения, можно сказать, что механизм формирования КРВ выявляет и закрепляет количественные связи между признаками, имеющие место внутри данного класса ситуаций: если какие-то группы признаков являются связанными, более частыми (или наоборот), то это отразится на форме соответствующей КРВ и, что, собственно, и важно, - будет учтено решающим правилом (7). Разработчик может не знать существующих условных вероятностей между признаками, - вводимая стадия обучения (и – соответствующее правило принятия решения) снимают эту проблему.

По характеру немонотонности этих кривых можно судить о величине зависимостей между признаками (и, кстати, использовать КРВ как инструмент для работы с ними – отбора, ранжирования и т. д.).

Однако этого мало. Мы можем продолжить вторую стадию обучения: будем взвешивать объекты класса M_j на эталонах/строках других классов (обозначим для этого случая текущие номера других классов через *r: r* =1, *…, s)* и сформируем соответствующие КРВ. Теперь описание каждого класса построено как бы по принципу дополнительности, - в памяти решающей системы имеются слепки событий, пропущенные не только через «свои» входы (в нашем случае – это взвешивание на своих эталонах), но и через входы/эталоны других классов, как бы через другие рецепторные каналы. При этом ситуация является симметричной.

Теперь эталоном класса является совокупность

$$B_j = \{P(X_{kr} / M_j)\} \tag{8}$$

кривых (функций) распределения вероятностей оценок X_{kr} векторов этого класса (т. к. j фиксировано) на каждой строке матрицы C_n.

Многомерный вариант решающей функции теперь принимает вид:

$$f_k \in M_j / R = \max_j \prod_r \left[P(X_{kr} / M_j)\right] \tag{9}$$

Согласно этому решающему правилу процедура распознавания неизвестного вектора $f_k = \{ v_1 ,…, v_i ,…, v_n \}$ состоит в следующем:

а) находятся интегральные оценки X_{kj} для данного вектора f_k для всех j;

б) для каждого j определяем по кривым B_j (8) величины $P(X_{kr})$ для всех r;

в) для каждого j определяются произведения найденных в п. (б) величин;

г) максимальное значение произведения указывает номер класса, к которому следует отнести неизвестный вектор.

В силу ряда обстоятельств имеет место тот факт, что разные векторы f_k получают одну и ту же оценку X_{kj}. Происходит это вследствие представления вероятностного диапазона конечным числом градаций, из-за наличия p_{ji} = 0,5 и т. п. Вследствие этого даже при достаточно представительной обучающей выборке система не сможет в процессе экзамена узнавать каждый вектор «в лицо», а только по вероятности оценки для группы векторов. Возможные при этом ошибки в распознавании при применении правила (7) являются, таким образом, платой за отказ от полного перебора.

Однако, если эти разные векторы, одинаково оцениваемые на одной строке матрицы C_n, взвешивать на разных строках, как это предусматривается выражением (9), то они будут получать различные оценки X_{kr}. Поэтому решающее правило, учитывающее вероятности оценок X_{kr} (9), обеспечивает меньшее количество ошибок при распознавании, чем правило (7). Это происходит потому, что многомерные распределения (8), получаемые во второй стадии обучения, более полно реализуют информацию о классах, содержащуюся в матрице C_n , чем это происходит в правиле (7), - более полно используется информация, которая имеется в обучающей последовательности.

Рассмотрим теперь величину

$$\mathrm{Y_{kj}} = \left|\mathrm{X^0_{kj}} - \mathrm{X_{kj}}\right|,$$

которая характеризует отклонение веса, подсчитанного для неизвестного вектора f_k , от наиболее вероятного в этом классе значения X^o_{kj}:

$$\mathrm{P}(\mathrm{X^o_{kj}}) = \max_{k} \mathrm{P_{rk}}(\mathrm{X_{kj}} / \mathrm{M_j}); \qquad k \in j.$$

Тогда, распространив эту оценку на многомерный случай:

$$\mathrm{Y^o_{kj}} = \{\left| (\mathrm{X^o_{kj}} / \mathrm{M_r}) - (\mathrm{X_{kj}} / \mathrm{M_r})\right|\}; \qquad r, j = 1,\dots, s,$$

можно предложить такую модификацию правила (9):

$$\mathrm{f_k} \in \mathrm{M_j} / \mathrm{R^o} = \min_{j} \prod_{r} \left[\left|(\mathrm{Y^o_{kj}} / \mathrm{M_r}) - (\mathrm{Y_{kj}} / \mathrm{M_r})\right|\right], \tag{10}$$

Упрощенным вариантом правила R^o (10) является выражение:

$$\mathrm{R_r} = \min_{r} \mathrm{Y_{kj}}, \tag{11}$$

которое при определенном характере КРВ (например, наличие одного экстремума) может явиться вполне приемлемой заменой правила (10), давая при этом значительную экономию памяти, времени и т. д.

В дальнейшем при конструировании решающего устройства или программного обеспечения задач принятия решений необходимо оценить, оправданы ли для достижения выигрыша в уменьшении ошибок необходимые усложнения (запоминание кривых, увеличение объема вычислений и т. д.). При использовании данных по выбору числа градаций вероятности можно получить достаточно простую систему, тем более, если реализуется правило (11). В дальнейшем очень большой интерес представляет сравнение предлагаемых модификаций решающих правил в различных задачах на больших массивах исходных данных.

ЗАМЕЧАНИЯ О ПРАКТИЧЕСКОЙ РЕАЛИЗАЦИИ И ПРИМЕРЫ ПРИМЕНЕНИЯ

При реализации любого из рассмотренных решающих правил представляется целесообразным применить некоторые технические рабочие приемы, повышающие качество решающей системы, облегчающие общение с пользователем и т. д.

Первый из таких приемов связан с желанием предоставить системе возможность отказываться от решения в сомнительных ситуациях, если, например, при чтении текста устройство вместо буквы видит кляксу.

В этом случае можно рекомендовать установление порогов снизу на величину X_{kj}, требование ощутимого расстояния до ближайшего конкурента и т. д. Введение та-кого порога дает возможность системе еще и оценивать свои ответы: при каких-то зна-чениях X_{kj} давать определенный ответ, при других – сопровождать его сигналом сомне-ния, при третьих – отказываться от ответа. Сам такой отказ означает введение «нулево-го» класса, то есть j = 0, 1, …., s. Подобный прием в различных вариантах реализуется сейчас в разных распознающих программах.

Далее, в ряде случаев (при диагностике, при социально-экономическом анализе и др.) полезно, кроме основного ответа, указывать ближайшее конкурирующее реше-ние. Реализация этого приема существенно расширяет возможности системы, особенно, если к тому же выдавать количественные оценки предлагаемых решений.

Отдельно должен быть рассмотрен вопрос о том, какое число N градаций веро-ятности появления признака должна различать система. Практика показывает, что для многих практических применений вполне достаточно трех-пяти градаций с неравно-мерным разбиением всего диапазона вероятностей (среднюю градацию выбирать больше, крайние – поменьше; сами эти величины легко могут быть рассчитаны и т. д.).

В качестве одного из вариантов этого последнего предложения может быть рас-смотрена для решающих правил (7) и (9) ступенчатая аппроксимация полученных КРВ, например, П-образной формой.

Описанные в настоящей работе решающие правила и их различные модифика-ции были применены при решении ряда задач из различных предметных областей.

В задачах медицинской дифференциальной диагностики с применением ре-шающих правил (6) и (7) осуществлялся диагноз туберкулеза легких [8], активного и неактивного ревматизма [9], других нозологических форм. В первом случае достовер-ность распознавания составила (для S=3, n=31, решающее правило (6)) около 90% при 6,3% отказов от распознавания. Во втором (для S=5, n=26, весовые коэффициенты для симптомов по Джонсу – 2:1, экзаменационная выборка – 170 историй болезни) было получено для модифицированного варианта решающего правила (7) 93,3% правильных ответов при 8,3% отказов от постановки диагноза; врачебный диагноз на том же мате-риале дал 82,4% и 11% соответственно.

Значительное количество опытов было связано с распознаванием машинопис-ных и рукописных букв и цифр. Для этих знаков составлялось описание в терминах геометрических признаков (например, вертикальные, горизонтальные и наклонные линии), а затем проводилось обучение и классификация с помощью различных РП. Опишем кратко результаты одного из этих экспериментов. На различных рукописных цифрах, вписанных в габаритную рамку десятью различными людьми, не знавшими о цели эксперимента, без особых ограничений на написание, были опробованы РП (6), (7) и (9) в различных модификациях. Число знаков, взятых для эксперимента, - 1000; число признаков n = 36; v =2; S = 10. При формировании КРВ использовался дискретный вариант представления; производилось выделение признаков, составление эталонов и последующее распознавание. Результаты распознавания для трех названных РП соответственно составили 84,5%, 85,9% и 93,1% [3, 5].

Кроме того, отдельно видоизмененное решающее правило (5) было опробовано при распознавании слабостилизованных рукописных цифр с применением идеологии и аппарата размытых множеств. Эта идеология была применена на двух уровнях принятия решений: для выделения признаков и для узнавания знаков. Результат распознавания составил 84% правильных ответов [10]. Помимо того, что для слабостилизованных рукописных цифр этот результат хорош и сам по себе, - он еще иллюстрирует возможности применения строгих методов в плохо формализованных областях.

Наряду с описанными были проведены эксперименты по распознаванию с по-мощью РП (6) и (7) звуковых команд (данные ИППИ АН СССР) и по разработке реко-мендаций для капитанов сухогрузных судов по нормативам скорости их движения в рейсе (данные Черноморского Морского Пароходства). И здесь качество ответов со-ставляло 90 - 95% [5].

Среди других областей применения, кроме перечисленных, следует назвать рас-познавание нефтеносных и водоносных слоев по данным геофизической разведки (145 групп замеров в пластах Бугульминского месторождения; применялись две модифика-ции решающего правила (7)), а также некоторые задачи экологического и социально-экономического анализа, в которых решающие правила, сходные с (6) и (7), использо-вались в целях квалиметрии [3, 6, 7, 11]. Определенная группа вопросов методологии, в частности, - по построению решающих правил, рассмотрена в работах [11 - 15].

Во всех перечисленных случаях, хоть и в разной степени, введение второй ста-дии обучения и, соответственно, применение решающих правил, учитывающих зави-симости между признаками, - существенно уменьшало число ошибок и повышало каче-ство принимаемых решений.

Полученные результаты показывают, что предложенные алгоритмы количест-венного обоснования решений успешно могут работать в различных сложных ситуаци-ях и найдут применение в задачах управления и диагностики, в экономике, природо-охранной деятельности и т. д.

ЛИТЕРАТУРА

1. R. Schlaifer. Analysis of Decisions under Uncertainty. McGrow Hill, 1979.

2. R.B. Banerji. A Language for Pattern Recognition. Patt. Rec., 1978, 1.

3. А.Д. Крисилов. Модифицированные решающие функции для распознавания сложных ситуаций. Тр. УII Всесоюзн. Симп. по теоретич. киберн., Т.3, Тбил., 1974.

4. V. Gladun, N. Vashchenko. Analytical Processes in Pyramidal Networks. Proc. of 5-th Intern. Conf. ITA 2000, Foi-Comm., Sofia, 2000.

5. А.Д. Крисилов. Алгоритмы количественного обоснования решений для систем, работающих в условиях неопределенности. Знание, Киев, 1984.

6. A. Krissilov, V. Krissilov, A. Shutko. Decision Making Procedures that Operate with Dependent Features and Their Environmental Applications. Proc. of 5-th Intern. Conf. ITA 2000, Foi-Comm., Sofia, 2000.

7. V.A. Krissilov, A.D. Krissilov, R.A.Tarasenko. Transformation of Object Feature Space Under the Goal of Evaluation. Proc. of Conf. IPMU’98, Paris, 1998, pp.1901-.1908.

8. А.И. Штейнберг, А.Д. Крисилов. Решение задачи дифференциальной диагности-ки туберкулеза легких с помощью гибридных решающих функций. «Кибернети-ка и вычислительная техника», вып. 59, Киев, 1983.

9. A.A. Popov, A.D. Krissilov et al. On Automation of Medical Diagnostic Procedures. Proc. of IFIP-71 Congr., TA-7, Nederl., 1971, pp.841-847.

10. А.Д. Крисилов, И.В. Хихловская, Н.Н. Гогунская. Экспериментальная проверка некоторых алгоритмов описания и распознавания применительно к стилизован-ным рукописным знакам. Тр. 5-ой Всесоюзн. конф. «Автоматизация ввода пись-менных знаков в ЦВМ», Т.2, Каунас, 1984.

11. A.D. Krissilov. Towards a New Econo-Ecological Order for the Black Sea Region.// Int. Leadership Sem., Internat. Ocean Institute – Black Sea Operational Centre, Mamaia, Romania, Sept., 1999, pp.87-96.

12. А. Крисилов, Е. Соловьева, А. Уемов. Краткий методологический меморандум – ч. I. Information Science & Computing. International Book Series: Knowledge – Dialogue – Solution, N. 15. – ITHEA, Sofia, 2009.

13. А. Крисилов, А. Уемов. О шагах системного синтеза (продолжение «Методологического меморандума» – часть II). Information Models of Knowledge, v. 19. Proc. of Intern. Confer. "Knowledge – Dialogue – Solution" (KDS – 2010), Kiev, Sept. of 2010, ITHEA, Kiev – Sofia, 2010.

14. А. Крисилов. «Шаги системного синтеза и построение Интеллектуального Агента в задачах представления знаний», Труды Междун. Конфер. «КДС - 2010». Варна, июнь 2010.

15. А. Крисилов. «Содержание некоторых этапов системного анализа и синтеза (продолжение «методолог. меморандума» – ч. III») Proc. of Intern. Confer. "III – 2011"), ITHEA, Sofia, 2011.

Информация об авторов

Крисилов А.Д. *– Институт проблем рынка и экономико-экологических исследований НАН Украины, 65044, Одесса-44, Французский б-р, 29,*

e-mail: adkrissilov@list.ru

Крисилов В.А. *– Одесский Государственный Политехнический Университет, 65044, Одесса-44, пр. Т. Шевченко, 1;*

e-mail: krissilovva@gmail.com

Чумаченко В.В. *- ООО “GERC” Одесса-29, ул. Канатная, 48,*

e-mail: viachesv@chumachenko.net

UNIVERSAL DECISION RULES: SUPPORT AND DECISION-MAKING IN RECOGNITION, DIAGNOSIS, SELECTION, CLASSIFICATION

Krisilov A.D., Krisilov V.A., Chumachenko V.V.

Abstract*: The decision-making functions are described, that allow take into account relations and de-pendence among the features in such tasks as classification, choice, pattern reco-gnition and evaluation. The algorytms proposed in this paper provide with high level of decision support and situations recognition under uncertain conditions. Some examples are given.*