



УДК 004.9:692.78

В. В. Деревянко

канд. физ.-мат. наук

*Институт вычислительного моделирования СО РАН, г. Красноярск,
Красноярский край, Россия*

ПРИМЕНЕНИЕ DATA MINING В КОСМИЧЕСКИХ ПРИЛОЖЕНИЯХ

Представлен обзор направлений использования Data Mining в различных приложениях космической тематики: контроль качества изготовления микросхем, анализ телеметрических данных, мониторинг работы космических аппаратов в процессе полета, предпусковой анализ космических аппаратов, прогнозирование поломок, анализ данных на борту космического аппарата в процессе полета и т.д.

Ключевые слова: Data Mining, KDD, поиск аномалий, контроль качества, космические аппараты.

V. V. Derevyanko

*Institute of Computational Modeling, Siberian Branch of the Russian
Academy of Sciences, Krasnoyarsk, Russia*

USE OF DATA MINING IN SPACE APPLICATIONS

There is a review of tendencies of data mining use in the different applications in space area: quality control of microcircuit production, analysis of telemetry data, monitoring of spacecraft operation during the flight, preoperational analysis of spacecraft, breakdown prognostification, data analysis on board of the spacecraft during the flight, etc.

Key words: Data Mining, KDD, anomaly search, quality control, spacecraft.

При использовании любого устройства в космических приложениях – непосредственно в космических аппаратах или в наземном оборудовании – одной из ключевых проблем является обеспечение надежности работы устройства. Надежность работы требует тщательного контроля: контроля качества производства, контроля производительности в процессе работы, своевременной диагностики и устранения возникающих неисправностей и т.д. Подобный контроль выполняется на основе информации, поступающей с датчиков, контролирующих работу устройства. Прогресс в развитии микроэлектроники за последние 10–15 лет привел к тому, что датчики стали существенно дешевле, легче и меньше по размерам. Это вызвало увеличение количества используемых датчиков и рост объемов

телеметрической информации. Естественно, ручная обработка больших объемов информации слишком трудоемка – нужны средства автоматизации.

Задачи автоматизации поиска знаний решаются средствами интеллектуального анализа данных – Data Mining. Фактически Data Mining – это набор технологий поиска скрытых закономерностей в больших необработанных объемах данных. Data Mining является частью процесса KDD (Knowledge Discovering in databases), включающем, помимо поиска закономерностей, этапы сбора, подготовки данных и последующего анализа полученных результатов. К настоящему времени разработано множество алгоритмов и технологий Data Mining. Характерно, что универсального алгоритма для извлечения знаний из данных не существует. Каждое конкретное практическое приложение, обладающее специфическими

характеристиками, требует либо адаптации технологий Data Mining, либо разработки новой технологии обработки данных.

Целью настоящей статьи является обзор вариантов применения Data Mining в космических приложениях. На наш взгляд, все работы в этом направлении можно условно разделить на три категории. К первой относятся работы, посвященные поиску аномалий в телеметрических данных и мониторингу состояния технических систем. Ко второй – работы, посвященные повышению качества производства, тестирования компонентов системы и прогнозированию возникновения неисправностей в используемых системах. К третьей – использование Data Mining на борту летательных аппаратов. Настоящий обзор не является всеобъемлющим и основан исключительно на публикациях, доступных через Интернет.

Одним из ключевых направлений применения технологий Data Mining является автоматизация поиска аномалий. Поиск аномалий – это поиск шаблонов данных, не соответствующих ожидаемому поведению [1]. Поиск аномалий широко применяется в задачах мониторинга состояния технических систем. Для решения подобных задач используются системы ISHM (Integrated Systems Health Management). В ISHM состояние системы контролируется по показаниям датчиков. Если в работе системы возникает неисправность, в данных, поступающих с датчиков, возникают аномалии, сигнализирующие об отклонении поведения системы от нормального поведения. Типичные задачи, которые решают подобные системы мониторинга, это определение факта возникновения аномалии, локализация ее местонахождения, диагностирование возникшей неисправности и прогнозирование возникновения неисправностей.

Традиционно системы ISHM используют одновременно несколько методов диагностики аномалий, в частности [2]:

- проверку выхода значения параметра за установленные пределы;
- экспертную систему, содержащую набор правил, описывающих нормальное поведение системы (rule-based);
- математическую модель, описывающую требуемое поведение системы (model-based).

Общий принцип у традиционных алгоритмов примерно один и тот же. Вначале эксперты задают модель поведения системы,

представляющую набор правил, характеризующих поведение системы. В процессе работы системы поступающие телеметрические данные проверяются на соответствие модели. Если поведение данных начинает отклоняться от модели, то оператору, контролирующему работу системы, поступает тревожный сигнал о возможной неисправности.

У всех традиционных алгоритмов есть общий недостаток – они требуют интенсивной работы экспертов. Эксперты задают набор правил, конструируют математическую модель, устанавливают допустимые пределы значений параметров. Возрастает количество данных – возрастает количество работы, которую необходимо проделать экспертам прежде, чем система мониторинга сможет работать.

Методы, основанные на Data Mining (data-driven методы), от этого недостатка свободны. Data-driven методы строят модель поведения системы автоматически, на основе данных о нормальном поведении системы. Для обучения метода обычно достаточно несколько десятков-сотен точек нормальных данных.

Data-driven методы имеют ряд преимуществ по сравнению с традиционными:

- не требуют априорно заданных знаний о работе системы;
- не требуют системного анализа, чтобы определить соотношения между параметрами;
- способны обрабатывать телеметрические данные, поступающие от работающей системы, в режиме реального времени и очень быстро реагировать на появление аномалии; модель поведения системы очень компактна и позволяет вести работу в режиме реального времени;
- позволяют устанавливать и отслеживать взаимосвязь между большим количеством параметров;
- способны обнаруживать коллективные и контекстные аномалии [1];
- дают возможность автоматически обрабатывать архивы накопленных данных и извлекать из них полезную информацию;
- позволяют легко учитывать новые данные о нормальном поведении системы и обновлять ранее построенную модель ее поведения.

Разработки data-driven методов мониторинга активно ведутся в Японии [2] и в США [3, 4]. К настоящему времени разработано множество методов: IMS, Orca, GritBot, GMM,

LDS, одноклассовый SVM и т.д. Как показано в работе [5], в общем случае разные методы находят различный набор аномалий, так что в IMHS целесообразно применять комбинации различных методов.

Одним из наиболее перспективных data-driven методов является технология IMS, разработанная в NASA [4]. IMS использует технологию кластеризации данных и вводит понятие расстояния между векторами данных. В процессе обучения IMS анализирует данные, полученные в процессе нормальной работы системы, и строит модель поведения системы – набор кластеров. Каждый кластер определяет ограничения на значения каждого параметра в каждом конкретном входящем векторе данных. Если в базе знаний нет кластера, содержащего близкие по значению данные, это означает, что система ведет себя неожиданным способом, т.е. возникла аномалия. В случае если расстояние между текущими и нормальными данными превышает пороговое, система выдает тревожный сигнал [3].

В настоящее время IMS активно применяется в NASA. В частности, разработчики IMS проанализировали данные, полученные с шатла Колумбия во время его последнего полета. Как известно, шатл потерпел катастрофу из-за отрыва куска изоляционной обшивки, пробившей термоизоляцию на левом крыле. Отрыв произошел во время старта корабля. Однако о проблемах с термоизоляцией стало известно лишь через 17 дней, во время приземления шатла. База знаний IMS строилась на основе анализа данных предыдущих пяти полетов Колумбии. IMS выдала сигнал о возникновении неисправности в течение двух минут с момента ее возникновения.

IMS успешно используется в центре управления полетов МКС для поиска аномалий в гироскопе с управляющим моментом. IMS успешно используется для мониторинга состояния главного двигателя шатла [6]. Есть и другие примеры успешного применения [9]. Компания iSagacity разработала на основе IMS коммерческий программный продукт Process Data Miner, способный обнаруживать в данных тренды, являющиеся ранними предвестниками начинающихся изменений в процессах или возникновения проблем в оборудовании.

Как известно, NASA собирается прекратить полеты шатлов в 2010 году. На замену шатлам приходят новые аппараты – Ares. Система диагностики Ares 1-X использует

data-driven алгоритмы [7] совместно с математическими моделями и экспертными системами. Здесь необходимо отметить важный момент. По правилам NASA любые компьютерные системы, предоставляющие данные для критических решений при полетах с участием человека, должны были сертифицированы. Несертифицированные системы могут использоваться лишь в качестве «советчиков» – их результаты в обязательном порядке перепроверяются на сертифицированных системах. При использовании на Space Shuttle data-driven алгоритмы работали именно в режиме «советчиков». Система диагностики Ares изначально разрабатывалась с учетом требований сертификации. Разработчики системы диагностики Ares I-X рассчитывают на то, что им удастся ее сертифицировать.

Для первого полета Ares 1-X не было данных, накопленных при предыдущих полетах. Между тем data-driven алгоритмы использовались уже при первом полете (успешно состоявшемся 28.10.2009). Для обучения метода были частично задействованы данные, полученные при полетах шатлов для ускорителя с РДТТ, отклоняемого вектора тяги, а также данные по наземной гидравлике [8].

О том, насколько для космических аппаратов важно качество используемых компонентов, – говорить излишне. В наше время Data Mining активно используется в промышленности для повышения качества производства, оптимизации тестирования и прогнозирования выхода компонентов из строя [13].

В работе [14] авторы применили Data Mining для поиска причин выхода из строя печатных плат. При сборке печатных плат под некоторыми компонентами образуются пузырьки припоя. Подобные пузырьки могут приводить к преждевременному выходу компонента (и блока в целом) из строя. 70–80 % дефектов плат связаны с пузырьками, 20–30 % – с компонентами. Авторы показали, что большинство подобных дефектов образуется под компонентами определенного вида. Полученная информация позволила разобраться в том, что может приводить к таким дефектам, и понять, каким образом следует модифицировать процесс производства.

В работе [15] предложен метод оптимизации процесса производства интегральных микросхем. Интегральные микросхемы фабрикуются на пластинах, содержащих сотни отдельных чипов. Далее каждый чип подвергают

длительному и дорогостоящему тестированию. После чего пластины разрезают и прошедшие проверку чипы объединяют в пакеты, которые вновь подвергают тестированию. Авторы применили технологию Data Mining, позволяющую снизить количество тестов отдельных чипов. В [17] использовали технологию распознавания образов для проверки качества изделий, отлитых из алюминия. Распознавание образов – это то же Data Mining.

В работе [11] технология Data Mining была применена в FMECA-анализе. FMECA-анализ – технология анализа возможности возникновения дефектов в авиационной микроэлектронике. Традиционно такой анализ проводится на основе FMECA-таблиц, которые обычно составляются экспертами вручную. Авторы использовали Data Mining для автоматизации составления FMECA-таблицы. Результат – FMECA-таблица получилась более точной, более полной, а время ее создания сократилось с нескольких месяцев до пары секунд.

К выходу электроники из строя в процессе эксплуатации приводят не только дефекты производства. К настоящему моменту производители электроники практически полностью автоматизировали процесс производства, так что характеристики производимых изделий минимально варьируются от образца к образцу. Важнейшей причиной возникновения неисправностей являются индивидуальные характеристики эксплуатации конкретного модуля. Среднее время эксплуатации и другие традиционные характеристики надежности крайне неточны. Для более точной оценки необходимо учитывать реальную историю эксплуатации конкретного модуля.

В работе [10] предпринята попытка разработки метода точной оценки вероятности выхода из строя авиационного и аэрокосмического оборудования на основе истории его эксплуатации. Анализ данных проводился с помощью технологий Data Mining.

В работе [3] описан способ применения IMS для предсказания возникновения неисправностей в процессе предпусковой диагностики. В предпусковых тестах могут остаться незамеченными мелкие отличия, появившиеся в поведении модуля. Технологии мониторинга фокусируются в основном на различиях в значении одного параметра. Метод IMS умеет анализировать совместное поведение множества параметров и способен определить, что хотя значения всех параметров находятся

в допустимых пределах, поведение модуля в целом начало меняться и отличается от поведения, зафиксированного в ранее проведенных тестах. Такие изменения могут служить ранним сигналом о возникновении неисправности. Кроме того, результаты такого анализа помогут выявить условия снижения производительности тестируемого компонента уже после выхода компонента из строя. Наконец, результаты IMS-анализа могут являться входными данными для алгоритмов, прогнозирующих будущее поведение компонента [12].

Использование бортовых систем Data Mining [16] актуально минимум по трем причинам. Во-первых, такие системы дают возможность быстро реагировать на происходящие события. Например, на космическом корабле, направляющемся к Марсу, задержка между моментом возникновения события и поступлением соответствующей команды из центра управления полета может составлять 20 мин. Во-вторых, бортовые системы Data Mining могут грамотно расставлять приоритеты при передаче данных в центр управления. Произошло нетривиальное событие – информация о нем будет передана в первую очередь. В-третьих, бортовые системы Data Mining могут осуществлять черновую обработку данных и передавать на землю «выжимку», значительно меньшую по объему, чем исходные данные. В условиях удаленности системы от центра управления полета, больших объемов телеметрических данных и ограниченной ширины пропускания каналов связи все это делает бортовые системы Data Mining весьма актуальными. В работе [18] рассмотрены три технологии Data Mining, которые планируется использовать на борту летательного аппарата, летящего на Марс.

Можно выделить два основных направления использования технологий Data Mining в космических приложениях: для анализа архивных данных и для анализа телеметрических данных в реальном времени.

Анализ архивных данных позволяет:

- автоматически, без участия экспертов, построить модель нормального поведения системы;
- найти скрытые закономерности в данных, позволяющие исследователям постфактум проанализировать возникновение той или иной неисправности и понять, какие симптомы в данных проявлялись перед ее возникновением, каковы причины неисправ-

ности, что можно сделать, чтобы исключить ее повторение.

Анализ телеметрических данных в реальном времени служит:

- для быстрого диагностирования появления аномалии в данных;
- для отслеживания возникновения трендов в данных;
- для диагностирования появления тонких различий в поведении системы, являющихся ранними предвестниками возникновения проблем.

Ключевыми преимуществами технологией Data Mining, отличающих их от ручного анализа данных, являются:

- возможность быстрой обработки больших объемов данных;
- возможность анализа совокупности параметров системы;
- быстрота реакции на возникновение проблемы.

В большинстве случаев технологии Data Mining используются в качестве «советчиков» – результаты анализа используются человеком для принятия решений. Однако прослеживается четкая тенденция к автоматизации процесса принятия решений.

Технологии Data Mining находят все более широкое применение в космических приложениях. Фактически это становится нормой и обусловлено, прежде всего, непрерывным ростом объемов архивных данных, количества систем, генерирующих телеметрические данные, и объемов телеметрических данных. Текущие тенденции позволяют сделать вывод о том, что в дальнейшем роль технологий Data Mining в применении к космическим приложениям будет только возрастать, а круг направлений использования – активно расширяться.

Библиографические ссылки

1. Chandola V., Banerjee A., Kumar V. Anomaly Detection: A Survey, ACM Computing Surveys, Vol. 41(3), Article 15, July 2009.
2. Yairi T., Kato Y., Hori, K. Fault Detection by Mining Association Rules from House-keeping Data, Proc. of International Symposium on Artificial Intelligence, Robotics and Automation in Space, 2001.
3. Iverson D. L., Martin R., Schwabacher M. et al. General Purpose Data-Driven System Monitoring for Space Operations, AIAA Infotech@Aerospace Conference, 2009.
4. Iverson D. L. Inductive System Health Monitoring, Proceedings of The 2004 International Conference on Artificial Intelligence (IC-AI'04), CSREA Press, Las Vegas, NV, 2004.
5. Martin R. A., Schwabacher M., Oza N., Srivastava A. Comparison of Unsupervised Anomaly Detection Methods for Systems Health Management Using Space Shuttle Main Engine Data, Proceedings of the 54th Joint Army-Navy-NASA-Air Force Propulsion, Meeting, Denver, CO, May 2007.
6. Schwabacher M., Oza N., Matthews B. Unsupervised Anomaly Detection for Liquid-Fueled Rocket Propulsion Health Monitoring, Proceedings of the AIAA Infotech@Aerospace Conference, AIAA, Reston, VA, 2007.
7. Schwabacher M., Waterman R. Pre-Launch Diagnostics for Launch Vehicles, IEEE Aerospace Conference, 2008.
8. Schwabacher M., Martin R. A., Waterman R. et al. Ares I-X Ground Diagnostic Prototype, AIAA Infotech@Aerospace Conference, 2010.
9. Inductive System Monitors Tasks, Spinoff 2008, pp. 138–139.
10. Skormin V. A., Gorodetski V. I. and Popyack I. J., 2002, “Data Mining Technology for Failure of Prognostic of Avionics”, IEEE Trans. Aerosp. Electron. Syst., 38_2_, pp. 388–403.
11. Tso K. S., Tai A. T., Chau S. N., Alkalai L. On Automating Failure Mode Analysis and Enhancing its Integrity, PRDC 2005: 287–294.
12. Schwabacher M., Goebel K. A Survey of Artificial Intelligence for Prognostics, Working Notes of 2007 AAAI Fall Symposium: AI for Prognostics, 2007.
13. Harding J. A., Shahbaz M., Srinivas S., Kusiak A. Data Mining in Manufacturing: A Review, ASME Transactions: Journal of Manufacturing Science and Engineering, Vol. 128, No. 4, 2006, pp. 969–976.
14. Kusiak A., Kurasek C. Data Mining of Printed-Circuit Board Defects, IEEE Transactions on Robotics and Automation, Vol. 17, No. 2, 2001, pp. 191–196.
15. Fountain T., Dieterich T., Sudyka B. Mining IC test data to optimize VLSI testing, KDD '00: Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, New York, NY, USA, ACM Press (2000) 18–25.
16. Tanner S., Stein C., Graves S. J. On-board Data Mining in Scientific Data Mining and Knowledge Discovery by M. M. Gaber (Editor), Springer Verlag GmbH, 2009, pp. 345–376.
17. Hernández S., Saez D., Mery D. Neuro-Fuzzy Method for Automated Defect Detection in Aluminium Castings, ICIAR (2) 2004: 826–833.
18. Castano R. et al. On-board analysis of uncalibrated data for a spacecraft at Mars in Proceedings of the Thirteenth International Conference on Knowledge Discovery and Data Mining, 2007, pp. 922–930.