

**МЕТОДОЛОГИЯ ПОСТРОЕНИЯ АДАПТИВНЫХ
АЛГОРИТМОВ ОБРАБОТКИ ИНФОРМАЦИИ В СРЕДСТВАХ
ОБНАРУЖЕНИЯ ДВИЖУЩИХСЯ НАЗЕМНЫХ ОБЪЕКТОВ
(Ч. 2. МЕТОДЫ ОБРАБОТКИ ИНФОРМАЦИИ
ДЛЯ ПОСТРОЕНИЯ АДАПТИВНЫХ
СРЕДСТВ ОБНАРУЖЕНИЯ)**

Аннотация.

Актуальность и цели. Объектом исследования являются адаптивные средства обнаружения движущихся наземных объектов. Предметом исследования являются методы построения адаптивных алгоритмов обработки информации. Цель – разработка методов, составляющих основу методологии построения адаптивных средств обнаружения движущихся наземных объектов.

Материалы и методы. Исследования выполнены с использованием методов прогнозирования временных рядов и распознавания образов.

Результаты. Предложены методы обработки информации, применяемые для создания адаптивных средств обнаружения. На экспериментальных данных показана их эффективность.

Выводы. Предложенные методы обработки информации в составе методологии построения адаптивных средств обнаружения движущихся наземных объектов могут быть использованы при разработке новых и модернизации существующих средств обнаружения для повышения их эффективности при работе в условиях воздействия изменяющихся внешних факторов.

Ключевые слова: средства обнаружения, решающие правила, предсказание.

М. А. Mitrokhin

**METHODOLOGY OF BUILDING ADAPTIVE
ALGORITHMS FOR DATA PROCESSING
BY MOVING GROUND OBJECT SENSORY DEVICES
(PART 2. Data processing methods for adaptive sensory devices)**

Abstract.

Background. The research object is adaptive devices for sensing moving ground objects. The research subject is the methods of building adaptive data processing algorithms. The aim of the work is to develop the methods being the basis of the moving ground object adaptive sensory devices building methodology.

Materials and methods. The research was carried out using the method of time series prediction and image recognition.

Results. The author has suggested data processing methods to be used in creation of adaptive sensory devices. The efficiency thereof has been proved by experimental data.

Conclusions. The suggested methods of data processing in the structure of the moving ground object adaptive sensory devices building methodology can be used in development of the new and modernization of the existing sensory devices in order to improve efficiency thereof when operating under the influence of changing external factors.

Key words: sensory device, decision rule, prediction

Введение

В первой части статьи проведен анализ влияния внешних факторов на вариацию информативных параметров и показатель эффективности решения задачи классификации. Сформулированы цель и основные задачи методологии построения адаптивных средств обнаружения (СО), предложены подходы к построению адаптивных СО, функционирующих в изменяющихся условиях и структурные схемы для их реализации.

Во второй части статьи приводятся основные результаты по разработке методов обработки информации, реализующих предложенные подходы к построению адаптивных СО.

1. Процедура мониторинга

Важным элементом двух схем построения адаптивного СО, описанных в первой части статьи, является элемент, реализующий процедуру мониторинга. Этот элемент решает задачи определения изменений, происходящих в признаковых описаниях классов со временем, определение количественных критериев величины изменений и определение необходимости начала процесса адаптации.

Обрабатываемые в СО признаки можно рассматривать как наборы случайных величин. Поэтому наибольшее количество информации о признаках содержится в их законах распределения, а наиболее полное описание происходящих изменений может быть получено из анализа эволюции плотностей распределения признаков во времени. Так, например, на рис. 1 показано изменение выборочной плотности распределения логарифма амплитуды огибающей сигнала во времени, вычисленной по недельной записи сейсмического фона при воздействии ветра и осадков.

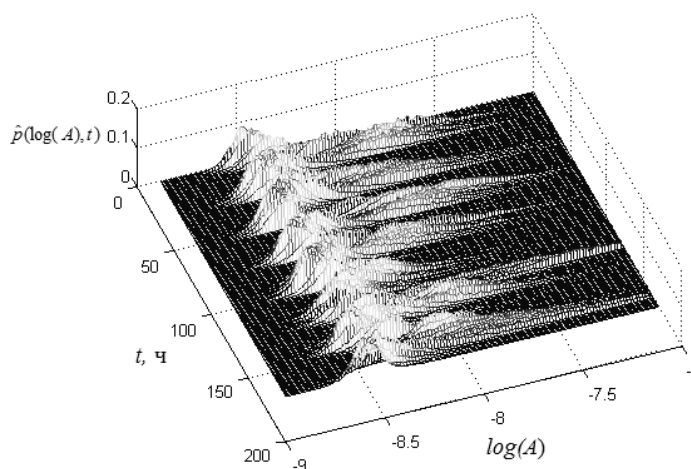


Рис. 1. Изменение выборочной плотности распределения логарифма амплитуды огибающей сигнала во времени

Плотность распределения на рис. 1 изменяется в достаточно широких пределах, особенно при воздействии ветра и осадков, кроме того, прослеживается периодичность изменений, связанная с суточной антропогенной активностью, оказывающей влияние на амплитуду сейсмического фона.

Для выявления расхождения между плотностями распределений двух последовательных выборок отсчетов признака предлагается использовать расстояние Кульбака – Лейблера [1]:

$$KL(p_1, p_2) = \sum_{x \in X} p_1(x) \ln \frac{p_1(x)}{p_2(x)},$$

где $p_1(x)$ и $p_2(x)$ – плотности вероятности отсчетов анализируемого признака, взятые для двух последовательных выборок.

Данный функционал можно использовать как меру удаленности друг от друга двух вероятностных распределений. Использование расстояния Кульбака – Лейблера не накладывает ограничений на вид и способ задания плотностей распределения. Это означает, что плотности по конечным выборкам могут восстанавливаться в том числе и быстрыми непараметрическими методами.

На рис. 2 приведен график изменения расстояния Кульбака – Лейблера, вычисленного для последовательности выборочных плотностей распределения признака, показанной на рис. 1.

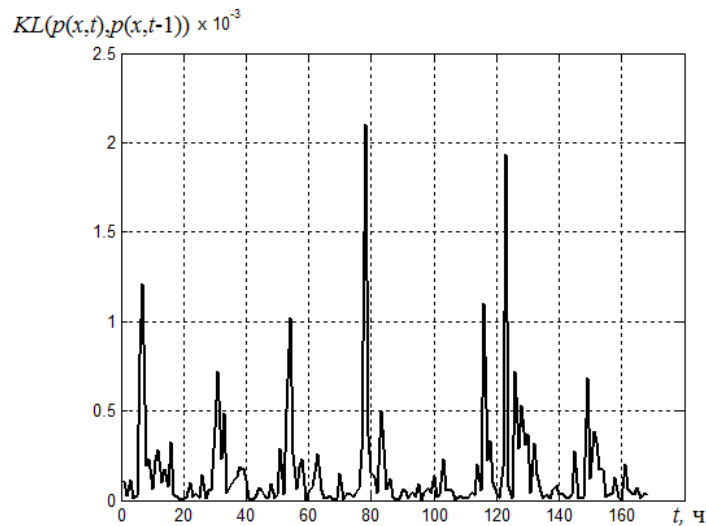


Рис. 2. Изменения расстояния Кульбака – Лейблера, вычисленного для последовательности выборочных плотностей распределения логарифма амплитуды огибающей сигнала

На рис. 2 близкие к нулю значения расстояния Кульбака – Лейблера указывают на отсутствие изменений в распределении признака, большие значения указывают на наличие изменений, вызванных одним из влияющих факторов (смена времени суток, воздействие ветра, дождя и т.д.).

Для определения наличия изменений необходимо задать пороговое значение этого критерия, при превышении которого изменения считаются существенными и необходимо начать процесс адаптации. Конкретное значение порога для расстояния Кульбака – Лейблера зависит от решаемой задачи, используемых признаков и воздействующих факторов и требует предварительного исследования априорной информации.

1. Метод адаптации СО на основе прогнозирования временных рядов

Как известно из байесовского подхода к распознаванию образов, вероятность ошибок классификации есть вероятность того, что признак превысит (или не превысит) соответствующий порог (рис. 3).

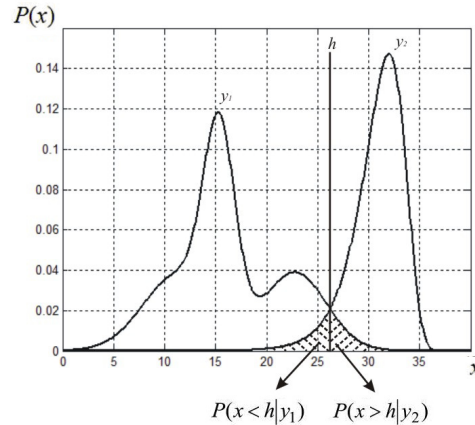


Рис. 3. Ошибки принятия решений при двухклассовой классификации с известными плотностями распределений классов

Оценку вероятности ошибок классификации можно получить, опираясь на неравенство Чебышева:

$$P\left(\left|x - \mu_x^{y_1}\right| \geq h \mid y_1\right) \leq \frac{\left(\sigma_x^{y_1}\right)^2}{h^2}, \quad P\left(\left|x - \mu_x^{y_2}\right| < h \mid y_2\right) \geq \frac{\left(\sigma_x^{y_2}\right)^2}{h^2},$$

где $\mu_x^{y_1}$ и $\mu_x^{y_2}$ – математические ожидания распределений признака x классов y_1 и y_2 ; $\sigma_x^{y_1}$ и $\sigma_x^{y_2}$ – среднеквадратичные отклонения распределений признака x классов y_1 и y_2 .

Из приведенных неравенств видно, что вероятность ошибки классификации зависит от дисперсий распределений признаков классов. В случае, когда значения x являются реализациями случайного процесса, зависящими от внешних факторов, каждый из них вносит определенный вклад в вариацию x . Имея модель зависимости признака x от параметров внешних факторов, можно оценить значение x в текущий момент времени с учетом этого влияния. Очевидно, что учет влияющих факторов при вычислении значений x уменьшает дисперсию σ_x^2 . Построение модели зависимости признака от значений параметров помеховых факторов может рассматриваться в рамках различных подходов: регрессионного анализа [2], сингулярного спектрального анализа [3, 4], локальной аппроксимации [5] и др.

Таким образом, разработав модели зависимостей признаков, используемых в СО, от изменяющихся условий и помеховых факторов, можно использовать в алгоритме распознавания прогнозные значения признаков, полученные по соответствующим моделям и использовать стандартные алгоритмы распознавания образов.

Так, при решении задачи обнаружения человека по сейсмическому сигналу, описанной в части 1 настоящей статьи, для прогнозирования значений признаков можно использовать метод сингулярного спектрального анализа, основанный на вычислении собственных векторов ковариационной матрицы многомерного представления исходного ряда значений признака. Собственные векторы, являющиеся основными составляющими ряда значений признака, используются для определения прогнозных значений. Модель прогнозирования строилась по сигналам класса «помехи». При прогнозировании признаков были использованы три собственных вектора, имеющих наибольшие собственные числа.

На рис. 4 приведены выборочные плотности распределения прогнозных значений логарифма амплитуды огибающей сигналов и количества экстремумов, вычисленные по сигналам объектов и помех.

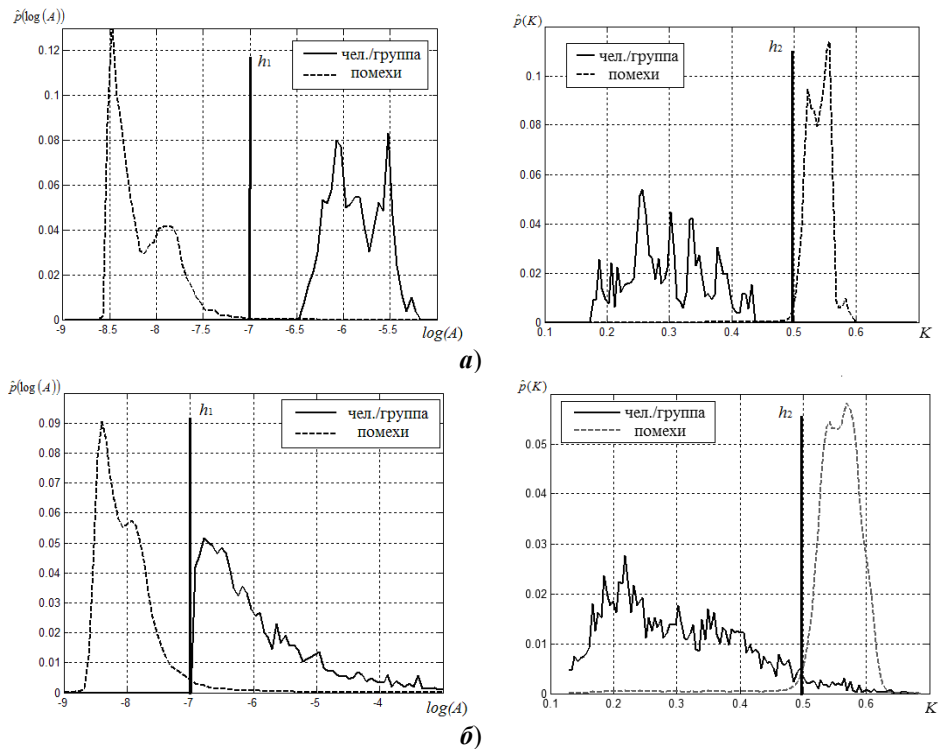


Рис. 4. Выборочные плотности распределения прогнозных (а) и исходных (б) значений логарифма амплитуды огибающей сигналов и количества экстремумов, вычисленные по сигналам объектов и помех

Как видно из рис. 4, плотности распределения прогнозных значений признаков имеют меньшую дисперсию по сравнению с распределениями исходных признаков, что приводит к значительному уменьшению ошибок классификации даже при использовании пороговых решающих правил.

При использовании Байесовского подхода к распознаванию для адаптации на основе прогнозирования может быть использовано прогнозирование не значений признаков, а непосредственно плотностей распределения классов [6].

2. Смешанное обучение в задаче адаптации СО

Разработка модели прогнозирования значений информативных параметров требует наличия большого объема данных, собранных при влиянии различных помеховых факторов. Сбор этой информации является длительным и требует значительных материальных затрат.

При недостатке априорной информации на этапе разработки СО применим подход, основанный на построении адаптивных решающих правил в рамках концепции смешанного обучения. Концепция предусматривает создание алгоритмов распознавания в условиях, когда обучающая выборка состоит из размеченных (принадлежность к соответствующим классам известна) и неразмеченных (принадлежность к классам неизвестна) данных. Так, при проектировании СО на этапе разработки алгоритма имеется небольшая обучающая выборка размеченных данных, а в процессе работы поступают неразмеченные данные, которые необходимо использовать для корректировки решающего правила.

Постановка задачи распознавания образов в рамках концепции смешанного обучения выглядит следующим образом. Существует пространство объектов X , характеризуемых набором признаков, и пространство имен классов объектов Y , существует также неизвестная функция $y^* : X \rightarrow Y$, осуществляющая отображение пространства объектов в пространство имен классов. Значения функции y^* известны лишь на некотором количестве N объектов обучающей выборки $X^M = (x_i, y_i)_{i=1 \dots M}$, остальные $M - N$ объектов являются неразмеченными. Необходимо построить алгоритм распознавания $a : X \rightarrow Y$, аппроксимирующий зависимость $y^*(x)$ на всем множестве X , используя при этом весь набор размеченных и неразмеченных примеров X^M .

Для построения адаптивного алгоритма a предлагается использовать следующий подход. На первом этапе по размеченным данным каждого класса производится кластеризация. Для кластеризации предлагается использовать метод C -средних [7], в котором каждый объект может принадлежать нескольким кластерам одновременно, но с разной степенью принадлежности. Существенным недостатком алгоритма C -средних является априорно задаваемое количество кластеров, что требует интерактивности разработки алгоритма.

Для автоматического определения количества кластеров используется метод субтрактивной кластеризации [8]. Метод субтрактивной кластеризации осуществляет поиск областей в пространстве признаков с наибольшей плотностью объектов, в результате его применения определяются центры кластеров, которые являются начальными центрами для алгоритма C -средних. В итоге получается множество кластеров каждого класса, содержащих размеченные данные.

На втором этапе осуществляется восстановление плотности распределения примеров в каждом из кластеров. Восстановление плотности распределения каждого кластера позволяет использовать расстояние Кульбака – Лейблера в качестве индикатора смещения концепта и формировать решающее правило на основе Байесовского подхода. При этом решающее правило будет выглядеть следующим образом:

$$a(x) = \arg \max_{y \in Y} \lambda_y P_y \sum_{j=1}^K p_y^j(x),$$

где λ_y – величина потери от неправильной классификации объекта класса y ; P_y – вероятность появления объектов класса y ; $p_y^j(x)$ – плотность распределения в j -м кластере объектов класса y .

На этом процесс обучения алгоритма заканчивается.

Процедура мониторинга реализует функции определения смещения концепта, его типа и определяет направление адаптации. При этом возможно несколько вариантов:

- новые данные принадлежат существующему кластеру и находятся вблизи его центра (смещение концепта отсутствует);
- новые данные принадлежат существующему кластеру и находятся на его границе (эволюционное смещение концепта);
- новые данные не принадлежат ни одному из существующих кластеров (скачкообразное смещение концепта).

Для каждой из описанных ситуаций предусмотрена своя процедура адаптации. Если новые данные принадлежат существующему кластеру и смещение концепта отсутствует, то такие данные дублируют имеющиеся размеченные и не используются для адаптации кластерной структуры, однако принадлежность текущих данных кластеру указывает на его «активность», т.е. такой кластер должен иметь больший вес в принятии решения по сравнению с остальными. Для этого в Байесовское решающее правило вводятся веса, определяющие вклад каждого кластера в принятие решения:

$$a(x) = \arg \max_{y \in Y} \lambda_y P_y \sum_{j=1}^K w_j p_y^j(x), \quad \sum_{j=1}^k w_j = 1, \quad w_j \geq 0.$$

В случае эволюционного смещения концепта в одном или нескольких кластерах адаптация заключается в повторном восстановлении плотности распределения в кластерах, в которых смещение концепта установлено. При этом новые данные включаются в обучающую выборку.

В случае скачкообразного смещения концепта, когда новые данные выходят за пределы существующих кластеров, адаптация заключается в формировании нового кластера. Принадлежность нового кластера к одному из классов определяется его близостью к кластерам соответствующих классов. В случае если новый кластер нельзя достоверно классифицировать на основе меры близости, процедура его разметки должна осуществляться учителем (разработчик, оператор и т.д.).

Рассмотрим работу предложенного адаптивного алгоритма в применении к задаче обнаружения человека и группы людей, описанной в первой части настоящей статьи. На этапе обучения в классе помех алгоритму предъявляются признаки, вычисленные по суточной записи сейсмического сигнала без воздействия дождя и ветра. На рис. 5 показаны сформированная по обучающим данным разделяющая граница и восстановленные плотности распределения кластеров.

На рис. 5 видно, что в классе помех сформировано два кластера, соответствующих дневным и ночным сигналам сейсмического фона.

На этапе функционирования в классе помех появляются примеры, вычисленные по сигналам дождя и ветра. Эти данные формируют новый кластер в классе помех (рис. 6).

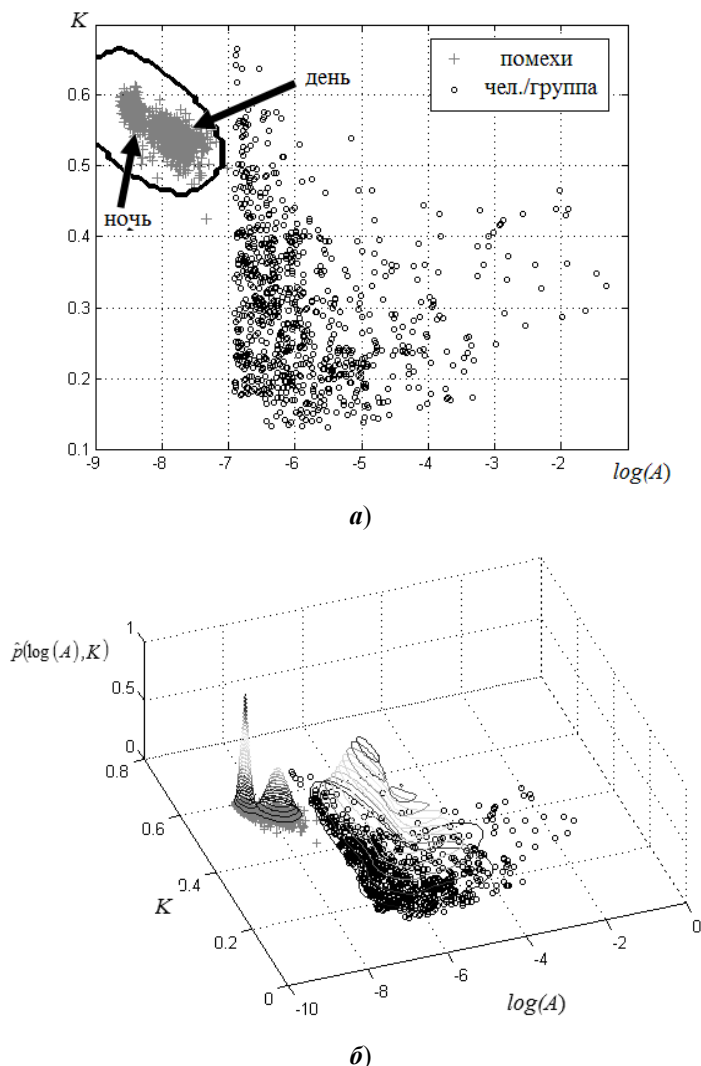


Рис. 5. Сформированная по обучающим данным разделяющая граница (а) и восстановленные плотности распределения кластеров (б)

Кроме того, поступающие данные используются для задания весов имеющимся кластерам. Кластеру, в который попадают текущие данные, назначается больший вес при принятии решения. На рис. 7 показаны сформированные на этапе функционирования разделяющие границы, соответствующие признакам, вычисленным в дневное, ночное время и при воздействии помех в виде дождя и ветра.

Таким образом, поступающие в процессе функционирования данные используются для адаптации как структуры, так и параметров решающего правила, что позволяет сократить вероятность ошибок принятия решений. Так, в приведенном примере «статичное» решающее правило позволяет обеспечить вероятность ложного срабатывания на уровне 0,046, тогда как адаптивное решающее правило – на уровне 0,031 при неизменной вероятности правильного обнаружения.

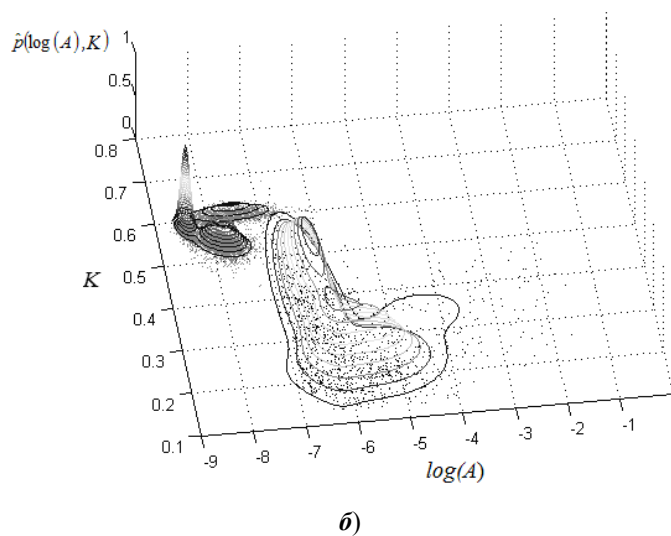
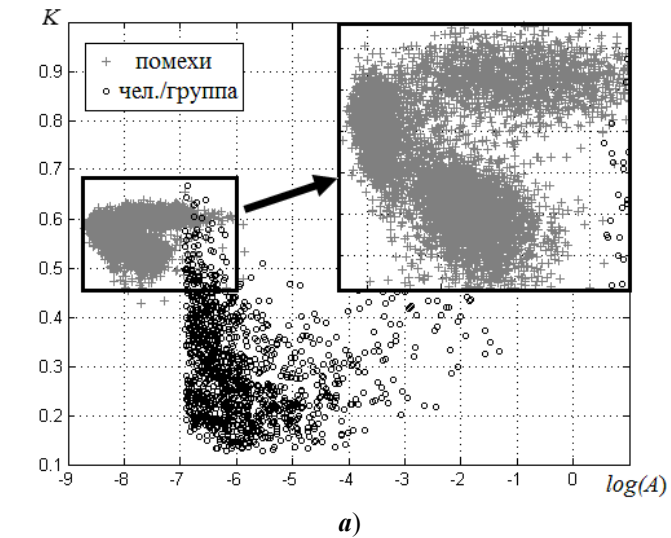
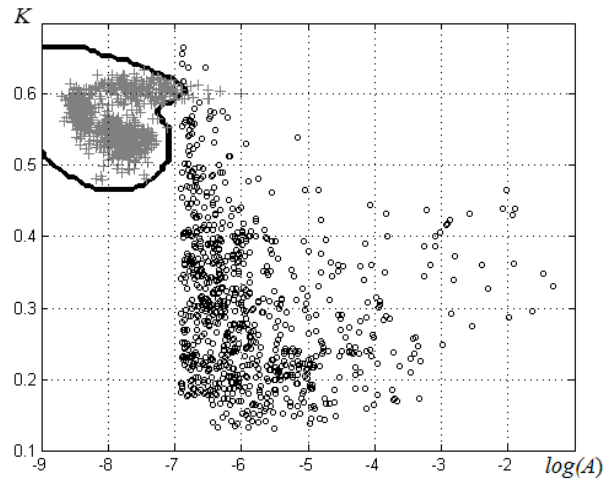


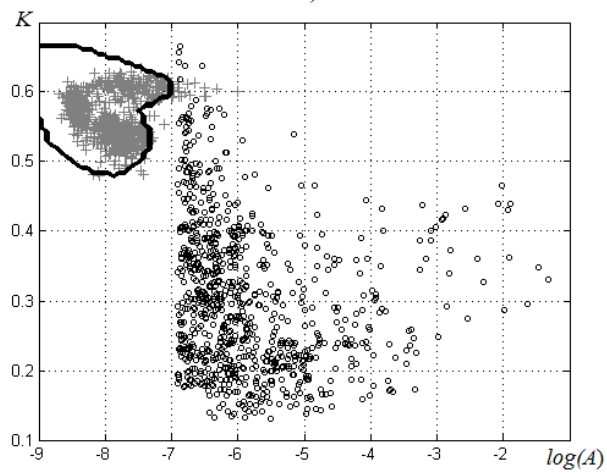
Рис. 6. Сформированные на этапе функционирования кластеры (а) и их плотности распределения (б)

Заключение

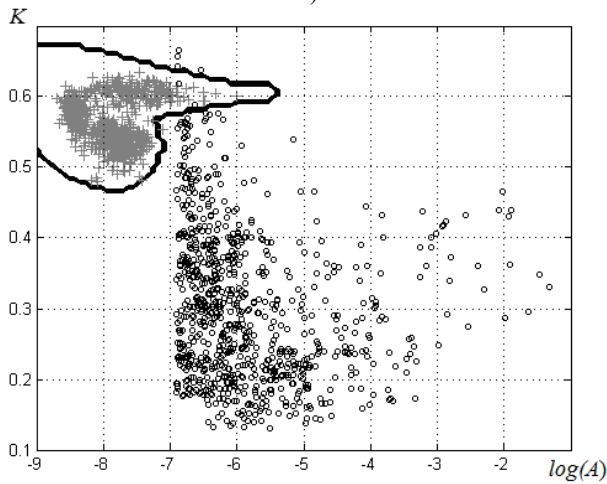
Результатом исследования является совокупность методов, реализующих два подхода к построению адаптивных алгоритмов обработки информации для СО движущихся наземных объектов в изменяющихся условиях функционирования. Предложен способ оценки величины изменений признаков описаний классов, основанный на вычислении расстояния Кульбака – Лейблера и используемый для реализации процедуры мониторинга. Метод, основанный на предобработке траекторий динамических описаний классов, требует значительного объема априорной информации для разработки модели прогнозирования информативных параметров. Метод, реализованный в рамках концепции смешанного обучения, предполагает использование информации, поступающей при функционировании СО для изменения решающего правила.



а)



б)



в)

Рис. 7. Сформированные на этапе функционирования в дневное (а), ночное (б) время и при воздействии помех (в) разделяющие границы

Список литературы

1. **Kullback, S.** On information and sufficiency / S. Kullback, R. A. Leibler // *The Annals of Mathematical Statistics*. – 1951. – Vol. 22, № 1. – P. 79–86.
2. **Айвазян, С. А.** Прикладная статистика: Исследование зависимостей / С. А. Айвазян, И. С. Енюков, Л. Д. Мешалкин ; под ред. С. А. Айвазяна. – М. : Финансы и статистика, 1985. – 487 с.
3. **Голяндина, Н. Э.** Метод «Гусеница»-SSA: прогноз временных рядов : учеб. пособие / Н. Э. Голяндина. – СПб. : Изд-во СПбГУ, 2004. – 52 с.
4. Главные компоненты временных рядов: метод «Гусеница» / под ред. Д. Л. Данилова, А. А. Жиглявского. – СПб : Изд-во СПбГУ, 1997. – 307 с.
5. **Лоскутов, А. Ю.** Применение метода локальной аппроксимации для прогноза экономических показателей / А. Ю. Лоскутов, Д. И. Журавлев, О. Л. Котляров // *Вопросы анализа и управления риском*. – 2003. – Т. 1, № 1. – С. 21–31.
6. **Hansen, B.** Autoregressive conditional density estimation / B. Hansen // *International Economic Review*. – 1994. – Vol. 35. – P. 705–730.
7. **Bezdek, J. C.** Pattern Recognition with Fuzzy Objective Function Algorithms / James C. Bezdek // PlenumPress. – New York, 1981. – 256 p.
8. **Yager, R.** Essentials of Fuzzy Modeling and Control / R. Yager, D. Filev. – USA : John Wiley & Sons, 1984. – 387 p.

References

1. Kullback S., Leibler R. A. *The Annals of Mathematical Statistics*. 1951, vol. 22, no. 1, pp. 79–86.
2. Ayvazyan S. A., Enyukov I. S., Meshalkin L. D. *Prikladnaya statistika: Issledovanie zavisimostey* [Applied statistics: investigation of dependencies]. Moscow: Finansy i statistika, 1985, 487 p.
3. Golyandina N. E. *Metod «Gusenitsa»-SSA: prognoz vremennykh ryadov: ucheb. posobie* [The “Caterpillar”-SSA method: prediction of time series: tutorial]. Saint-Petersburg: Izd-vo SPbGU, 2004, 52 p.
4. *Glavnye komponenty vremennykh ryadov: metod «Gusenitsa»* [Main components of time series: the “Caterpillar” method]. Eds. by D. L. Danilov, A. A. Zhiglyavskiy. Saint-Petersburg: Izd-vo SPbGU, 1997, 307 p.
5. Loskutov A. Yu., Zhuravlev D. I., Kotlyarov O. L. *Voprosy analiza i upravleniya riskom* [Problems of risk analysis and management]. 2003, vol. 1, no. 1, pp. 21–31.
6. Hansen B. Autoregressive conditional density estimation. *International Economic Review*. 1994, vol. 35, pp. 705–730.
7. Bezdek J. C. Pattern Recognition with Fuzzy Objective Function Algorithms. *Plenum-Press*. New York, 1981, 256 p.
8. Yager R., Filev D. *Essentials of Fuzzy Modeling and Control*. USA: John Wiley & Sons, 1984, 387 p.

Митрохин Максим Александрович
кандидат технических наук, доцент,
кафедра автономных информационных
и управляющих систем, Пензенский
государственный университет (Россия,
г. Пенза, ул. Красная, 40)

E-mail: aius@pnzgu.ru

Mitrohin Maksim Aleksandrovich
Candidate of engineering sciences, associate
professor, sub-department autonomous
information and control systems,
Penza State University (40 Krasnaya
street, Penza, Russia)

УДК 004.93::004.942

Митрохин, М. А.

Методология построения адаптивных алгоритмов обработки информации в средствах обнаружения движущихся наземных объектов (Ч. 2. Методы обработки информации для построения адаптивных средств обнаружения) / М. А. Митрохин // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2016. – № 1 (37). – С. 16–27.