

Классификация состояний беспроводной сенсорной сети с использованием методов машинного обучения

М.Н. Юлдашев¹, А.П. Адамов², А.А. Адамова¹

¹Московский государственный технический университет им. Н. Э. Баумана,

mikhail.yuldashev@gmail.ru, arina.adamova@rambler.ru

²Дагестанский государственный технический университет, apadamov@mail.ru

Аннотация — Данная статья посвящена перспективному направлению использования беспроводных сенсорных сетей в сочетании с методами машинного обучения. Основное внимание уделено задачам классификации при работе беспроводных сенсорных сетей, а именно принятию корректного решения о состоянии сети на основе информации от датчиков. Кратко рассмотрены основные задачи, решаемые беспроводными сенсорными сетями и их корреляция с задачами машинного обучения. В статье проведен анализ существующих решений и оценка эффективности внедрения методов анализа данных. В результате исследования выявлены критерии оптимизации беспроводных сенсорных сетей и предложены векторы их развития с точки зрения применения методов машинного обучения

Ключевые слова — беспроводные сенсорные сети, машинное обучение, принятие решений, оптимизация сетей, анализ данных, нагрузка сети.

I. ВВЕДЕНИЕ

С ростом вычислительных возможностей современных программно-управляемых устройств, а также наличие множества различных сенсоров сделало возможным построение сенсорных сетей обработки информации, обеспечивающих автоматизацию всевозможных процессов. В простейшем случае беспроводная сенсорная сеть представляет собой совокупность взаимосвязанных по беспроводной сети сенсоров, расположенных по контролируемой среде, где сенсор – это модуль, состоящий из чувствительного элемента, блока обработки данных, приемопередатчика и элемента питания.

Подобные системы позволяют собирать при помощи чувствительного элемента всевозможные данные из окружающей среды и затем соответствующим образом на них реагировать. При этом одна из систем может управлять поведением другой по заранее разработанным алгоритмам. Важным результатом объединения подсистем является синергетический эффект. На данный момент сенсорные сети применяются в промышленности, сельском хозяйстве, медицине и других областях.

В области разработки сенсорных сетей существует значительное количество прикладных задач, по которым представлен ряд исследований. В работах Е.А. Бакина предложены методы повышения эффективности сбора информации в беспроводных сенсорных сетях на основе оптимизации расписания; работы А.Д. Фомина охватывают исследования надежных методов агрегации данных; в работах Е.М. Линского рассматриваются подходы к управлению передачей пакетов. Также решением задач, связанных с оценкой, анализом и эффективным управлением информационных потоков в беспроводных сенсорных сетях, занимаются ученые: Чинг-Чуан Чанга, Д.А. Молчанова, А.А. Захарова, К.А. Аксенова, Л.Г. Волков, М.М. Комаров, С.Г. Ефремов и др. Большое внимание уделяют анализу и решению проблем разработки надежных сетей, а также созданию сопутствующих программных и аппаратных средств такие ведущие компании, как Philips, Ember, Samsung, IBM, Motorola, Freescale Semiconductor, NEC, LG, OKI, «Высокотехнологические системы» и др. Активные исследования и разработки в области сенсорных сетей говорят об актуальности данной тематики.

II. ПОСТАНОВКА ПРОБЛЕМЫ

Функционирование сенсорной сети заключается в сборе отдельными узлами информации об окружающей среде посредством датчиков, которая затем передается главному модулю по беспроводной сети. Главный модуль агрегирует полученные данные, после чего отображает обработанную информацию оператору, а также может принимать решения на основе заранее определенных алгоритмов. При этом оператору нет необходимости знать о показании всех сенсоров (их количество может достигать до сотен и тысяч единиц). Однако совокупная информация является наиболее важной, чтобы принимать корректные решения для последующих действий. Вследствие этого необходимо использовать алгоритмы, обеспечивающие наиболее надежное и достоверное вычисление.

При решении задач классификации использование разветвляющихся алгоритмов с постоянными

предикатами приводит к тому, что зачастую системы подстраиваются под конкретные ситуации. Перед использованием алгоритмов требуется проводить аналитическую работу по изучению данных, определению наиболее приоритетных параметров и выработке конечного алгоритма. Получившийся алгоритм зачастую находит применение в узкой области использования и является чувствительным к малейшим изменениям как самой сети, так и окружающей среды.

Сенсорные сети являются нестабильными системами, где в любой момент произвольный сенсор в силу своей неисправности может не посылать данные, или что хуже — посылать недостоверные данные. При этом линейные алгоритмы вероятнее всего будут выдавать неверные решения, т.е. система в целом не будет выполнять свою основную обязанность, что крайне недопустимо.

Для решения задач по принятию решения и классификации следует использовать алгоритмы, которые будут помехоустойчивыми и способными подстраиваться под изменяющуюся среду. К ним можно отнести гибкие алгоритмы, построенные на вероятностных и эвристических вычислениях.

III. БЕСПРОВОДНЫЕ СЕНСОРНЫЕ СЕТИ И МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

В последнее время стремительно набирают популярность методы машинного обучения (Machine Learning). Являясь математической дисциплиной, использующей разделы математической статистики, численных методов оптимизации, теории вероятностей, дискретного анализа, машинное обучение оперирует с большим количеством структурированных данных и решает задачи классификации, ранжирования и регрессии. Широко применяется в различных сферах жизнедеятельности, таких как медицина, банковское дело, IT и т.д.

Данные в беспроводных сенсорных сетях являются структурированными, основываются на локальных метаданных. Автором предлагается использование методов машинного обучения для анализа данных с сенсоров и последующего решения таких задач, как принятие решения о передаче данных, классификация состояний системы и предсказание её поведения в будущем.

Также предлагаемые методы обработки информации повышают потенциал для решения следующих задач: повышения помехоустойчивости сенсорной сети, адаптация к изменениям окружающей среды.

IV. АНАЛИЗ ПОТОКА ДАННЫХ В БЕСПРОВОДНЫХ СЕНСОРНЫХ СЕТЯХ

Первостепенным этапом при принятии решения является сбор большого объема информации (объектов). Это всевозможные данные с датчиков, представленные в определенной структуре с указанием временных меток, идентификаторов сенсора и самих

показаний (признаков объекта). Первоначально объекты поступают в модуль и помещаются в его внутреннюю память. В простейшей модели данные фактически сразу попадают в очередь для передачи по сети в главный модуль, вследствие чего возникает большая нагрузка на сеть. Главный модуль агрегирует данные со всех датчиков, анализирует и принимает на их основе решения по заранее определенным алгоритмам. Данные алгоритмы оперируют с информацией, поступившей в текущий момент и представляют собой поведенческое описание в виде конечного автомата. В результате система теряет свою гибкость, адаптацию к внешним факторам, имеет низкую помехоустойчивость.

Методы машинного обучения ориентированы на построения математической модели поведения внешней среды, где параметры данной системы определяются путем обучения на выборке данных с сенсоров и результатов принятия решения главным модулем или оператором. В последующем, применение данной модели позволит усовершенствовать процесс принятия решений.

Основной поставленной задачей при использовании машинного обучения в сенсорных сетях является классификация состояния системы в целом. Систему сенсорной сети можно описать с помощью ограниченного количества состояний. В простейшем случае достаточно обозначить два состояния: «исправное» и «неисправное». Каждый набор значений сопоставлен с конкретным состоянием системы, тем самым образуются непересекающиеся подмножества. Данная задача классификации хорошо решается с помощью дерева принятия решений. Листьями дерева будут являться заданные состояния системы, а ветви — критерии отбора. Пример подобного дерева представлен на рис. 1.

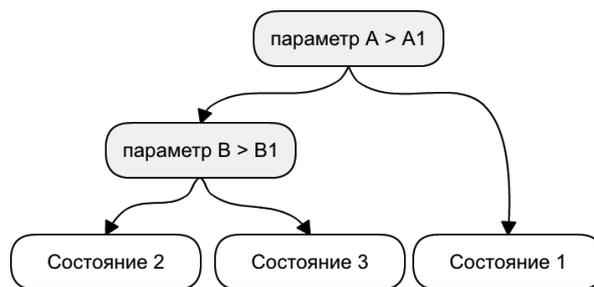


Рис. 1. Пример решающего дерева для сенсорной сети

Для определения параметров предикатов необходимо набрать обучающую выборку. Качество построенного решающего дерева напрямую зависит от качества этой выборки. Предлагаются 2 варианта получения данных для выборки:

- 1) классификация объектов в соответствии с решениями, принятыми оператором;
- 2) принятие решения на основе положительной обратной связи от сенсорной сети через определенный

промежуток времени или до определенного диапазона изменений.

На рис. 2 показан пример обучающей выборки, в результате анализа которой будет построено обучающее дерево, представленное на рис. 1.

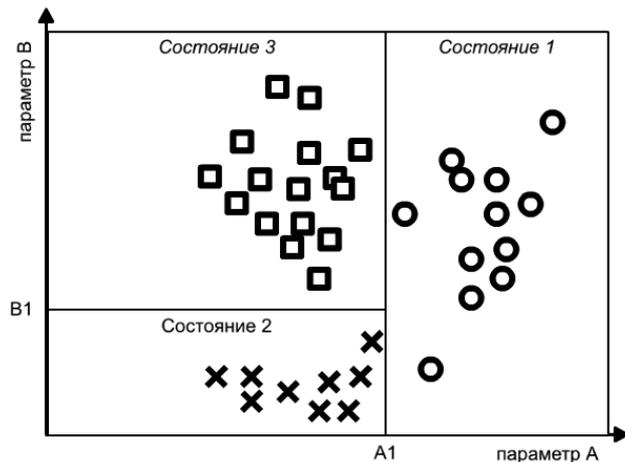


Рис. 2. Пример множества объектов классификации с разбиением на состояния

В качестве технологии построения решающего дерева можно использовать алгоритм ID3. Данный алгоритм принимает локально оптимальные решения на каждом этапе, поэтому количество вычисления в ID3 минимально, что положительно сказывается на энергозатратах узлов сенсорной сети.

Преимущества использования решающих деревьев, построенных по алгоритму ID3, заключается в том, что:

- 1) трудоемкость алгоритма принятия решения линейна по длине выборки;
- 2) не бывает отказов от классификации;
- 3) возможно принятие решения на основе неполной информации от датчиков (объекты с пропусками);
- 4) дерево является достаточно гибким с точки зрения количества параметров.

Недостатками метода ID3 является переусложнение структуры дерева и большая вероятность переобучения системы, поэтому необходимо применять методы для снижения данного эффекта (например, проводить перекрестную проверку, строить независимую композицию решающих деревьев – решающий лес).

V. РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Беспроводные сенсорные сети являются актуальной тематикой для исследования. Ученые всего мира работают над тем, как можно усовершенствовать данную систему. В статье представлена одна из задач, имеющая место в сенсорных сетях: классификация состояния системы в целом. Оператору необходимо

знать в каком состоянии находится система в текущий момент, чтобы принять соответствующие действия. При этом сенсорные сети представляют собой нестабильные системы, где в любой момент из строя может выйти узел, отвечающий за передачу данных к главному модулю. Использование в этих случаях разветвляющихся алгоритмов не даст желаемых результатов.

Предлагается использование методов машинного обучения для задачи классификации состояния системы. Помимо классификации, станет возможным решение таких задач, как упреждающее принятие решения, повышение помехоустойчивости, адаптация к изменениям окружающей среды.

Предлагаемый алгоритм построения дерева принятия решений ID3 обладает своими преимуществами и недостатками. Преимущества алгоритма значительны для поставленной задачи, а недостатки устраняются аналитическими методами, такими как усечение построенного дерева.

Последующие исследования по тематике статьи предполагают формализацию признаков изучаемых объектов, построение математической модели, определение параметров модели на основе сформулированных признаков системы и исследование корректности принимаемых решений.

ЛИТЕРАТУРА

- [1] Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: классификация и снижение размерности. — М.: Финансы и статистика, 1989.
- [2] Mitchell T. Machine Learning. — McGraw-Hill Science/Engineering/Math, 1997.
- [3] Ryszard S. Michalski, Jaime G. Carbonell, Tom M. Mitchell (1983), Machine Learning: An Artificial Intelligence Approach.
- [4] Грешилов А.А. Математические методы принятия решений : учеб. пособие. — 2-е изд., испр. и доп. — М. : Изд-во МГТУ им. Н. Э. Баумана, 2014.— 647, [1] с. : ил.
- [5] Краснобрыжий Б.В. др. Анализ беспроводных сенсорных сетей на основе стандарта ZigBee // Сборник докладов XV молодежной научно-технической конференции "Наукоемкие технологии и интеллектуальные системы" - М.: МГТУ им.Н.Э.Баумана. 2013. - С. 306-314.
- [6] Лавров А.В. и др. Исследования сенсорной сети датчиков давления // Датчики и системы. 2013. № 9 (172). С. 51-55.
- [7] Денисов А.А., Кальнов В.А., Шахнов В.А. Проектирование Наносенсоров - М. Изд-во МГТУ им. Н.Э. Баумана, 2011. Сер. Библиотека "Наноинженерия". Том 6. 126 с. ил.:
- [8] Шахнов В.А. и др. Способ функционирования беспроводной сенсорной сети // Патент РФ на изобретение №2556423 от 05.07.2013.

Support methods of reliability on the wireless sensor networks by criterion of the network load

M.N. Yuldashev¹, A.A. Adamova¹, A.P. Adamov²

¹Bauman Moscow State Technical University, mikhail.yuldashev@gmail.ru,
arina.adamova@rambler.ru

²Dagestan State Technical University, apadamov@mail.ru

Keywords — wireless sensor networks, machine learning, decision making, optimization of networks, data analysis, network load, classification.

ABSTRACT

Growth of computing opportunities of the modern program-driven devices, and also existence of a set of different sensors made possible creation of sensor networks of information processing are providing automation of various processes. Similar systems allow to collect various data from environment and then to react to them as appropriate. Such leading companies as Philips, Ember, Samsung, IBM, Motorola, Freescale Semiconductor, LG, etc. pay great attention to the analysis and solution of problems of development of reliable BS and hardware. The active researches in the field of sensor networks show relevance of this subject.

Scientists of the whole world try to improve this system. In the article, one of the tasks taking place on the sensor networks is provided: classification of status of system in general. The operator needs to know in what status system is at the moment to accept the appropriate actions. At the same time, sensor networks represent unstable systems where the node, which is responsible for data transfer to the principal module at any time, can fail. The use of the branching algorithms in these cases does not yield desirable results.

The main objective when using machine training on the sensor networks is classification of status of system in general. The system of a sensor network can be described by means of a limited number of statuses. In the elementary case, it is enough to designate two statuses: "operational" and "faulty". Each set of values is compared with specific status of system, non-overlapping subsets are formed thereby. This problem of classification is well solved with the help of decision tree. Leaves of the tree will be the given statuses of the system, and the branches – search criterions.

For determination of parameters of predicates, it is necessary to collect learning selection. Quality of the constructed decisive tree directly depends on quality of this selection.

It is possible to use ID3 algorithm as the technology for creation of decisive tree. This algorithm makes locally optimal solutions at each stage; therefore, the number of

computation in ID3 is minimum that positively affects energy consumption of nodes of sensor network.

Advantages of use of the decisive trees constructed by ID3 algorithm are as follows: labor input of algorithm of decision-making is linear on selection length, there are no failures from classification, decision-making on the basis of incomplete information from sensors is possible (objects with gaps).

Use of methods of machine training for the task of classification of status of system is offered. In addition to classification, solution of such tasks as anticipatory decision-making, noise immunity increase, adaptation to changes of environment will become possible.

The offered algorithm of creation of decision tree of ID3 has advantages and shortcomings. Advantages of the algorithm are considerable for objective, and defects are eliminated by analytical methods, such as truncation of the constructed tree.

REFERENCES

- [1] Ajvazjan S.A., Buhshtaber V.M., Enjukov I.S., Meshalkin L.D. *Prikladnaja statistika: klassifikacija i snizhenie razmernosti*. Moscow, Finansy i statistika, 1989. 272 (in Russian).
- [2] Mitchell T. *Machine Learning*. McGraw-Hill Science/Engineering/Math, 1997.
- [3] Ryszard S. Michalski, Jaime G. Carbonell, Tom M. Mitchell (1983), *Machine Learning: An Artificial Intelligence Approach*, Tioga Publishing Company.
- [4] Greshilov A. A. *Matematicheskie metody prinjatija reshenij : ucheb. posobie. 2-e izd., ispr. i dop.* Moscow, Izd-vo MGTU im. N. Je. Baumana, 2014. 647 p. fig 272 (in Russian).
- [5] Krasnobryzhij B.V. *Analiz besprovodnyh sensoryh setej na osnove standarta ZigBee // Sbornik dokladov XV molodezhnoj nauchno-tehnicheskoy konferencii "Naukoemkie tehnologii i intellektual'nye sistemy"* Moscow, MGTU im.N.Je.Baumana. 2013. pp. 306-314 (in Russian).
- [6] Lavrov A.V. *Issledovanija sensornoj seti datchikov davlenija // Datchiki i sistemy*. 2013. No. 9 (172). pp. 51-55.
- [7] Denisov A.A., Kal'nov V.A., Shahnov V.A. *Proektirovanie Nanosensorov*. Moscow, Izd-vo MGTU im.N.Je.Baumana, 2011. Ser. Biblioteka "Nanoinzhenerija". Vol. 6. 126 p. fig. 272 (in Russian).
- [8] Shahnov V.A. *Sposob funkcionirovanija besprovodnoj sensornoj seti// Patent RF na izobretenie No. 2556423 from 05.07.2013* (in Russian).