Компьютерная математика

УДК 519.68

А.С. САМОСЕНОК

МЕТОДЫ АВТОМАТИЧЕСКОЙ ОБРАБОТКИ НЕПОЛАДОК В КОМПЬЮТЕРНЫХ СЕТЯХ

Описываются наиболее распространенные математические методы для решения задачи поиска первопричины неполадки – использование нейронных сетей и экспертных систем на основе набора правил. Предложена модель эффективной системы обработки неполадок и корреляции сообщений об ошибках, основанная на нейросетевом подходе.

Введение. В конце 80-х в начале 90-х годов прошлого века развитие телекоммуникационных технологий получило качественно новый импульс. Основной причиной такого роста стало развитие Интернета и построение больших корпоративных систем. Индустриальные провайдеры начали использовать инновационные технологии передачи данных и как результат значительно осложнились методы мониторинга компьютерных систем.

Одной из наиболее важных задач, с которыми столкнулись разработчики систем поддержки работоспособности сети, конечно же является эффективное управление сбоями и ошибками, возникающими во время функционирования компьютерных систем. Благодаря высокой пропускной способности современных магистральных каналов передачи данных даже кратковременный выход со строя сети может привести к задержке огромных объемов данных и как результат к катастрофическим последствиям. Таким образом, построение быстрых и эффективных механизмов для определения и локализации поломок является наиболее актуальной областью применения математических алгоритмов при поддержке безотказной передачи информации. Методы диагностики зависят от сообщений, сигнализирующих о поломке, уровня сетевого взаимодействия (физического в случае разрыва кабеля или логического в случае неправильной настройки сетевого оборудования) и степени критичности неполадки (будь-то потеря одного пакета или выход со строя локальной сети целого офиса).

В зависимости от месторасположения и технических возможностей устройств мониторинга, система регистрации сообщений об ошибках может как получить огромное количество избыточных сигналов для некоторой поломки, так и не получить никакой информации, касающейся остальных сбоев. Нужно также учитывать, что сообщение о единичной поломке на физическом уровне может распространиться по всей длине канала и вызвать большое количество одновременных сигналов об ошибке на логическом уровне. Это усложняет и без того не тривиальное задание определения точного местонахождения источника неисправности. Функционирование такой системы заключается в сборе сообщений об ошибках непосредственно с сетевого оборудования, их первичная обработка и корреляционный анализ (определение связей между различными сообщениями, а также с элементами сетевой топологии) и определение истинной первопричины (поломки) аварийной ситуации (так называемый root-cause analysis) [1]. Каждый из перечисленных этапов заслуживает отдельного подробного рассмотрения, но мы в этой работе сосредоточим внимание на математических алгоритмах и опишем технические подробности лишь в мере, достаточной для понимания обычного читателя.

Корреляционный анализ (correlation analysis)

В математической статистике и теории вероятности корреляция является мерой линейной зависимости двух случайных величин. Математической мерой корреляции двух случайных величин является коэффициент корреляции. В контексте систем мониторинга сети под корреляцией понимают наличие взаимосвязи между сигналами об ошибке или разбиение всего множества сообщений на классы (события) в соответствии с общей для них первопричиной сбоя [2]. То есть, первоочередной задачей является установление наличия факта связи между различными сообщениями об ошибках, а не расчет количественных характеристик таких зависимостей.

Эту проблему сейчас в большей или меньшей мере решают несколько систем, используя различные методы обработки данных. Такие системы используют топологические или не топологические шаблоны для объединения сообщений об ошибке, которые сгенерированы в результате одной общей неисправности в одно событие. В случае не топологической корреляции система не осуществляет привязку зарегистрированных сообщений к элементам топологии сети. Как результат, сигналы обрабатываются достаточно быстро, но в одно логическое событие могут быть объединены несколько абсолютно независимых сетевых проблем. Логично, что в таком случае автоматизированная локализация повреждения сети практически невозможна.

При использовании методов топологической корреляции сообщения об ошибках ставятся в соответствие к элементам топологии компьютерной сети. Это позволяет различить несколько независимых друг от друга проблем пусть даже произошедших в одно время (с точностью до некоторого временного интервала). Каждое сообщение об ошибке, которое проходит этап топологической корреляции, предварительно ассоциируется с одним или несколькими топологическими объектами. Таким образом, благодаря анализу сообщений, полученных от соседних объектов сети, проводится их корреляция. Однако в ситуациях с неполными входными данными (когда не все сгенерированные ошибкой сообщения регистрируются системой мониторинга) однозначная идентификация поломки достаточно проблематична и зачастую

сводится к определению некоторого «проблемного» сегмента сети. Ответственность за дальнейшее принятие решения ложится на системного администратора со всеми вытекающими из этого последствиями.

Корреляция, основанная на системе правил (Rule-based correlation)

Одним из наиболее ранних и простых предложенных алгоритмов является корреляция на основе правил. В функционирующей системе, которая обеспечивает мониторинг сети провайдеров услуг магистрального уровня существует большое количество сценариев корреляции. Для удобного управления, редактирования и добавления новых сценариев такие системы используют корреляционные правила. Сценарий корреляции может состоять как минимум из одного правила. Процесс обработки событий контролируется набором правил, критериям которых удовлетворяют полученные сообщения. Согласно [3] структура такого корреляционного модуля состоит из трех уровней:

данные (Data level) – в нашем случае это непосредственно сообщения об ошибке (или другие данные о возникнувшей ошибке);

уровень правил (Knowledge level) – набор правил или база знаний, которая составлена на основе экспертных оценок, для обработки неполадок;

контрольный уровень (Control level) – механизм, обеспечивающий корректную обработку сообщений о проблеме, исходя из существующих правил.

В отличие от традиционных алгоритмов, контрольный уровень и набор правил разделены и, соответственно, могут быть расширены или отредактированы без изменения программного кода самого корреляционного модуля. Как упоминалось выше для анализа правил необходима некоторая управляемая логика. В простейшем случае все правила могут оцениваться по очереди до тех пор, пока не будут выбраны наиболее релевантные. В более сложных системах возможны и более комплексные решения.

Преимуществом этого алгоритма является достаточно эффективное использование существующих баз знаний и простота описания (особенно если используется условный оператор «если – то»), не требующая высокой квалификации администратора. Такие алгоритмы успешно применяются для решения простых проблем, возникающих, как правило, на физическом уровне, когда взаимодействия объектов системы значительно сложнее (при проблемах на уровне приложений, а также сетевом и транспортном уровне) такие алгоритмы должны учитывать очень большое количество сценариев. Что касается других недостатков, то здесь в первую очередь следует отметить сложность добавления и редактирования правил, так как их набор – это результат работы эксперта, обладающего знаниями предметной области, и инженера, способного внедрить новые правила в систему. Соответственно корреляционные модули, основанные на системе правил подвержены сбоям в случае возникновения новых, непредвиденных ранее ситуаций.

Корреляция на основе нейронных сетей

Разнообразные корреляционные алгоритмы на основе нейронных сетей используют их способность прогнозировать поведение нелинейных динамических систем и имеют следующие преимущества [4]:

– обучение на примерах – позволяет нейронной сети выявлять скрытые и сложные зависимости в сети, успешно бороться с шумами и неполнотой информации;

– обобщающие возможности – нет необходимости в детальной модели сети и знании полной информации об элементах сети и связях между ними, в процессе обучения нейронная сеть имеет способность выделять существенные признаки для корреляции и абстрагироваться от малозначимых;

– эффективность – архитектура нейронной сети изначально ориентирована на параллельную обработку, что позволяет достичь высокой эффективности;

– дообучение – возможность дообучения в процессе эксплуатации для корректного отслеживания изменений в сети.

Поиск причины неисправности (Root cause analysis)

Рассмотрим более подробно частичный случай мониторинга неисправности сети состоящей из оптоволоконных каналов передачи данных (SONET – Synchronous Optical Networking). Не углубляясь в технические подробности процесса передачи данных в оптоволоконных каналах связи предложим эффективный алгоритм поиска неисправности с использованием аппарата нейронных сетей.

Предполагается, что данный алгоритм, основанный на теории нейронных сетей, позволит идентифицировать аварийный участок в условиях неполноты исходных данных. То есть, когда поток пришедших алармов не дает возможности со стопроцентной вероятностью говорить о единственно верном решении задачи (приняв общее количество алармов, инициированных разрывом одного участка кабеля за 100 %, ставится задача добиться верного решения о первопричине сбоя на основе, 80 % от общего количества алармов, что соответствует реальным условиям).

Схема работы корреляционного модуля выглядит следующим образом:

после исследования топологии информация о сетевых устройствах сохраняется в базе данных. На основе этих данных генерируется обучающая выборка для настройки весов нейронной сети. Этот механизм подчинен определенным правилам, основанных на реальном взаимодействии элементов сети. Фактически, генератор обучающей выборки для данной топологии решает обратную задачу к задаче поиска неисправности, т. е., сгенерировав возможную неполадку фиксирует все возможные сообщения об ошибках, возникнувших в системе;

на основе обучающей выборки производится обучение и тестирование нейронной сети;

обученная нейронная сеть принимает сообщения о реальных неполадках и осуществляет поиск их первопричин в реальном времени.

Принцип работы системы обработки неполадок показан на рисунке.

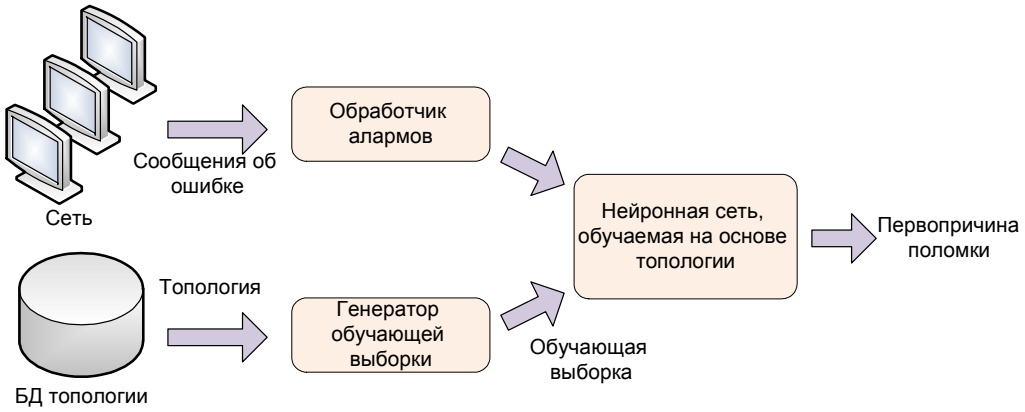

РИСУНОК. Принцип работы системы обработки неполадок на основе нейронной сети

Под топологией сети будем понимать ее логическую структуру, состоящую из оптоволоконных соединений. Представим ее в виде графа (**Sites**-вершины, **Cables**-ребра). В каждом **Cable** размещаются **Fiber**-ы (жилы). Эти данные формируют физическую топологию сети. Кроме того на графе выделим каналы передачи данных (подграфы) – **Networks**, т. е., некоторые маршруты на топологическом графе, проложенные по оптическим жилам. В отличие от ip сетей такие маршруты в нашем случае перманентны и исключают динамическую маршрутизацию. Вся сеть полностью состоит из таких каналов (любая оптическая жила должна принадлежать одному определенному каналу). Именно для каналов передачи данных регистрируются сообщения об ошибках – алармы (**alarms**). Задача состоит в том, чтобы зафиксировав поток таких сообщений об ошибке, определить первопричины возникновения аварийной ситуации на физическом уровне (выяснить точное место разрыва линии передачи данных с точностью до идентификатора участка кабеля).

В задаче корреляции приходящих из сети сообщений можно выделить следующие трудности [5]:

– шумы – из сети приходит много ненужных, избыточных, повторяющихся, случайных сообщений, происходят частые осцилляции (исправно-неисправно);

– скрытые зависимости – любая модель упрощает предметную область, поэтому появляются "невидимые" в данной модели элементы и связи, и их неисправность проявится в генерировании сообщений на другие элементы сети;

– сложные зависимости – реальные зависимости между элементами сети могут быть сложнее и разнообразнее, чем это представлено в модели;

– потеря сообщений;

– интенсивность потока сообщений: в сети может одновременно возникать большое количество неисправностей, что порождает поток сообщений высокой интенсивности (более 50 сообщ. / сек.). Это дополнительно усложняет корреляцию и накладывает строгие требования на производительность системы обработки сетевых неполадок;

– большое количество объектов сети – благодаря высокой степени детализации модели, в больших сетях база топологии может содержать несколько миллионов объектов и занимать порядка терабайта дискового пространства.

Предположения модели:

– топология полносвязна;

– каждый cable связан только с двумя site-ами;

– каждый fiber относится только к одному Cable;

– каждый fiber ассоциируется с одним маршрутом (network).

Данные предлагается представлять в виде четырех бинарных матриц, однозначно задающих топологию сети.

1. Матрица SitesxCables – матрица связности ребер и вершин топологического графа.

В каждом столбце могут быть только две единицы (каждый cable связан только с двумя вершинами), тогда как в строке их количество не ограничивается (из одной вершины могут выходить несколько ребер). Полносвязность графа подразумевает наличие между двумя любыми вершинами некоторого маршрута.

2. Матрица FibersxCables – матрица соответствия жил ребрам.

Ограничения:

– любой fiber может находится только в одном Cable (в каждой строчке матрицы только одна «1»);

– в каждом Cable может размещаться как минимум один fiber, как максимум, maxFib (16), т. е. в каждом столбике количество единиц варьируется от 1 до 16.

3. Матрица FibersxNetworks показывает маршруты (Networks) проложенные по жилам.

4. Матрица AlarmsxNetworks – фактически показывает в каких каналах связи зарегистрированы сообщения об ошибках.

Первые три матрицы однозначно задают топологию сети и используются для генерации обучающей выборки. Последняя, фактически, является матрицей входных данных нейронной сети, для которой и производится поиск первопричины сетевого сбоя.

Для реализации модели была выбрана нейронная сеть каскадно-корреляционной архитектуры [6]. Согласно этому алгоритму в начале строится минимальная нейронная сеть, состоящая из двух слоев, и затем в процессе обучения последовательно добавляются новые нейроны, формируя, таким образом, многослойную структуру. После добавления в сеть очередного нейрона значения его входящих весов фиксируются и процесс обучения продолжается. Каскадно-корреляционная архитектура имеет ряд преимуществ перед аналогичными алгоритмами, среди которых следует отметить достаточно быстрое обучение и формирование структуры сети непосредственно во время обучения.

Заключение. Рассмотрены общие принципы архитектуры систем автоматической обработки неполадок компьютерных сетей и описан нейросетевой подход к корреляции сообщений о неисправностях для компьютерных сетей со статической маршрутизацией. Предложена полноценная модель построения корреляционного модуля. Показаны преимущества нейросетевого подхода к корреляции сообщений о неисправностях по сравнению с классическими топологическими методами.

О.С. Самосьонок

МЕТОДИ АВТОМАТИЧНОЇ ОБРОБКИ НЕСПРАВНОСТЕЙ
В КОМП‘ЮТЕРНИХ МЕРЕЖАХ

Описано найбільш розповсюджені математичні методи для розв‘язку задачі пошуку першопричини несправності – використання нейронних мереж і експертних систем на основі набору правил. Запропонована модель ефективної системи обробки несправностей і кореляції повідомлень про помилки на основі теорії нейронних мереж.

A.S. Samosonok

MATHEMATICAL METHODS IN FAULT MANAGEMENT SYSTEMS

A model of fault management system is described as a result of implementation of different mathematical algorithms such as neural networks and rule-base reasoning. New effective fault management system architecture based on neural network algorithm is proposed.

1. *Andreas Muller*. Event correlation Engine. – Zurich: Institut für Technische Informatik und Kommunikationsnetze, 2009. – P. 35–45.
2. *Zeng H. and Huang C.* Fault detection and path performance monitoring in the meshed all-optical networks. // IEEE GLOBELCOM `04. – 2004. – N 3. – P. 2014 – 2018.
3. *Robert N. Cronk, Paul H. Callahan, and Lawrence Bernstein.* Rule based expert systems for network management and operations: An introduction // IEEE Network Magazine. – September 1988. – P. 7 – 21.
4. *Уоссерман Ф.* Нейрокомпьютерная техника. – М.: Мир, 1992. – С. 1 – 184.
5. *Meira D.M*. A Model for Alarm Correlation in Telecommunications Networks. – Belo Horizonte, 1997. – P. 1 – 169.
6. *Scott E. Fahlman, Christian Lebiere*. The Cascade-Correlation Learning Architecture. – Pittsburgh: Carnegie Mellon University, 1991.

Получено 07.11.2010

Об авторе:

Самосенок Александр Сергеевич,
аспирант Института кибернетики имени В.М. Глушкова НАН Украины.