

УДК 004.891.2:004.93'1

С.В. Страхова, ст. преподаватель

Крымский юридический институт

Национального университета внутренних дел

ул. Стевена, 14, г. Симферополь, Украина, 950053

E-mail: svestr@univ.crimea.ua

ОЦЕНИВАНИЕ ВЕРСИЙ В ПРОЦЕССЕ РАССЛЕДОВАНИЯ ПРЕ- СТУПЛЕНИЙ КАК ЗАДАЧА РАСПОЗНАВАНИЯ ОБРАЗОВ

Проводится формальная постановка задачи оценивания версий, выдвигаемых при расследовании преступлений, как задачи распознавания образов; предлагается и обосновывается использование вероятностного подхода к оцениванию версий; рассматриваются проблемы обеспечения корректности использования процедуры Байеса и снижения размерности пространства признаков, указывается на один из возможных путей их решения.

Одной из приоритетных задач противодействия преступности является повышение эффективности процесса расследования преступлений. В современных условиях, характеризующихся значительным возрастанием нагрузки на оперативно-следственных работников, усилением противодействия расследованию со стороны виновных и других лиц, уходом из следственного аппарата профессионалов, одним из направлений решения данной задачи является применение в процессе расследования специализированных систем обработки информации для поддержки принятия решений.

Основная функция таких систем состоит в предоставлении помощи следователю в выборе рациональных решений при определении приоритетных направлений расследования и необходимых следственных действий.

Актуальность подобных разработок особенно возрастает в связи с появлением большого числа преступлений, расследование которых отличается значительной ресурсоемкостью.

Анализ структуры существующих консультационных и экспертных систем, применяемых в рассматриваемой предметной области, свидетельствует о том, что их информационным ядром является база данных признаков, описывающих исходную следственную ситуацию. Как правило, это – признаки объекта, способа совершения, сокрытия и орудия преступления, обстановки на месте происшествия, следовая картина, а также признаки субъекта и мотив. Такая система признаков является основой для выдвижения версий по делу. Именно выдвижение версий, выбор наиболее правдоподобной из них в конкретной следственной ситуации является, в свою очередь, основой процесса расследования преступлений.

Анализ существующих разработок, в частности, информационно-справочной системы «Версия», созданной в Московском НИИ проблем за-

конности и правопорядка в 1993 г., и экспертной системы «Маньяк», созданной в российском Учебно-производственном центре «ЮрИнфор» в 1991 г., а также публикаций [1,2] показывает, что в настоящее время доминирующим при оценивании выдвигаемых версий является вероятностный подход. Так, информационно-справочная система «Версия» базируется на методике выдвижения типовых версий при расследовании убийств, авторами которой являются Н.А. Селиванов и Л.Г. Видонов. Указанная методика нашла отражение в справочно-статистических таблицах „Типовые версии по делам об убийствах”, где в относительных показателях выражена корреляция между элементами криминалистической характеристики преступлений. В качестве исходных данных в системе учтены 5 признаков преступлений, каждый из которых может принимать 5...10 значений. Результатом работы системы является выдвижение типовой версии о личности преступника.

Экспертная система «Маньяк» предназначена для выдвижения версий о личности преступника при расследовании преступлений по сексуальным мотивам. В качестве исходных данных используются более 30 признаков преступлений, каждый из которых может принимать от 3 до 5 значений. В результате работы система выдает перечень версий о личности преступника, проранжированных по величине их вероятности.

К недостаткам указанных систем можно отнести невозможность работы в условиях неполноты информации (когда какие-либо из признаков преступления неизвестны), невозможность выдвижения и оценки правдоподобности версий не только о личности преступника, но и других неизвестных признаках, отсутствие возможности использования альтернативной стратегии для выдвижения версий, отказы системы («Версия») при определенных сочетаниях значений признаков. Кроме того, отсутствуют достоверные сведения о том, каковы основания использования в системах заданного количества и состава исходных признаков, а также об адекватности используемой в информационно-справочной системе «Версия» методики при анализе разнотипных данных.

Целью статьи является постановка задачи оценивания версий при расследовании преступлений как задачи распознавания образов.

Традиционно задача распознавания образов состоит в отнесении предъявляемого объекта к одному из классов.

Пусть задано пространство признаков определенной разновидности преступлений X размерностью n , векторами которого являются совокупности признаков отдельных преступлений $\tilde{x} = (x_1, x_2, \dots, x_n)$, где x_i – значения признаков. Формально следователя интересуют признаки субъекта и мотив, соответствующие имеющейся совокупности остальных признаков преступления [2]. Для определенности назовем признаки субъекта и мотива обоб-

щенной версией. Сделаем также замечание о том, что в процессе расследования версия может быть выдвинута о неизвестном значении любого признака преступления. Тогда пусть также задано пространство обобщенных версий $H = (h_1, h_2, \dots, h_k)$, точками которого являются версии, по которым раскрывалась данная разновидность преступлений. Каждому $\tilde{x}$ поставим в соответствие класс-версию h_j , по которой данное преступление было раскрыто. Тогда необходимо определить формальное решающее правило, позволяющее для произвольного $\tilde{x}$ из X указать класс-версию h_j , к которой он принадлежит (фактически – по которой должно расследоваться данное преступление).

При выборе способа построения решающего правила следует учесть следующие обстоятельства: 1) вероятностный характер признаков преступления; 2) преимущество вероятностных рассуждений в достижении равнорациональных решений даже в условиях неполноты информации, что существенно для данной предметной области; 3) возможность возникновения в процессе расследования необходимости выдвижения и оценивания версии о неизвестном значении любого признака преступления, учтенного в системе. С этой точки зрения значение и возможности метода Байеса для построения решающего правила трудно переоценить.

Использование процедуры Байеса базируется на предположении о том, что известны априорные вероятности $P(h_j)$ для классов и априорные условные плотности распределения значений вектора признаков $P(\tilde{x} | h_j)$. В рассматриваемой задаче $P(h_j)$ могут быть вычислены на основании эмпирических данных исследования репрезентативного массива раскрытых уголовных дел определенной разновидности, а именно, как отношение количества уголовных дел, раскрытых по данной версии h_j к общему количеству исследуемых дел. Эмпирическим путем могут быть вычислены и априорные условные вероятности $P(x_i | h_j)$ появления каждого конкретного значения каждого признака $x_i = x_i'$ при определенном значении версии h_j : как отношение количества наблюдений данного значения x_i при расследовании по версии h_j , к общему количеству наблюдений данного значения в исследуемой выборке.

Таким образом, имеем установленное соответствие между каждым значением версии и параметрами признаков преступления. Указанное соответствие может быть выражено в виде матрицы условных вероятностей,

строки которой соответствуют заданным признакам, а столбцы – версиям, по которым были раскрыты уголовные дела.

Если сделать допущение о статистической независимости признаков преступления между собой, то для распознавания нового объекта $\tilde{x}' = (x'_1, x'_2, \dots, x'_n)$, подаваемого на вход системы распознавания, процедура Байеса будет выглядеть следующим образом [3]:

$$P(h_j | x'_1, x'_2, \dots, x'_n) = \frac{P(x_1 = x'_1 | h_j) \cdot P(x_2 = x'_2 | h_j) \cdot \dots \cdot P(x_n = x'_n | h_j)}{\sum_{j=1}^k P(h_j) \cdot P(x_1 = x'_1 | h_j) \cdot P(x_2 = x'_2 | h_j) \cdot \dots \cdot P(x_n = x'_n | h_j)} \quad (1)$$

Результатом распознавания является распределение апостериорной вероятности $P(h_j | x'_1, x'_2, \dots, x'_n)$, которая показывает, какова вероятность данной версии h_j при имеющемся векторе признаков $\tilde{x}$. Решение принимается в пользу той гипотезы, апостериорная вероятность которой окажется наибольшей.

Известно, что качество решения, полученного в результате статистической процедуры, зависит от количества и свойств исходных данных. В частности, для корректного выполнения процедуры Байеса должно соблюдаться условие независимости признаков. Кроме того, остается проблема размерности пространства признаков, связанная как с большим количеством групп признаков, так и самих признаков, и их градаций. Так, уголовные дела исследуются по специальным анкетам, в которых содержится до 600 вопросов. И хотя известно, что первоначальный состав признаков распознавания должен быть как можно более полным и разносторонним, так как с увеличением количества признаков вероятность их распознавания увеличивается, однако, например, в работе А.С. Вознесенского «Информационные критерии качества распознавания состояния объектов и выбор параметров для его осуществления» показано, что необоснованное использование признаков в системе может привести к ухудшению качества распознавания.

Один из возможных путей решения этих проблем состоит в следующем.

Предварительно перед выполнением процедуры распознавания на обучающей выборке проводится процедура кластеризации, которая разбивает все множество из n признаков на m ($m < n$) групп. Признаки группируются по принципу максимальной близости (зависимости) между собой. Тогда между отдельными группами обеспечивается максимальное расстояние (независимость). Выбрав по одному признаку из каждой группы, получим m -мерную подсистему, которая наилучшим образом соответствует решающему правилу, ориентированному на использование независимых признаков [4], что и требуется в рассматриваемой задаче. Одновременно достига-

ется существенное сокращение количества признаков в системе. Более того, в [4] показано, что в рамках данного алгоритма кластеризации существует прием, позволяющий сформировать наиболее информативную n -мерную подсистему признаков.

Тем не менее, подзадача кластеризации требует отдельного рассмотрения, так как связана не только с выбором алгоритма, который позволил бы сохранить юридический смысл полученных результатов, их интерпретируемость, но, в большей мере, с выбором меры близости между разнотипными признаками – наиболее важного фактора, влияющего на результат кластеризации.

Таким образом, поставлена задача оценивания версий, выдвигаемых в процессе расследования преступлений, как задача распознавания образов; предложено использовать метод распознавания, основанный на вероятностном подходе, а именно, на одной из интерпретаций процедуры Байеса; выделена подзадача обеспечения корректного использования процедуры Байеса для построения решающего правила и снижения размерности пространства признаков распознавания и указан один из возможных путей ее решения, который одновременно позволяет сформировать наиболее информативную подсистему признаков.

Как представляется, при разработке автоматизированных систем поддержки принятия решений в таких предметных областях как уголовный процесс и криминалистика должна предусматриваться возможность реализации в рамках одной системы альтернативных стратегий принятия решений и процедуры сравнения их эффективности с целью принятия наиболее обоснованного решения при оценивании версий. В связи с этим дальнейшие исследования будут связаны с поиском иных методов распознавания, основанных принципах эмпирической индукции – анализе и обобщении прецедентов. В частности, предполагается исследовать возможность применения логических методов.

Библиографический список

1. Надеев А.Т. Принятие статистических решений в процессе расследования уголовных дел / А.Т. Надеев, О.С. Данилова // Тр. 5-й науч. конф. по радиофизике, Нижний Новгород, 7 мая 2001 г. — Нижний Новгород, 2001. — С. 355–356.
2. Надеев А.Т. Формализация представления априорной информации в обучающейся экспертной криминалистической системе / А.Т. Надеев, О.С. Данилова, С.В. Кошелев // Тр. 6-й науч. конф. по радиофизике, Нижний Новгород, 7 мая 2002 г. — Нижний Новгород, 2002. — С. 316–317.
3. Щавелев Л.В. Способы аналитической обработки данных для поддержки принятия решений / Л.В. Щавелев. — <http://www.osp.ru/dbms/1998/04-05/03.htm>.

4. Загоруйко Н.Г. Прикладные методы анализа данных и знаний / Н.Г. Загоруйко. — Новосибирск: Изд-во ИМ СО РАН, 1999. — 270 с.

Поступила в редакцию 23.09.2005 г.