Применения байесовых сетей для оценки надежности автоматизируемых систем с особыми требованиями к обеспечению безопасности

И.М. Панасенко, В.М. Бабиков (ИПУ РАН)

Показано, что наилучшие результаты в процессе формализации и числовой оценки воздействия разного рода свидетельств на надежность системы могут быть получены в рамках подхода с построением байесовых сетей доверия (БС). Схематически изложены основные положения введения в теорию БС. Рассмотрены некоторые наиболее распространенные системы, в повышении надежности которых успешно используется методология БС.

Введение

Системы с особыми требованиями к обеспечению безопасности (СОТБ) представляют собой отдельный класс автоматизируемых систем. В первую очередь к нему следует отнести системы управления аварийно-опасными объектами и производствами. Наряду с оценкой и обеспечением надежности аппаратного обеспечения СОТБ одной из важнейших является проблема оценки влияния дефектов ПО на общую надежность системы. Именно они наиболее трудны для предсказания и выявления [1-3].

В СОТБ значимыми являются проектные дефекты, а также дефекты управления, напрямую связанные с ошибками человека. Как правило, число наблюдаемых неисправностей и дефектов такого рода слишком мало для получения достоверных оценок. Эти дефекты могут быть оценены экспертами только качественно, и зачастую такие оценки делаются на основании данных из разнородных источников. В комплексных оценках надежности СОТБ, включающих оценки надежности ПО, нельзя полагаться только на хорошо разработанные методологии, базирующиеся на статистических и прочих надежностных моделях, в которых невозможно учитывать такого рода дефекты. Одним из наиболее перспективных направлений в решении проблемы оценки надежности ПО сегодня представляется построение так называемых "случаев безопасности" или "аргументов безопасности", которые являются собранием различного рода фактов (свидетельств), относящихся как к процессу разработки, так и к конечному продукту. Виды собираемой информации и пути ее сбора в настоящее время относительно хорошо известны, например, посредством шаблонов документов и проверочных списков.

Заметные результаты в процессе формализации и числовой оценки воздействия разного рода свидетельств на надежность системы получены в рамках подхода с построением БС.

Для начала рассмотрим пример (рис. 1) из монографии [1], представляющей, по сути, первое на русском языке систематическое изложение теории БС в сравнении алгебраическими БС.

Ориентированный граф с вершинами-овалами представляет собой упрощенную, но вполне жизненную, логико-вероятностную зависимость между падением ПК со стола и возможными следствиями этого падения. Исходные таблицы условных вероятностей событий (в прямоугольниках на рис. 1) задаются специалистами в предметных областях. Такой граф является удобным инструментом для анализа как следствий: падения компьютера, появления на жестком диске трещин и пр., так и возможных причин того, что на экране монитора нет изображения, нарушен один из контактов и пр.

Обратимся к следующему примеру. Пусть W_1 и W_2 – два независимых индикатора. Допустим, стало известно, что индикатор X свидетельствует о совпадении значений W_1 и W_2 .Тогда значения индикаторов W_1 и W_2 становятся условно зависимыми, то есть значение одного из них однозначно определяет значение

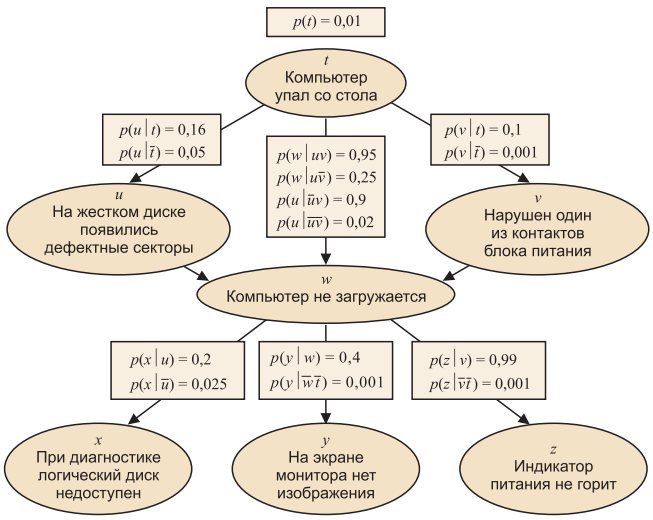

Рис. 1

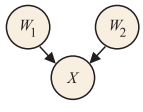

Рис. 2

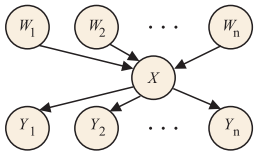

Рис. 3

второго. Граф, представленный на рис. 2, задает один из возможных видов причинно-следственных связей ("сходящихся") между переменными X, W_i и W_j. Цепочка вида W_i, X, Y_j задает связь "последовательного" вида между соответствующими переменными (рис. 3). И, наконец, связь между X, Y_k и Y_l называется связью "расходящегося" вида.

Таким образом, можно дать определение БС (BBN – Bayesian Belief Network), представляющей собой ориентированный граф без циклов, вершины которого (переменные X, W_i и Y_j) являются событиями с их возможными "ценностями" или "состояниями", составляющими полное пространство взаимно независимых событий. Дуги отражают зависимость (обычно причинную) между переменными. Эта зависимость задается таблицами условных вероятностей, приписываемыми каждой вершине и отражающими "цену" данной вершины при условии "цен" вершин, влияющих на данную вершину. Для построения таблиц могут быть использованы как статистические оценки, полученные из предыдущего опыта или модельных экспериментов, так и экспертные оценки, и индивидуальные суждения. Фрагмент БС и представлен на рис. 3.

Для различных типов анализа наиболее важная задача БС – модификация характеристик вершин после произошедших событий. Эпизодическая специфическая информация, связанная с событием, называется свидетельством. Существуют два типа свидетельств: прямое – свидетельство для вершины X, состояние которой зафиксировано в соответствии с произошедшим событием; косвенное – любое свидетельство, которое дает возможность модифицировать предшествующие вероятностные характеристики состояния вершины X.

Цель конструирования БС – нахождение вероятностей состояний событий X по ряду априорных свидетельств – e, и изменения в этих данных при появлении новых свидетельств, то есть расчет апостериорных вероятностей $Pr(x|e)$ для каждой вершины X, не имеющей свидетельств. При этом вероятности состояний событий определяются как степень доверия, корректируемая при поступлении новых данных. Вектор доверия BEL(X) означает апостериорное распределение вероятностей переменной X. На базе этого вектора эксперт (агент) выносит заключение об изменении в оценке надежности системы.

Каждая вершина делит сеть на два разобщенных дерева: одно включающее вершины, ассоциирующиеся с родителями W_j переменной X (причинное) и другое – с ее детьми Y_i (диагностическое). Все свидетельства e для вершины X, если ее состояние не основывается на прямых свидетельствах, можно разделить на: верхние ("родительские"), через выражение e_x^+ обозначаются свидетельства, передаваемые по дугам между узлом X и родителями; и нижние ("детские"), через выражение e_x^- обозначаются свидетельства, передаваемые по дугам между узлом X и детьми, и, используя правило Байеса, найти

$$BEL(x) = Pr(x \mid e_x) = Pr(x \mid e_x^-, e_x^+) = Pr(e_x^- \mid e_x^+, x)\ Pr(x \mid e_x^+)/Pr(e_x^- \mid e_x^+).$$

Естественным представляется рассмотрение случая, когда $Pr(e_x^- \mid e_x^+, x) = Pr(e_x^- \mid x)$.

Такое условие графически интерпретируется как d-разделение множеств e_x^- и e_x^+.

Если трактовать $1/Pr(e_x^- \mid e_x^+) = \alpha$ как нормализующий множитель, определяемый из условия:

$$\sum_x \mathrm{BEL}(x) = 1,$$

тогда исходное выражение для доверия BEL(x) в узле X приобретает симметричный вид:

$$\mathrm{BEL}(x) = Pr(x \mid e_x^-, e_x^+) = \alpha Pr(e_x^- \mid x)\ Pr(x \mid e_x^+).$$

Вводятся следующие понятия:

$\lambda(x) = Pr(e_x^- \mid x)$ – диагностический вектор (правдоподобие),

$\pi(x) = Pr(x \mid e_x^+)$ – вектор причины (априорных вероятностей).

Таким образом:

$$\mathrm{BEL}(x) = \alpha\, \lambda(x)\, \pi(x).$$

Для расчетов используется понятие "послания", поступающего от одной вершины к другой, и включающий это понятие алгоритм расчета вероятностей, характеризующих вершины. В настоящее время разработаны разнообразные методики создания БС в широком спектре предметных областей, эффективные алгоритмы расчетов доверия, реализованные в ряде пакетов прикладных программ. Рассмотрим некоторые наиболее распространенные системы оценки надежности СОТБ с использованием методологии БС.

Системы, готовые к коммерческому применению без предпродажной подготовки

Системы, готовые к коммерческому применению без предпродажной подготовки – COTS (Commercial-Off-The-Shelf), преимущественно разрабатываются в рамках коммерческого подхода, и тестируются в соответствии с коммерческими стандартами и спецификациями. Критерии выбора ПО могут охватывать три области: создание гарантий, проверка функций безопасности и работоспособности продукта [4].

COTS более функциональны и гибки относительно аналогов, для которых необходима предпродажная подготовка. В частности, для них возможно преодолеть существенный недостаток традиционного метода оценки безопасности систем, а именно невозможность учитывать "наложения", то есть нарушения, действующие совместно или одновременно.

COTS применяются в электроэнергетике, авиации, химическом производстве, автомобилестроении, военной технике, медицинском приборостроении и пр. Значительный интерес к потенциалу COTS наблюдается в ядерной энергетике. Так в работе [2] в качестве примера для иллюстрации правил применения и возможностей COTS выбран ядерный реактор.

Исследуются основные категории нарушений функционирования ядерного реактора, и сформулированы четыре стадии процесса моделирования системы:

1. трансформация измерений безопасности в вероятностные величины по критериям COTS;

2. исследование вероятностей нарушений с целью получения информации для построения БС;

3. использование данных из реальных систем для апостериорной модернизации БС;

4. рассмотрение результатов, которые могут оказать влияние на оценку безопасности COTS.

Большие перспективы имеет применение методологии БС для проведения независимого анализа качества ПО при помощи COTS. Такой анализ позволяет обнаружить и тестировать редко встречающиеся нарушения, значимость которых может быть так высока, что определение "достаточно близко" не приемлемо. Разработанные алгоритмы БС могут быть использованы в экспертных системах для оценки тестов, необходимых в гарантированной оценке надежности систем.

Использование сетей Байеса для оценки безопасности и риска

Основная задача системы SERENE (Safety and Risk Evaluation Using Bayesian Nets) состоит в развитии методологии и инструментария оценок безопасности ПО СОТБ [5-10]. Применение этой системы способствует согласованию различных типов суждений с использованием числовых расчетов, позволяющих сравнивать альтернативные стратегии безопасности. Базирующийся на свидетельствах подход, характерный для системы SERENE в значительной мере исключает недостатки подходов, основанных на стандартах.

При построении БС в данной системе используются идиомы – повторяющиеся в сети структуры [2,5]. Для конструирования аргументов безопасности могут также применяться шаблоны БС [6].

Эксперты отмечают следующие возможности системы SERENE:

1. приводить различные типы оценок к единой оценке;

2. подробно исследовать неопределенности, связанные с причинами нарушения безопасности;

3. фиксировать свойства, повышающие уровень безопасности;

4. усовершенствовать связи аргументов безопасности;

5. давать числовые оценки для сравнения стратегий безопасности;

6. выявлять слабости продукта и способствовать его усовершенствованию;

7. определять степени ограничений, связанных с предсказанием безопасности;

8. вносить рациональность в дискуссию экспертов;

9. создавать базу для гарантированной оценки доверия экспертов к безопасности и рациональности существующих стандартов, вытекающих из общей точки зрения на "лучшую практическую эксплуатацию".

В SERENE предусмотрены также возможности облегчения значительных трудностей, связанных с созданием БС, особенно на ранних стадиях проектирования, для чего предлагаются: способ иерархической декомпозиции и типовые образцы для построения БС снизу вверх. Дается формальное руководство в получении знаний от экспертов для заполнения таблиц, позволяющее избежать известных установок и предубеждений экспертов (например, эксперты лучше воспринимают частоту, чем вероятность).

Система SERENE успешно применена при разработке и оценке безопасности программируемых электронных систем по следующим базам свидетельств: данные о качестве и безопасности схем (планов), журналы риска, проектная документация, документация результатов тестирования, аудита и опыта управления, укомплектованность персонала [2].

В работе [3] рассматриваются возможности системы SERENE при проведении анализа безопасности или уменьшении рисков эксплуатации систем контроля и управления ядерными реакторами, и излагаются результаты работы на корпорации Electricite'de France (EDF) по построению БС. Модель БС такого объекта содержит большое число вершин (в рассматриваемом случае более 100 единиц), имеет сложную структуру, которую невозможно представить единым плоским графом. Создается глобальная структура, отражающая последовательность стадий развития процесса. Каждый шаг процесса является входом для следующего шага и включает проверку на "соответствие", которая охватывает достаточное число аспектов и характеризуется как исчерпывающая. Было обнаружено, что несмотря на ряд различий моделей аргументов безопасности, получаемых разработчиками проекта SERENE, в них можно выделить всего пять типичных образцов. Инструментарий SERENE поддерживает как иерархическую декомпозицию, так и эффективность механизма использования типовых образцов.

Следует отметить, что SERENE применима для любых особо опасных систем при условии доступности соответствующей документации и результатов тестирования. Естественно, для получения надежных прогнозов модель потребует многих лет калибровок. (Последнее заключение верно и для любых других систем оценки безопасности ПО с использованием БС). Она также окажет помощь в выявлении переменных, в наибольшей степени влияющих на безопасность в процессе проектирования СОТБ.

Проект SERENE может быть выбран для демонстрации возможностей использования БС в качестве способа представления аргументов безопасности. Были проанализированы возможности и преимущества трех наиболее общих подходов к представлению аргументов безопасности (естественные языки, табличные структуры, сети Байеса и образы, структурированные по цели) на предмет четкого описания и оценки уве-

ренности в аргументах безопасности при наличии свидетельств, и предпочтение было отдано БС.

Однако отмечаются некоторые проблемы и ограничения использования БС:

1. эксперты могут быть излишне оптимистичны в суждениях;

2. заполнение таблиц – обыкновенно субъективный процесс, и определение величин в таблице требует от экспертов владения большим запасом знаний о рассматриваемых системах и технологии БС;

3. до тех пор, пока система не будет находиться в эксплуатации в течение достаточно длительного времени, трудно быть уверенным в числовых данных таблицы;

4. если вероятностное описание некоторых вершин определяется как "мнения экспертов", уровень точности числовой оценки может быть значительно снижен;

5. подход БС может привести к ложному ощущению надежности, когда уровень риска определяется как "по-видимому, низкий" что не соответствует действительности.

Оценка надежности систем критической безопасности посредством объединения измеримых свидетельств

Система DATUM (Dependability Assessment of Safety Critical Systems Through the Unification of Measurable Evidence) предлагает строгий, основанный на эмпирической базе и логических заключениях подход, позволяющий улучшить искусство оценки надежности СОТБ, комбинированием различной относящейся к делу информации, и преодолеть серьезные проблемы, имеющие место при комбинировании различных свидетельств [2].

В DATUM решается проблема оценки надежности СОТБ при наличии ошибок при проектировании аппаратного и программного обеспечения. Для решения задачи предсказания надежности в системе используются данные о нарушениях. Выделяются три фазы выполнения работы:

1. анализ (по обзорам литературы) ошибок систем "человек-компьютер" и связанного с ними риска, а также значений когнитивных факторов в вопросах надежности оператора;

2. анализ случаев инженерной практики применения СОТБ;

3. комбинирование таксономии и результатов, полученных на первой фазе, со знаниями, приобретенными во второй.

В результате создается система, использующая формализм и учитывающая такие факторы, как: профессионализм разработчиков, эффективность средств (методов), эффективность проверок и тестирования, влияние спецификаций и программных языков, специфические трудности применения или специфику проекта, испытание "подобного" продукта и пр. Отмечено, что доверие к частным системам может быть установлено из прошлого опыта на других системах.

Система DATUM позволяет помочь разработчикам определить, каким образом различные способы усовершенствования способствуют общим аргументам достоверности, и повышает возможности оценки надежности СОТБ. Разработчики и специалисты, дающие оценки, в состоянии моделировать все свои варианты допущений и предположений, рассматриваемые в случаях обеспечения надежности и безопасности. В системе DATUM делается целый ряд допущений с целью сделать оценки надежности наглядными.

Оценка безопасности опасных индустриальных процессов при дефектах проекта

Система SHIP (Assessment of the Safety of Hazardous Industrial Processes in the Presence of Design Faults) ставит задачу оценки безопасности опасных индустриальных объектов при наличии ошибок проектировщиков, и позволяет получить аргументы для формирования соответствующих БС. При этом предполагается, что имеет место организация устойчивого процесса разработки ПО и происходит накопление данных о подобных предшествующих проектах, для которых уже сделаны текущие оценки [2].

Система SHIP обеспечивает надежность, базируясь на двух основополагающих аргументах: высоком качестве проектирования (практическом отсутствии ошибок) и безошибочном тестировании.

На первом этапе построения сети доверия заполняются таблицы условных вероятностей на базе наблюдаемых ранее и зафиксированных в существующих отчетах организаций разработчиков ПО данных. Затем свидетельства каждых вновь регистрируемых в журналах неисправностей, найденных во время отладки ПО и устранения технических дефектов, последовательно вводятся в сеть доверия как апостериорные свидетельства вершин БС.

Конечная цель метода, используя приобретаемые свидетельства, добиться, чтобы показатель "вероятность неисправностей на требование" был меньше заданной величины (например 0,0001).

Принцип максимума неопределенности в инженерии ПО

Система MUSE (The Maxim of Uncertainty in Software Engineering) дает детальное представление о неопределенностях при разработке ПО и может подтверждать некоторые характеристики поведения ПО СОТБ [2]. Задачи, поставленные перед системой MUSE, исходят из допущения, что процессу создания ПО и окончательному варианту ПО свойственны неизбежные неопределенности. В системе они должны быть смоделированы и наделены управляемостью. Сложная система ПО может быть эффективно наглядно управляема посредством модели БС, так как последняя дает возможность отвечать на вопросы, касающиеся инженерии ПО, например, когда следует прекращать тестирование, где узкие места проекта, или какие части проекта являются компонентами высокого уровня риска.

Выделено четыре типа неопределенностей, на которые не может быть дан четкий ответ "да – нет", а именно неопределенности, возникающие в процессах: анализа, перехода от системных требований к проектированию и кодированию, модернизации ПО и его повторном использовании.

Можно выделить три категории источников неопределенностей в инженерии ПО, возникающих в следующих областях:

- проблемной, когда существуют противоречия между действительностью и системными допущениями о том, где может существовать риск;
- решений, где может иметь место конкуренция в скорости принятия решения;
- когда неопределенности вызываются участием человека, когда ошибки и необъяснимые факты привносятся в течение всего цикла программирования.

Анализ неопределенностей, предлагаемый системой MUSE, применим для всех трех категорий источников неопределенностей.

Система MUSE дает детальное представление о неопределенностях при тестировании ПО. Она может быть использована для подтверждения некоторых характеристик ПО, например, подтверждается увеличение доверия к факту отлаженности ПО. Система MUSE может быть развита для оценки качеств, свойств и процессов проектирования/тестирования ПО, например, удовлетворительности процессов тестирования. MUSE дает наиболее ценный и самый жесткий способ предсказания свойств оцениваемой системы с ПО за счет исследования неопределенностей в новых системных требованиях и согласованности нового проекта. Например, можно предсказать качество системы, у которой предъявляются новые системные требования.

MUSE сконструирована для решения указанных выше вопросов в ПО системы управления лифтом. Однако метод может быть развит для оценки качества и свойств ПО или процесса тестирования разработки других систем.

Заключение

Методы, основанные на БС, созданы для описания сложного аргумента, его пересмотра и модернизации в связи с новыми наблюдениями и подтверждения соответствия вероятности нарушений требуемой вероятности. БС – это формализм, хорошо известный заложенными в него возможностями к выводам из неточно определенных фактов, результатов моделирования и формализации экспертиз.

БС обеспечивает выводы в значительно более точных терминах, чем выражения типа "очень вероятно", "невероятно", "незначительное увеличение" и т.п.

Эффективность БС заключается в последовательном распространении воздействия свидетельств. Надежность системы может быть пересмотрена в тот момент, когда становится доступным новая часть свидетельств. Это означает, что определенные тенденции могут быть оценены уже на ранних стадиях разработки системы. Модель БС, описывая все возможные выводы, касающиеся доверия, может выявить некоторые типичные ошибки (заблуждений) в выводах, полученных на основании ложного понимания вероятности.

Таким образом, метод БС способствует более строгой оценке безопасности систем. В этом методе заложен большой потенциал для более легкого комбинирования и проверки аргументов безопасности, что в конечном итоге делает их более достойными доверия.

Хотя для внедрения метода БС в практику требуется значительный комплекс расчетов, и необходимы ограничения перечня проблем и причин, активно развивающаяся технология БС позволяет обозначить подходы к решению проблемы совершенствования СОТБ.

Список литературы

1. *Тулупьев А. Л. И др.* Байесовские сети. Логико-вероятностный подход. – Санкт-Петербург: Наука, 2006.
2. *Yangyang Yu, Barry W. Johnson Bayesian* "Belief Network and Its Applications", Technical Report UVA-CSCS-BBN-001, Draft, May, 20, 2002.
3. *Bouissou M., Martin F., Ourghanlian A.* "Assessment of Safety-Critical System Including Software: a Bayesian Belief Network for Evidence Sources". Paper present at the RAMS'99 (Reliability and Maintainability Symposium), Washington, January 1999).
4. *Yu, Y., Johnson, W.D.,* "Modeling COTS Systems for Safety-Critical Applications Using System Safety Standards by Bayesian Belief Networks", the 6thInternational Conference on Probabilistic Safety Assessment and Management (PSAM-6), June. 2002.
5. The SERENE Method Manual version 1.0, 1999.
6. *Robert Andrew* Weaver, The safety of software – Constructing and assuring arguments, University of York, Department of Computer Science, September 2003.
7. *N. Fenton,* Overview of the SERENE Project, Computer Science Departmtnt, Faculny of Informatics and Mathematical Scinees, Queen Mary and Westfield Colltgt, University of London, (online) http://www.des.qmul.ac.uk/~norman/SERENE_Help/start.htm (last accessed September 2003), 2000.
8. Centre for Software Reliability, SERENE- Safety and Risk Evaluation using Bayesion Nets, Centre for Software Reliability, City University, London, (online) http://www.csr.city.ac.uk/csr_city/projects/serene.html (last accessed September 2003), 2000.
9. The Serene Project Partners, Safety and Risk Evaluation using Bayesion Nets: SERENE – Final Report, TSPRITEC Project No. 22187, ERA Technology, Leatherhead, UK, 1999.
10. Hugin Expert A/S, Bayesion Network Technology Developers, (on line) http://www.hugin.com/ (last accessed September 2003), 2003.

Бабиков Василий Макарович *– канд. техн. наук,*
Панасенко Ирина Михайловна *– канд. техн. наук,*
ст. научные сотрудники института проблем управления им. В.А. Трапезникова РАН.
Контактный телефон (495) 334-93-19.