

НЕЙРОСЕТЕВОЙ ПОДХОД К РЕШЕНИЮ ЗАДАЧИ АНАЛИЗА ВРЕМЕННЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Нестерук Ф.Г., Нестерук Г.Ф., Нестерук Л.Г.

Введение

В научных публикациях много внимания уделяется применению интеллектуальных средств в системах обеспечения информационной безопасности и обороноспособности [1, 2]. Исследуется возможность использования, как отдельных интеллектуальных средств [3 - 5], так и их комбинаций [2, 6] для обеспечения безопасности в условиях высокой динамики угроз [1, 7]. Известные интеллектуальные средства основаны на принципе подобия реализуемых функций аналогичным функциям биосистем [1, 7]. К часто используемым интеллектуальным средствам относят базы знаний в составе экспертных систем [6, 8], в том числе на основе байесовского метода [6, 9], нечеткой логики [10, 11], а также нейронные сети (НС) [12, 13], эволюционные методы [14] и гибридные системы [7, 15, 16].

Задача исследования

Интеллектуальные средства обладают, с одной стороны, сочетанием свойств, необходимых для решения задач, которым свойственны недостаточная полнота и не высокая достоверность исходных данных, с другой, эволюционными качествами обобщения, самоорганизации, адаптации, наследования и передачи накопленного опыта.

В ряде приложений стоит задача интеллектуального анализа данных, упорядоченных во времени, в частности разделение сложного потока цифровых данных по источнику их возникновения. В этой связи актуальна задача применения интеллектуальных средств для решения задачи классификации источников, информацию о которых необходимо выделить из сложного потока цифровых данных.

Анализ подходов к разделению сложного потока цифровых данных

Признаки, по которым разделяют сложный поток цифровых данных, можно разделить на две группы: во-первых, внутренняя специфика цифровых данных, во-вторых, наличие временной составляющей (периодичности) поступления отдельных цифровых данных в общий информационный поток. Рассмотрим подходы разделения цифровых данных по периодичности поступления отдельных слов данных в общий информационный поток.

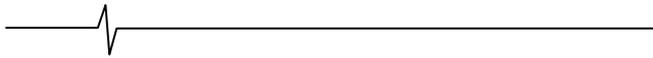
В случае, когда составляющие сложного информационного потока имеют постоянные периоды следования, достаточно применить аппаратно-программный способ временного разделения смешанной последовательности сигналов [17]. Способ состоит в том, что формируют смешанную дво-

Существует класс проблем, связанных с распознаванием образов, оперативным анализом временных последовательностей, которые характеризуются наличием неполных и зашумленных данных. Подобные проблемы свойственны таким сферам, как обеспечение безопасности стратегически важных систем. Для их решения привлекают интеллектуальные средства экспертных систем, нечеткой логики, нейронных сетей (НС), которым свойственны эволюционные качества самоорганизации, адаптации, возможности наследования накопленного опыта. В работе рассматривается подход, связанный с применением интеллектуальных средств для решения задачи классификации объектов по последовательности импульсных сигналов.

ичную последовательность, каждый бит которой соответствует минимальному отсчету системного времени и кодирует 1 или 0, соответственно, наличие или отсутствие импульса в последовательности сигналов. В подходе игнорируется смысловая составляющая компонентов информационного потока, а в упорядоченной последовательности битов отмечают единицами моменты времени включения каждого отдельного компонента в общую информационную последовательность. Разделение потока цифровых данных производится многоходовой схемой совпадения, которая объединяет единичные биты, выделенные из двоично-кодированной последовательности посредством линии задержки с рядом отводов. Временная задержка на каждой из групп отводов кратна целому числу периодов следования представленных импульсами сигналов [17].

Недостатки рассмотренного способа – ограниченное множество типов сигналов с постоянным периодом следования импульсов, что редко встречается в практических задачах, и чувствительность к пропуску отдельных компонентов потока, так как способ основан на конъюнктивном объединении трех или более битовых позиций анализируемого потока.

Большую практическую значимость имеет подход к разделению информационного потока, заключающийся в том, что формируют ряд двоично-кодированных последовательностей, а именно: смешанную последовательность (по общему информационному потоку) и конечное множество эталонных последовательностей. Для разделения потока цифровых данных осуществляют последовательное сравнение смешанной последовательности с каждой из эталонных последовательностей, исходя из результатов которого формулируют заключение о вхождении эталонных последовательностей в состав смешанной последовательности [18]. В способе использована идея объединения процессов распознавания и разделения импульсов на основе текущих статистических оценок, а именно: кросс-корреляции каждой эталонной последовательности со смешанным потоком цифро-



вых данных и автокорреляции импульсов в пределах каждой эталонной последовательности. Идентификация источников производится по максимальным значениям корреляции.

Для каждого источника сигналов создается библиотека бинарных последовательностей импульсов (bps - binary pulse sequence), кодирующих единицей отсчет в случае наличия импульса и нулем - в противном случае. Аналогично кодируется смешанный импульсный поток, формируя бинарную последовательность потока (bss - binary stream sequence). Рис. 1 иллюстрирует варианты кодирования для случаев а) bss и б) bps [18].

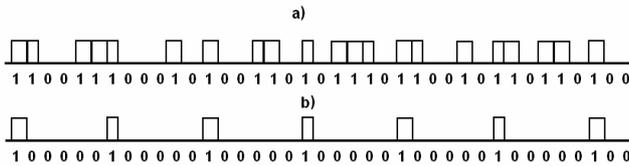


Рис. 1. Иллюстрация а) смешанного и б) эталонного бинарных потоков

Если число отсчетов в интервале наблюдения - N , то кросс-корреляция между bps n -го источника и принятой bss определена как

$$cross - correlation_n(i) = \sum_{k=0}^{N-1-i} bps_n(k)bss(k+i), \quad 0 \leq i \leq N-1$$

Автокорреляция bps n -го источника определена как

$$auto - correlation_n(i) = \sum_{k=0}^{N-1-i} bps_n(k)bps_n(k+i), \quad 0 \leq i \leq N-1$$

Если максимальное значение кросс-корреляции соответствует $i = h$, то отношение корреляций n -го источника, $R(n)$ определено как

$$R(n) = \frac{cross - correlation_n(h)}{auto - correlation_n(h)}$$

Данное отношение представляет процент импульсов от n -го источника, присутствующих в смешанном импульсном потоке. $R(n)$ сравнивают с порогом $T = 1 - A$, где A - допустимый процент пропущенных импульсов для bps. Рассмотренный процесс повторяют до тех пор, пока в bss остается число единиц, не меньшее оговоренного заранее.

Взаимная корреляция между bss и bps каждого из M источников, также как автокорреляция bps того же источника может быть реализована с использованием конъюнкторов и сумматоров для подсчета числа единиц в выходных сигналах. Выходные сигналы обоих корреляторов - фактическое число импульсов от n -го источника и число импульсов от данного источника в сложном импульсном потоке в течение некоторого интервала времени [18].

В качестве недостатка данного способа разделения сигналов смешанной последовательности следует назвать невысокое быстродействие из-за последовательного характера сравнения двоичных последовательностей сигналов.

Известен способ разделения импульсов смешанной последовательности сигналов по отдельным дорожкам с использованием нейросетевых классификаторов, например, Fuzzy ARTMAP (FAM), согласно которому вначале производится оцифровка внутриимпульсных параметров сигнала, а

затем, исходя из значений цифровых кодов параметров (первая группа признаков), к отдельным дорожкам относят подмножества импульсов из смешанной последовательности сигналов [2].

FAM реализуют в виде упрощенной модели (рис. 2) [19]. FAM используется для классификации источника по функциональным импульсным параметрам. Он состоит из двух уровней узлов (нейронов) с полными связями: M узлов входного слоя $F1$ и N узлов соревновательного слоя $F2$. Набор весов

$$W = \{w_{ij} \in [0, 1], i = 1, 2, \dots, M, j = 1, 2, \dots, N\}$$

соответствует межнейронным связям между слоями $F1$ и $F2$.

Каждый j -й узел слоя $F2$ представляет категорию распознавания, которой соответствует вектор прототипа

$$W_j = (w_{1j}, w_{2j}, \dots, w_{Mj}).$$

Слой $F2$, связанный через обученные ассоциативные связи с $1, \dots, L$ узлами выходного слоя, образующие поле преобразования F^{ab} , где L - число классов в выходном пространстве классификатора. Набор двоичных весов

$$W^{ab} = \{w_{jk}^{ab} \in \{0, 1\}, j = 1, 2, \dots, N, k = 1, 2, \dots, L\}$$

соответствует межнейронным связям между слоями $F2$ и F^{ab} . Вектор

$$w_j^{ab} = (w_{j1}^{ab}, w_{j2}^{ab}, \dots, w_{jL}^{ab})$$

связывает j -й узел слоя $F2$ с L узлами выходного слоя классификатора.

В течение обучения FAM выполняет управляемое обучение согласно нормализованным векторам входного набора обучения

$$a = (a_1, a_2, \dots, a_m), \quad 0 \leq a_i \leq 1,$$

и выходным меткам (output labels)

$$t = (t_1, t_2, \dots, t_L),$$

где $t_k = 1$, если k - целевая метка класса для a , и нулю в противном случае.

Вектор $A = (a, 1 - a)$ является $2m$ -мерным кодовым дополнением m -мерного вектора a , где 1 является m -мерным единичным вектором.

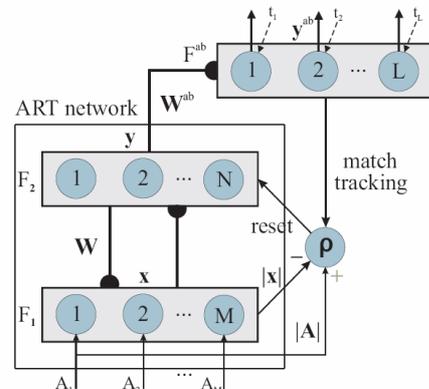
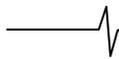


Рис. 2. Структура FAM

Недостатком способа разделения смешанной последовательности сигналов на базе FAM является использование на этапе разделения сигналов базы данных о перио-



дах следования импульсов, которая автоматически не выполняется в процессе работы системы.

Нейросетевой подход к разделению смешенного потока цифровых данных

Для повышения производительности процесса разделения сигналов смешанной импульсной последовательности следует использовать параллелизм обработки информации, свойственный НС. На рис. 3 представлена структурная схема нейросетевого разделителя сигналов смешанной последовательности.

На входы A_1, A_2, \dots, A_M подают двоичные последовательности разрядности M .

В процессе обучения подают конечное множество двоичных последовательностей разрядности M , на входы 1 НС, которая в процессе обучения формирует кластеры 2, взаимосвязанные с выходами 3 нейронной сети. Для каждого кластера формируют вектор весов входных межнейронных связей, координатам которого в процессе обучения НС по одному из известных алгоритмов присваивают значения, равные результатам выполнения операции «min» над двоичными последовательностями, соответствующими одному из разделяемых сигналов [19]. В процессе обучения НС для каждого кластера формируют код маски за счет выполнения операции «max» над двоичными последовательностями, соответствующими одному из разделяемых сигналов. Таким образом настраивают маску на выделение из смешанной последовательности участков, соответствующих данному сигналу.

В цикле эксплуатации подают на входы НС фрагмент смешанной двоичной последовательности разрядностью, равной разрядности эталонных двоичных последовательностей. При этом выполняют операцию «min» фрагмента с кодом маски, поступающим по обратной связи на вход блока 5. Результат обработки передается по входным межнейронным связям в каждый из N кластеров нейронной сети.

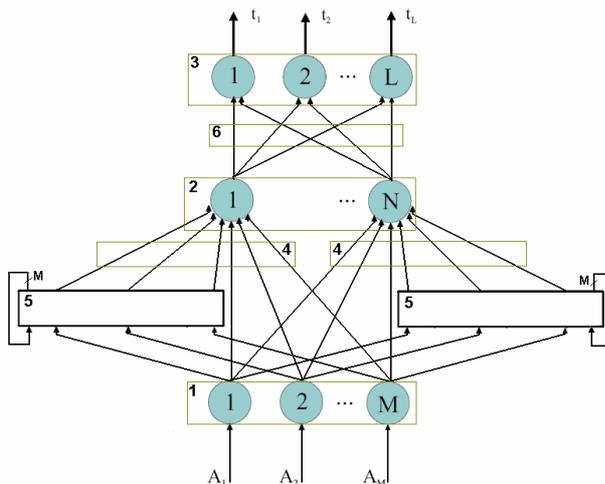


Рис. 3. Нейросетевой разделитель сигналов

В каждом из N кластеров проверяют соответствие результатов выполнения операции «min» (блок 5) и векторов весов входных межнейронных связей 4. Формируют результаты сравнения двоичных последовательностей в виде значений, которые накапливают в весах выходных межнейронных связей, например, аналогично, описанному в [20].

Выполняют операцию левого сдвига смешанной двоич-

ной последовательности. Операция сдвига позволяет выявить совпадение эталонной двоичной последовательности и фрагмента смешанной двоичной последовательности в случае, если их начальные моменты не совпадают во времени.

Описанные преобразования выполняют для всех фрагментов смешанной двоичной последовательности, включая конечный фрагмент смешанной двоичной последовательности. Формулируют заключение о наличии в составе смешанной импульсной последовательности каждой из эталонных импульсных последовательностей, исходя из результатов сравнения цифровых кодов весов выходных межнейронных связей с величиной порога. Результаты сравнения отражаются на выходах t_1, t_2, \dots, t_L нейронной сети.

Проиллюстрируем эффективность предложенного нейросетевого подхода к разделению смешенного потока цифровых данных.

Базовый алгоритм Fuzzy ARTMAP сравнивает входной вектор $a = (a_1, a_2, \dots, a_m)$, $0 \leq a_i \leq 1$ с каждой из категорий распознавания, которой соответствует вектор прототипа $W_j = (w_{1j}, w_{2j}, \dots, w_{mj})$.

1. В случае совпадения вышеназванных векторов (с заданной точностью) происходит отнесение входного вектора к классу данной категории и инкрементное обучение (коррекция) вектора прототипа $W_j = (w_{1j}, w_{2j}, \dots, w_{mj})$ для данной категории.

2. В противном случае определяется категория, максимально близкая к входному вектору, и исследуется возможность расширения ее границ (гиперпрямоугольная область в пространстве решений) для охвата точки пространства, соответствующей входному вектору. Если попытка неудачна, то она повторяется для ряда близ расположенных гиперпрямоугольных областей пространства решений других категорий.

3. В случае отрицательного результата по п. 2 создается новая категория с координатами вектора прототипа $W_j = (w_{1j}, w_{2j}, \dots, w_{mj})$, равными координатам входного вектора $a = (a_1, a_2, \dots, a_m)$.

Так как во входных цепях Fuzzy ARTMAP отсутствует операция предварительного маскирования над двоичной последовательностью смешанных импульсных сигналов, то входные векторы $a = (a_1, a_2, \dots, a_m)$ наряду с разрядами двоичных последовательностей, соответствующими разделяемым сигналам, содержат также разряды других последовательностей, наложившиеся на классифицируемую последовательность. В результате исход в виде отнесения входного вектора к классу одной из категорий (п.1) маловероятен, и алгоритм Fuzzy ARTMAP будет выполнять затратные по времени шаги в соответствии с п.п. 2 и 3.

В описанном варианте нейросетевого разделителя смешанных двоичных последовательностей блоки 5, выполняя операцию «min» над фрагментом двоичной последовательности и кодом маски, поступающим по обратной связи с выхода того же блока, обеспечивают поступление на вход нейросетевого классификатора только тех разрядов смешанной двоичной последовательности, которые совпали с кодами масок в блоках 5. В результате вероятность исхода в виде отнесения входного вектора к классу одной из категорий (п.1) близка к 1, и алгоритм Fuzzy ARTMAP не будет выполнять затратных по времени шагов в соответствии с п.п. 2 и 3.

Таблица 1

Поток, имп/с.	Ср. время 1 имп.	Время 100 имп.	Число каналов	Число классификаций	Ср. время одной классификации, с	Ср. время одной классификации, нс
300000	3,3333E-06	0,000333333	10	1000	3,33333E-07	333,3333

Таблица 2

Число категорий	22		11	
	1	2	1	2
Вариант				
Найдено совпадений	73	73	20	20
Среднее время классификации, мс	0,0059365	0,0059365	0,0041206	0,0041905
Минимальное время, мс	0,0053079	0,0050286	0,0032825	0,0032825
Максимальное время, мс	0,0120825	0,0119429	0,0150857	0,0141778

Проведено моделирование на ПЭВМ с двухядерным процессором Intel, тактовой частотой 2,4 ГГц и ОЗУ 2Гб для потока импульсных данных в соответствии с таблицей 1.

В таблице 2 содержатся результаты моделирования нейросетевого разделителя смешанных двоичных последовательностей для двух случаев: с числом категорий классификатора, равным, соответственно, 22 и 11.

Минимальное время классификации характерно для случаев совпадения входного вектора $a = (a_1, a_2, \dots, a_m)$ с вектором одного из прототипов $W_j = (w_{1j}, w_{2j}, \dots, w_{mj})$ в начале выполнения п.1, а максимальное – для совпадения входного вектора с вектором одного из прототипов в конце выполнения п.1. Среднее время классификации рассчитывалось по всем случаям классификации (73 для 22 категорий и 20 для 11 категорий).

Если учесть, что для удачной классификации в случае Fuzzy ARTMAP потребуется время близкое к максимальному времени из таблицы 2 (маловероятный исход правильной классификации потребует опроса практически всех категорий), то описанный в статье нейросетевой разделитель смешанных двоичных последовательностей будет оперировать согласно графе «Среднее время классификации». Таким образом, выигрыш в быстродействии составит от 2 до 3,3 раз при прочих равных условиях.

Заключение

Рассмотренный способ позволяет реализовать разделение сигналов смешанной импульсной последовательности в реальном масштабе времени, а также повысить точность решения задачи за счет исключения потери ряда импульсов смешанной последовательности по причине недостаточной производительности средств разделения сигналов.

Литература

1. Нестерук Г. Ф., Осовецкий Л. Г., Харченко А. Ф. Информационная безопасность и интеллектуальные средства защиты информационных ресурсов. - СПб.: СПбГУЭФ. 2003.
2. Granger E., Rubin M. A., Grossberg S., Lavoie P. A what-and-where fusion neural network for recognition and tracking of multiple radar emitters, *Neural Networks*, vol. 3, 2001. P. 325 - 344.
3. Porras P. A., Ilgun K., and Kemmerer R. A. State transition analysis: A rule-based intrusion detection approach. // *IEEE Trans. on Software Engineering*, 1995. SE-21. P. 181-199.
4. Логинов В. А. Методика активного аудита действий субъектов доступа в корпоративных вычислительных сетях на основе апарата нечетких множеств // *SCM'2003*. – СПб.: СПГЭТУ, 2003. т. 1. С. 240 - 243.
5. Ryan J., Lin M., Miikkulainen R. Intrusion Detection with Neural Networks. *AI Approaches to Fraud Detection and Risk Management: Papers from the 1997 AAAI Workshop* (Providence, Rhode Island), pp. 72-79. Menlo Park, CA: AAAI. 1997.
6. Negnevitsky M. *Artificial intelligence: a guide to intelligent systems* // Addison-Wesley. 2002.
7. Нестерук Г. Ф., Осовецкий Л. Г., Нестерук Ф. Г. О применении нейро-нечетких сетей в адаптивных системах информационной защиты // *Нейроинформатика-2005: Материалы VII всероссийской научно-технической конференции*. М.: МИФИ (ТУ), 2005. Ч.1. С. 163 - 171.
8. Waterman D.A. *A Guide to Expert Systems*. - MA: Addison-Wesley, Reading. 1986.
9. Прокопчина С. В. Байесовские интеллектуальные технологии для аудита и управления сложными объектами в условиях значительной неопределенности // *SCM'2002*. - СПб, 2002. т.1, С. 27 - 31.
10. Zadeh L.A., Kacprzyk, J. *Fuzzy Logic for the Management of Uncertainty*. – NY: John Wiley. 1992.
11. Li H., Gupta M. *Fuzzy Logic and Intelligent Systems*. – Boston: Kluwer Academic Publishers. 1995.
12. Fu L.M. *Neural Networks in Computer Intelligence*. - McGraw-Hill Book, Inc. 1994.
13. Круглов В. В., Борисов В. В. *Искусственные нейронные сети. Теория и практика*. - 2-е изд. – М.: Горячая линия - Телеком, 2002.
14. Davis L. *Handbook on Genetic Algorithms*. – NY: Van Nostrand Reinhold. 1991.
15. Fuller R. *Neural Fuzzy Systems*. - Abo: Abo Akademi University, 1995.
16. Рутковская Д., Пилиньский М., Рутковский Л. *Нейронные сети, генетические алгоритмы и нечеткие системы*. - М.: Горячая линия –Телеком, 2004.
17. Мельников Ю. П. *Воздушная радиотехническая разведка (методы оценки эффективности)*. – М.: Радиотехника, 2005.
18. Hassan H. E. A new algorithm for radar emitter recognition. *Proceedings of the 3rd International Symposium on Image and Signal Processing and Analysis*. 2003. P. 1097 – 1101.
19. Carpenter G. A., Grossberg S., Markuzon N., Reynolds J. H. and Rosen D. B. "Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps," *IEEE Transactions on Neural Networks*, vol. 3, no. 5, pp. 698-713, 1992.
20. Carpenter G.A., Markuzon N. ARTMAP-IC and medical diagnosis: Instance counting and inconsistent cases. // *Neural Networks*, 11, 1998. P. 323 – 336.