С. Д. Кулик, Д. А. Никонец

ЗАДАЧА ИДЕНТИФИКАЦИИ ИНФОРМАЦИОННОГО ОБЪЕКТА В СЛУЧАЕ БОЛЬШОГО КОЛИЧЕСТВА ИДЕНТИФИЦИРУЮЩИХ ПРИЗНАКОВ

При разработке систем для решения задач информационной безопасности возникает необходимость идентификации информационного объекта или принятия решения о тождестве объектов, например идентификации пользователя системы. Задачу идентификации можно решать различными способами. Однако необходимо заметить, что в случае идентификации информационного объекта с большим количеством идентифицирующих признаков могут потребоваться значительные вычислительные и временные затраты. Алгоритм идентификации и его модификация, описанная в данной работе, позволяют частично решить данную проблему.

В настоящее время алгоритм используется при разработке АРМ эксперта-криминалиста «FHWE v. 1.0» [1]. Один из режимов работы АРМ представляет собой автоматизированный вариант методики идентификации исполнителя рукописного текста по прописным буквам.

Подробная математическая постановка задачи дана в [1]. В основе алгоритма лежит получение с заданным уровнем надежности интервальной оценки вероятности появления комплекса признаков и последующее сравнение оценки с заданным порогом. Также предполагается статистическая независимость появления признаков информационного объекта, что не всегда верно на практике.

Порядок использования алгоритма для принятия решения о тождестве двух объектов следующий:

1) выделение признаков, совпадающих в представленных информационных объектах;

2) определение с заданным уровнем надежности интервальной оценки вероятности появления комплекса признаков;

3) принятие решения о тождестве информационных объектов в случае, когда интервальная оценка меньше или равна выбранному пороговому значению.

Также возможна реализация алгоритма в виде специального устройства определения фальшивых рукописных документов на русском языке [2, 3].

Для экспериментальной проверки алгоритма была использована выборка из 691 рукописного образца, выполненных разными людьми. В качестве идентифицирующих признаков были использованы около 2000 признаков букв. Были сформированы пары исполнителей для проверки. Общее число пар равно $N = C_{691}^{2} = 238395$. Ошибочный вывод о том, что документы выполнены одним лицом, был сделан для 14,1 % от общего числа пар.

Было сделано предположение, что достаточно большой уровень ошибок связан с тем, что существует ряд наборов признаков, которые статистически зависимы. Для уменьшения количества ошибок были использованы алгоритмы поиска ассоциативных правил [4], что позволило выявить наборы зависимых признаков. Для поиска ассоциативных правил была использована модификация AprioriTid хорошо известного алгоритма Apriori. Таким образом, были обнаружены более 100 наборов зависимых признаков, каждый из которых содержал от 2 до 5 признаков.

Модификация алгоритма идентификации информационного объекта заключается в замене всех зависимых признаков, входящих в такой набор, на один новый логический признак. Т. е. сколько бы ни встретилось в исследуемом образце зависимых признаков из одного набора, считается, что встретился только этот один логический признак.

Ошибочный вывод о том, что документы выполнены одним лицом, при использовании модифицированного алгоритма был сделан для 0,54 % пар.

Следующей планируемой модификацией алгоритма идентификации будет поиск наборов зависимых признаков не для всех возможных признаков, а только для тех, которые встречаются в исследуемых информационных объектах, что уменьшит количество возможных наборов и, соответственно, снизит временные затраты. Все это в итоге позволит обеспечить заданный уровень информационной безопасности.

Выводы

Использование предложенного алгоритма позволяет с заданным уровнем доверия решить задачу идентификации объекта в случае большого количества признаков с относительно небольшими вычислительными затратами.

Модификация алгоритма идентификации позволяет снизить уровень ошибок и повысить качество принимаемых решений, в том числе и для обеспечения информационной безопасности.

СПИСОК ЛИТЕРАТУРЫ:

1. *Кулик С. Д., Никонец Д. А.* Примеры использования нейросетевого алгоритма в методиках для эксперта-почерковеда // Нейрокомпьютеры: разработка и применение. 2009. № 9. С. 61–85.
2. *Кулик С. Д., Никонец Д. А., Ткаченко К. И., Жижилев А. В.* Устройство определения поддельных документов // Безопасность информационных технологий. 2009. № 1. С. 114–115.
3. *Кулик С. Д., Никонец Д. А., Ткаченко К. И., Жижилев А. В.* Патент на полезную модель № 73750, Российская Федерация (RU), л. МПК7 G 07 D 7/00. Устройство определения фальшивых рукописных документов на русском языке / С. Д. Кулик, Д. А. Никонец, К. И. Ткаченко, А. В. Жижилев (Россия). Заявка № 2007147832/22; Заяв. 25.12.2007; Зарегистр. 27.05.2008; Приоритет от 25.12.2007. Опубл. Бюл. № 15. Ч. 3. С. 860. (РОСПАТЕНТ).
4. *Webb G. I.* Discovering Significant Patterns // Machine Learning. Vol. 68 (1). Netherlands: Springer, 2007. P. 1–33.

С. Д. Кулик, К. И. Ткаченко

РАЗРАБОТКА ГЕНЕРАТОРОВ ДЛЯ ОБЕСПЕЧЕНИЯ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

На практике при обучении экспертов-криминалистов часто возникает проблема недостатка образцов документов с искаженной информацией из-за малой наполненности учебного фонда необходимыми документами. В рамках решения данной проблемы было предложено на основе имеющихся образцов и правил составления подлинных документов различных типов генерировать образцы искаженных (фальшивых) документов [1–8]. Были разработаны и внедрены генераторы искажения информации для фальшивых векселей и фальшивых инструкций к лекарственным средствам [2–5, 8]. Обобщенная схема функционирования данных генераторов искажения информации [3–6] представлена на рис. 1.

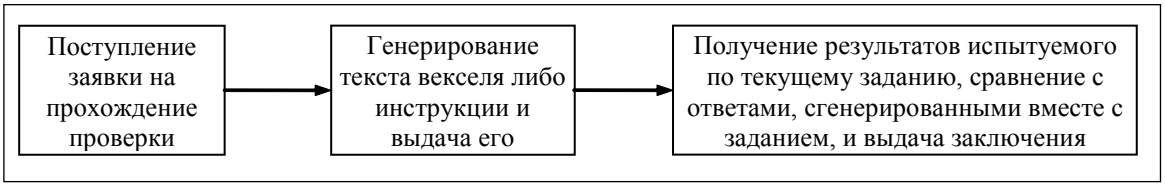

Рис. 1. Обобщенная блок-схема функционирования разработанных генераторов