

МЕЖДУНАРОДНЫЙ
НАУЧНО-ТЕХНИЧЕСКИЙ ЖУРНАЛ

ПРОБЛЕМЫ УПРАВЛЕНИЯ И ИНФОРМАТИКИ

6'2010

**Институт кибернетики им. В.М. Глушкова
НАН Украины**

**Институт космических исследований
НАН Украины и НКА Украины**

Уважаемые читатели!

Продолжается подписка на «Международный научно-технический журнал «Проблемы управления и информатики» (старое название — «Автоматика») на 2011 год. Подписной индекс 74002. В каталогах журнал представлен под буквой «М».

Журнал является единственным в Украине периодическим изданием, более пятидесяти лет (основан в январе 1956 года) публикующим работы фундаментального и прикладного характера в широком спектре проблем автоматического управления и информатики.

Ведущие тематические разделы: проблемы динамики управляемых систем; методы идентификации и адаптивного управления; оптимальное управление и методы оптимизации; математическое моделирование и исследование сложных управляемых систем; общие проблемы исследования космоса; управление физическими объектами и техническими системами; управление и оптимизация систем с распределенными параметрами; методы управления и оценивания в условиях неопределенности; качественные методы в теории управляемых систем; методы обработки информации; космические информационные технологии и системы; технические средства для измерений и управления; космический мониторинг; экономические и управленческие системы; управление в биологических и природных системах; роботы и системы искусственного интеллекта; проблемы защиты информации. В журнале регулярно публикуются труды научных конференций и семинаров по автоматическому управлению, по перспективным космическим исследованиям.

Журнал издается при творческом участии

Украинской Ассоциации по автоматическому управлению; Национального космического агентства Украины; академических и отраслевых научных учреждений; ведущих вузов Украины и стран СНГ; ученых и специалистов стран дальнего зарубежья.

Главный редактор журнала — почетный директор Института космических исследований НАН Украины и НКА Украины, Президент Украинской ассоциации по автоматическому управлению академик НАН Украины КУНЦЕВИЧ Всеволод Михайлович.

Международный научно-технический журнал «Проблемы управления и информатики» в полном объеме переиздается на английском языке с 1968 года, в настоящее время издательской фирмой Begell House, Inc. (США) под названием «Journal of Automation and Information Sciences» (www.begellhouse.com).

Журнал включен в перечень профильных изданий ВАК Украины; регулярно реферируется в отечественных и зарубежных изданиях: «Джерело» (Украина, Киев), «Реферативный журнал» (Россия, Москва, ВИНТИ), «Математическое ревью» («Math. Review», США, Американское математическое общество), «Прикладная механика» («Applied mechanics Reviews», США, Американское общество инженеров-механиков); журнал введен в базу данных американского Института научной информатики Томсона (ISI), в реферативную базу данных Scopus.

Информация о журнале включена в:

Каталог периодических изданий Украины по подписке на 2011 год; Каталог Агентства «Роспечать» на 2011 год; Сводный каталог периодических изданий академий наук — членов МААН на 2011 год; Экспортный каталог на 2011 год.

Если вы не успели оформить подписку, журнал можно приобрести непосредственно в редакции. Наш адрес:

03680 ГСП Киев 187, проспект Академика Глушкова, 40, корпус 4/1,
Институт космических исследований НАН Украины и НКА Украины,
редакция Международного научно-технического журнала «Проблемы управления и информатики», комн. 106;

тел. 526-22-29, 522-58-46. E-mail: red@nonnared.kiev.ua

Редакторы: Л.И. Лесько, О.И. Жудра, Е.И. Мезенцева, И.И. Баданина
Компьютерная группа: Г.В. Зорько, Т.В. Иванова, А.А. Туроверов

Подл. в печ. 04.10.2010. Формат 70×108/16. Бум. офсет. Гарнитура Таймс. Офс. печать. Усл. печ. л. 14.0.
Усл. кр.-отт. 14.71. Уч.-изд. л. 12. Тираж 250 экз. Заказ 79. Цена 18 грн. (Украина), 5 у.е. (Россия и страны СНГ)
Свидетельство о государственной регистрации КВ № 15605-4077 ПР от 09.09.2009

Напечатано на полиграфическом участке
Института кибернетики им. В.М. Глушкова НАН Украины
03680 Киев 187, проспект Академика Глушкова, 40

СОДЕРЖАНИЕ

Методы идентификации и адаптивного управления

| | |
|---|----|
| <i>Сергиенко И.В., Дейнека В.С.</i> Идентификация градиентными методами параметров задач диффузии вещества в нанопористой среде | 5 |
| <i>Сальников Н.Н.</i> Точность оценивания параметров линейной регрессии при погрешностях в переменных | 19 |
| <i>Скобелев В.Г.</i> Восстановление вектора начального состояния нелинейных автоматов над конечным кольцом | 31 |

Математическое моделирование и исследование сложных управляемых систем

| | |
|---|----|
| <i>Булавацкий В.М., Скопецкий В.В.</i> Об одной неизотермической консолидационной математической модели геоинформатики | 35 |
| <i>Булат А.Ф., Киселева Е.М., Пичугов С.А., Блюсс О.Б.</i> Модификация метода с-средних на основе нелинейного векторного критерия | 46 |
| <i>Спиридонов В.Н.</i> Метод моделирования дискретных динамических систем с постоянной структурой | 55 |

Методы обработки информации

| | |
|---|----|
| <i>Бидюк П.И., Давиденко В.И., Трофименко Д.В., Терентьев А.Н.</i> Сравнительный анализ методов оценки взаимосвязи между вершинами при построении байесовских сетей | 64 |
| <i>Иванов В.Г., Любарский М.Г., Ломоносов Ю.В.</i> Сжатие изображения текста на основе выделения символов и их классификации | 74 |
| <i>Вагис А.А.</i> Байесовские процедуры распознавания на сетях | 85 |

Космические информационные технологии и системы

| | |
|---|----|
| <i>Ефименко Н.В.</i> Магнитная система ориентации и стабилизации космического аппарата «Egyptsat-1» | 91 |
|---|----|

Космический мониторинг

| | |
|--|----|
| <i>Куссуль Н.Н., Соколов Б.В., Зельк Я.И., Зеленцов В.А., Скакун С.В., Шелестов А.Ю.</i> Оценка рисков стихийных бедствий на основе разнородной геопространственной информации | 97 |
|--|----|

Экономические и управленческие системы

| | |
|---|-----|
| <i>Бойчук М.В., Шмурыгина Н.М.</i> Моделирование динамической обобщенной эколого-экономической системы межотраслевого баланса с запаздыванием | 111 |
| <i>Галушко В.Г.</i> Сравнение двух стратегий планирования перевозок скоропортящихся грузов | 131 |

Проблемы защиты информации

| | |
|--|-----|
| <i>Кошкина Н.В.</i> Метод встраивания цифровых водяных знаков в аудиосигналы на основе вейвлет- и Фурье-преобразований | 134 |
|--|-----|

| | |
|---|-----|
| Содержание «Международного научно-технического журнала «Проблемы управления и информатики» за 2010 год | 148 |
|---|-----|

| | |
|---|-----|
| Содержание «Journal of Automation and Information Sciences», Т. 41, №№ 1–12, 2009 г. | 154 |
|---|-----|

| | |
|-------------------------------------|-----|
| Информация для авторов | 160 |
|-------------------------------------|-----|

УДК 62-50

П.И. Бидюк, В.И. Давиденко, Д.В. Трофименко, А.Н. Терентьев

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ ОЦЕНКИ ВЗАИМОСВЯЗИ МЕЖДУ ВЕРШИНАМИ ПРИ ПОСТРОЕНИИ БАЙЕСОВСКИХ СЕТЕЙ

Введение

Развитие и всестороннее применение вычислительной техники приводит к широкому использованию средств цифровой обработки данных почти во всех сферах деятельности человека. Полученная информация не может использоваться в исходном виде и требует некоторой обработки. Именно поэтому популярны методы интеллектуального анализа данных (data mining), которые позволяют находить закономерности в данных, описывающих процессы различной природы. Одной из технологий интеллектуального анализа данных являются байесовские сети (БС) (Bayesian networks). Они дают возможность определять причинно-следственные связи между факторами и получать вероятности возникновения той или иной ситуации при изменении состояния любого узла сети.

Методика построения БС состоит из двух основных задач: построение оптимальной топологии БС, т.е. нахождение направленного ациклического графа, который лучше всего соответствует эмпирическим обучающим данным; вычисление значений таблиц условных вероятностей для всех вершин этого графа.

1. Постановка задачи

Данная статья посвящена сравнительному анализу методов оценки взаимосвязи между вершинами при нахождении оптимальной структуры БС с использованием эвристического алгоритма.

Эвристический алгоритм построения БС [1] заключается в следующем:

- 1) построение упорядоченного множества вершин;
- 2) выполнение целенаправленного поиска с использованием оценочной функции на основе принципа описания минимальной длины;
- 3) повторение итераций до получения структуры сети заданного качества.

В данном алгоритме построение упорядоченного множества вершин — один из ключевых моментов, поскольку наличие такого множества существенно сокращает пространство возможных нециклических структур. Однако в этом случае появляется новая нетривиальная задача: как по множеству обучающих данных получить упорядоченное множество вершин сети. Самый очевидный способ — привлечение экспертов. Однако может возникнуть необходимость моделирования данных такой предметной области, в которой квалифицированных экспертов нет. Именно поэтому актуальна задача автоматического построения упорядоченного множества вершин и, как следствие, выбора эффективного метода оценивания степени зависимости между факторами или произвольными случайными переменными.

2. Методы оценки взаимосвязи между факторами

Рассматриваются следующие методы оценивания: меры на основании статистики χ^2 -квадрат (коэффициенты Пирсона, Крамера, Чупрова), значение обоюдной информации и коэффициент Гудмана.

Коэффициенты связи, основанные на статистике χ^2 . Поскольку критерий χ^2 служит основой для проверки независимости, стоит рассмотреть возможность использования значений χ^2 как меру связи.

При рассмотрении коэффициентов связи, основанных на значении χ^2 , исходим из предпосылки о том, что чем больше объем выборки N , тем легче получить статистически значимую величину критерия даже при очень слабой взаимосвязи переменных (т.е. при больших объемах выборки даже слабые связи будут статистически значимыми) [2]. Чтобы элиминировать влияние объема выборки N , К. Пирсон предложил в качестве меры связи коэффициент среднеквадратической сопряженности:

$$\varphi^2 = \frac{\chi^2}{N}.$$

Значения χ^2 рассчитываются на основе таблицы частот взаимного сопряжения переменных $X^{(1)}$ и $X^{(2)}$ (когда переменные сети имеют произвольное число состояний m и n соответственно):

$$\chi^2 = N \left(\sum_{i=1}^m \sum_{j=1}^n \left(\frac{k_{ji}}{\sum_{i=1}^m k_{ji} \cdot \sum_{j=1}^n k_{ji}} \right) - 1 \right), \quad N = m + n.$$

В табл. 1 приведены значения частот взаимного сопряжения переменных $X^{(1)}$ и $X^{(2)}$, где $x_q^{(p)}$ — q -е состояние p -й переменной; k_{ji} — частота взаимного сопряжения j -го состояния переменной $X^{(1)}$ с i -м состоянием переменной $X^{(2)}$.

Таблица 1

| Переменные | $X^{(1)} = x_1^{(1)}$ | ... | $X^{(1)} = x_n^{(1)}$ | Сумма |
|-----------------------|-----------------------|-----|-----------------------|------------------------------------|
| $X^{(2)} = x_1^{(2)}$ | k_{11} | ... | k_{n1} | $\sum_{j=1}^n k_{j1}$ |
| ... | ... | ... | ... | ... |
| $X^{(2)} = x_m^{(2)}$ | k_{1m} | ... | k_{nm} | $\sum_{j=1}^n k_{jm}$ |
| Сумма | $\sum_{i=1}^m k_{1i}$ | ... | $\sum_{i=1}^m k_{ni}$ | $\sum_{i=1}^m \sum_{j=1}^n k_{ji}$ |

Стремясь нормировать меру связи к единому диапазону, С. Крамер видоизменил формулу коэффициента среднеквадратической сопряженности для своего коэффициента Крамера [3]: $V^2 = \frac{\varphi^2}{\min(n-1, m-1)}$.

А. Чупров [2] предложил похожую формулу — полихорический коэффициент сопряженности (polychoric correlation coefficient) — коэффициент Чупрова:

$$T^2 = \frac{\varphi^2}{\sqrt{(n-1)(m-1)}}.$$

Нетрудно заметить, что T^2 и V^2 эквивалентны, когда число столбцов равно числу строк, в иных случаях V^2 всегда больше T^2 . Для таблицы сопряженности 2×2 обе меры равны φ^2 .

Коэффициент контингенции Пирсона [4] имеет вид: $C^2 = \frac{\varphi^2}{1 + \varphi^2}.$

Перечисленные коэффициенты, основанные на χ^2 , принимают область допустимых значений от нуля до единицы. Значение ноль говорит о том, что переменные независимы, а единица означает, что значение одной переменной можно точно спрогнозировать по другой, т.е. случай полной линейной зависимости.

Информационный подход к оценке связи. Традиционные меры связи, основанные на χ^2 , представляют собой сугубо эвристические конструкции. Но интерпретация и математико-статистическое обоснование не являются совершенными. Поэтому стоит также рассмотреть оценки мер связи, в основе которых лежит теоретически-информационный подход к анализу таблиц частот.

Современная трактовка статистической связи между переменными A и B сводится к оценке количества информации:

$$I(A, B) = H(B) - H(B | A),$$

которое устраняет неопределенность того, какое значение примет B , если известно значение A . Таким образом, мера сопряженности оценивается как доля сопряженного разнообразия величины B , включенной в систему (A, B) , по сравнению с разнообразием (энтропией) B , рассматриваемой отдельно:

$$R_{B|A} = I(A, B) / H(B).$$

Практическая трудность построения мер, относящихся к этому семейству, заключается в сложности выбора конкретных дефиниций и формальных выражений понятий «неопределенность» и «информация» из множества возможных

Л. Гудман и В. Крускал (L. Goodman, W. Kruscal) в своем принципе пропорциональной предикации [5] выразили мнение, что мерой связи должно служить относительное уменьшение вероятности ошибки предсказания значения зависимого признака по значению независимого. Если в рамках этой концепции оценивать относительную ошибку, используя отношения правдоподобия и минимизацию числа ошибочных предсказаний, получим меру связи, известную как коэффициент Гудмана:

$$\lambda_{B|A} = \frac{\sum_i p_{i\cdot} \cdot \max_j p_{j|i} - \max_j p_{\cdot j}}{1 - \max_j p_{\cdot j}},$$

где $\max_j p_{\bullet j}$ — вероятность, соответствующая модальной категории маргинального распределения B ; $\max_j p_{j|i}$ — вероятность модального значения B при i -м значении A .

Значения λ лежат в пределах от 0 до 1: ноль означает невозможность предсказания зависимого признака, а 1 означает, что по значению независимого признака можно уверенно предсказать значение зависимого признака. Необходимо отметить, что нулевое значение λ эквивалентно не состоянию статистической независимости, а тождеству модальных категорий переменных A и B .

Кроме несимметричных мер λ в [2] предложен симметричный коэффициент Гудмана:

$$\lambda_{A,B} = \frac{\left[\sum_i p_i \cdot \max_j p_{j|i} - \max_j p_{\bullet j} + \sum_j p_j \cdot \max_i p_{i|j} - \max_i p_{i\bullet} \right] / 2}{1 - \left(\max_i p_{i\bullet} + \max_j p_{\bullet j} \right) / 2}.$$

Значение обоюдной информации (ЗОИ). Для оценки степени зависимости двух произвольных переменных x^i и x^j в работе [6] Шоу и Лью в 1968 г. впервые предложили использовать значение обоюдной информации (mutual information) $MI(x^i, x^j)$. Для расчета они предложили следующее выражение:

$$MI(x^i, x^j) = \sum_{x^i, x^j} P(x^i, x^j) \cdot \log \left(\frac{P(x^i, x^j)}{P(x^i) \cdot P(x^j)} \right),$$

где $P(x^i, x^j)$ — значение совместной вероятности между x^i и x^j , $P(x^i)$ и $P(x^j)$ — значения вероятностей $P(x^j)$ соответственно.

По своей сути значение обоюдной информации — аналог корреляции, но по своему содержанию это оценка количества информации о переменной x^j , содержащейся в переменной x^i . Данный показатель измеряет, насколько знание одной переменной уменьшает неопределенность другой. ЗОИ принимает неотрицательные значения $MI(x^i, x^j) \geq 0$. Если вершины x^i и x^j полностью независимы одна от другой, $MI(x^i, x^j) = 0$, так как $P(x^i, x^j) = P(x^i) \cdot P(x^j)$, следовательно,

$$\log \left(\frac{P(x^i, x^j)}{P(x^i) \cdot P(x^j)} \right) = \log \left(\frac{P(x^i) \cdot P(x^j)}{P(x^i) \cdot P(x^j)} \right) = \log(1) = 0.$$

3. Оценивание качества построения байесовских сетей

При проведении вычислительных экспериментов для оценивания качества построения БС используется число лишних, отсутствующих и реверсивных дуг в полученной БС сравнительно с оригинальной БС, а также структурная разница или перекрестная энтропия между полученной и оригинальной БС.

Структурная разница. Для вычисления структурной разницы используют формулу симметричной разницы структур [7]:

$$\delta = \sum_{i=1}^n \delta_i = \sum_{i=1}^n \text{card}(\Pi^{(i)}(B) \Delta \Pi^{(i)}(A)) =$$

$$= \sum_{i=1}^n \text{card}((\Pi^{(i)}(B) \setminus \Pi^{(i)}(A)) \cup (\Pi^{(i)}(A) \setminus \Pi^{(i)}(B))),$$

где B — полученная БС; A — оригинальная БС; n — количество вершин сети; $\Pi^{(i)}(B)$ — множество предков i -й вершины, полученной B ; $\Pi^{(i)}(A)$ — множество предков i -й вершины оригинальной сети A ; $\text{card}(\xi)$ — мощность конечного множества ξ , которое определяется как количество элементов, принадлежащих множеству ξ .

Перекрестная энтропия. Перекрестная энтропия — это расстояние между распределением построенной и оригинальной БС [8]. Пусть $p(v)$ — совместное распределение оригинальной БС, $q(v)$ — совместное распределение построенной БС. Тогда перекрестная энтропия вычисляется так:

$$H(p, q) = \sum_v p(v) \cdot \log \frac{p(v)}{q(v)} =$$

$$= \sum_{j \in J} \sum_{s \in S(j, g)} \sum_{a \in A^{(j)}} p(X^{(j)} = a | \Pi^{(j)} = s) \cdot \log \frac{p(X^{(j)} = a | \Pi^{(j)} = s)}{q(X^{(j)} = a | \Pi^{(j)} = s)}.$$

Интегральный показатель. Для оценивания качества построения БС разработан и реализован интегральный показатель, который учитывает число лишних, отсутствующих и реверсивных дуг в построенной БС при сравнении с оригинальной БС.

Интегральный показатель вычисляется по формуле

$$R = M_{\text{add}} + M_{\text{del}} + 0,5M_{\text{dir}},$$

где M_{add} — количество добавленных связей; M_{del} — количество удаленных связей; M_{dir} — количество связей с измененным направлением (в формуле этот показатель имеет коэффициент 0,5, который указывает на меньшую критичность измененного направления по сравнению с отсутствием дуги или наличием лишних).

4. Описание вычислительных экспериментов

Цель экспериментов — сравнение эффективности использования разных метрик при построении БС эвристическим методом. Для проведения экспериментов применялось два метода: генерирование случайным образом выборки по заданной структуре исходной сети и непосредственное применение метрик при построении БС по реальным эмпирическим данным. Также в рамках данных экспериментов исследовались возможности использования способа псевдослучайной генерации сетей для поиска взаимосвязей между размером выборки, количеством вершин в сети, количеством состояний в каждой вершине и качеством восстановления сети.

Классические БС. Первый эксперимент заключался в использовании разных методов оценки взаимосвязи между вершинами при построении БС по реальным эмпирическим данным.

Рассматривались следующие сети: Asia, CarStarts и Alarm. Они представляют собой классические байесовские сети и используются исследователями для тестирования новых и существующих алгоритмов построения сетей. Данные сети входят в стандартный набор примеров почти всех более-менее известных программных продуктов для работы с БС.

Сеть Asia состоит из 8 вершин и 8 дуг. Размер выборки 7200 обучающих записей.

Сеть Car Starts имеет 18 вершин и 17 дуг. Размер выборки 10 тыс. обучающих записей.

Сеть Alarm имеет 37 вершин и 46 дуг. Размер выборки 10 тыс. обучающих записей (выборка сгенерирована).

Значения разных метрик оценки взаимосвязи между вершинами при построении данных сетей приведены в табл. 2.

Таблица 2

| Сеть | Метрики | M_{add} | M_{del} | M_{dir} | R | δ |
|------------|-----------------|-----------|-----------|-----------|------|----------|
| Car Starts | ЗОИ | 1 | 2 | 3 | 4,5 | 9 |
| | Метрика Пирсона | 1 | 2 | 3 | 4,5 | 9 |
| | Метрика Чупрова | 1 | 2 | 3 | 4,5 | 9 |
| | Метрика Крамера | 1 | 2 | 3 | 4,5 | 9 |
| | Метрика Гудмана | 1 | 2 | 3 | 4,5 | 9 |
| Asia | ЗОИ | 0 | 0 | 3 | 1,5 | 6 |
| | Метрика Пирсона | 0 | 0 | 3 | 1,5 | 6 |
| | Метрика Чупрова | 0 | 0 | 3 | 1,5 | 6 |
| | Метрика Крамера | 0 | 0 | 3 | 1,5 | 6 |
| | Метрика Гудмана | 0 | 1 | 3 | 2,5 | 7 |
| Alarm | ЗОИ | 12 | 2 | 22 | 25 | 58 |
| | Метрика Пирсона | 12 | 2 | 13 | 20,5 | 40 |
| | Метрика Чупрова | 14 | 3 | 14 | 24 | 45 |
| | Метрика Крамера | 15 | 2 | 13 | 23,5 | 43 |
| | Метрика Гудмана | 16 | 3 | 16 | 27 | 51 |

Псевдослучайная генерация байесовской сети. Второй эксперимент заключался в генерировании по заданной структуре исходной сети псевдослучайным образом.

Эксперимент можно разделить на четыре этапа:

- 1) генерация сети случайным образом;
- 2) генерация выборки;
- 3) построение структуры сети по выборке;
- 4) сравнение исходной и восстановленной сетей.

Генерация сети случайным образом. Для того чтобы исключить изолированные вершины, первым этапом было объединение всех вершин в цепь. При этом направление дуги выбирается случайным образом. Количество связей, построенных таким образом, равно $N - 1$.

Оставшиеся $M - N + 1$ связей строятся путем выбора случайным образом родительской и дочерней вершин. При добавлении связи осуществляется проверка на цикличность. Если новая связь образует цикл, то она отбрасывается и избираются новые родительская и дочерняя вершины.

Таблицы условных вероятностей задаются путем заполнения их случайными числами в диапазоне от 0 до 1 со следующей нормализацией (приведение суммы вероятностей по строке до 1).

Генерация выборки. Генерация выборки происходит следующим образом. Сначала делается вероятностный вывод в сети без инстанцированных вершин $P(S_{ij})$, $i = 1, \dots, N$; $j = 1, \dots, S$, затем выбирается вершина N_{i^*} и инстанцируется одно из ее состояний $S_{i^*j^*}$ с вероятностью этого состояния $P(S_{i^*j^*})$. Пересчитываются вероятности состояний вершин после инстанцирования $P(S_{ij} | S_{i^*j^*})$. Далее выбирается следующая вершина. Эта операция повторяется пока не останется неинстанцированных вершин. Инстанцированные состояния образуют запись в выборке $(S_{i_1}, \dots, S_{i_N})$, $i_k \in (1, \dots, S)$. Алгоритм повторяется до тех пор, пока не будет создано необходимого количества записей V в выборке.

Построение структуры сети по выборке. На основе полученной выборки строится сеть с использованием выбранных метрик.

Сравнение исходной и восстановленной сетей. Вычисляются показатели, характеризующие отличия восстановленной сети от исходной, а также оценивается качество восстановления.

Проведение эксперимента: количество вершин (N): 10, количество связей (M): 15, количество состояний вершины (S): 2, размер выборки (V): 1000, количество итераций (I): 100.

Для уменьшения статистической ошибки эксперимент был повторен I раз для каждой метрики и были найдены средние значения показателей (табл. 3):

$$\bar{R} = \frac{\sum_{i=1, I} R_i}{I}, \bar{M}_{add} = \frac{\sum_{i=1, I} M_{add_i}}{I},$$

$$\bar{M}_{del} = \frac{\sum_{i=1, I} M_{del_i}}{I}, \bar{M}_{dir} = \frac{\sum_{i=1, I} M_{dir_i}}{I}, \bar{\delta} = \frac{\sum_{i=1, I} \delta_i}{I},$$

где R_i , M_{add_i} , M_{del_i} , M_{dir_i} , δ_i — значения показателей на i -й итерации.

Таблица 3

| Метрики | \bar{M}_{add} | \bar{M}_{del} | \bar{M}_{dir} | \bar{R} | $\bar{\delta}$ |
|-----------------|-----------------|-----------------|-----------------|-----------|----------------|
| ЗОИ | 0,96 | 4,4 | 4,13 | 7,425 | 13,62 |
| Метрика Пирсона | 0,97 | 4,41 | 4,12 | 7,44 | 13,62 |
| Метрика Чупрова | 0,97 | 4,41 | 4,12 | 7,44 | 13,62 |
| Метрика Крамера | 0,97 | 4,41 | 4,12 | 7,44 | 13,62 |
| Метрика Гудмана | 1,15 | 6,25 | 3,51 | 9,155 | 14,42 |

Связь между параметрами байесовской сети и качеством ее восстановления. Цель эксперимента — определить характер зависимости качества восстановления сети от размера выборки. Для этого генерировалась сеть с заданными параметрами (количество узлов, связей, состояний). На основании этой сети генерировались выборки различного объема. С помощью сгенерированных данных с использованием эвристического алгоритма с метрикой в виде обоюдной информации сеть была восстановлена. Качество восстановления оценивалось инте-

гральным показателем. Для получения более качественных результатов для заданного набора параметров сеть генерировалась несколько раз и рассчитывалось среднее значение интегрального показателя.

На рис. 1–3 показаны результаты экспериментов для сетей с различными параметрами. Количество повторений экспериментов для каждой ситуации, заданной соответствующим набором параметров, равнялось 10. Для отдельных наборов эксперимент был повторен 100 раз, ошибка по сравнению с экспериментами, состоящими из 10 повторов моделирования ситуаций, не превысила 5 %.



Рис. 1



Рис. 2



Рис. 3

Исходя из графиков, можно сделать ряд выводов. Во-первых, зависимость качества восстановления сети от размера выборки имеет характерную форму и может быть представлена в виде функции. Параметры функции также будут зависеть от параметров сети (количество вершин, состояний, связей и т.д.). Нужно отметить, что зависимость между ошибкой восстановления и размером выборки имеет обратно пропорциональный характер, т.е. значение ошибки уменьшается при увеличении выборки.

Во-вторых, видно, что при определенном объеме выборки дальнейшее ее увеличение не несет заметного улучшения качества восстановления. Таким образом, существует некий порог размера выборки, превышение которого нецелесообразно. На основании полученных результатов проведенных 630 вычислительных экспериментов для определения размера обучающей выборки, обеспечивающей высокую точность построения топологии БС (состоящей не более чем из 50 вершин), авторы предлагают функцию вида

$$V(N, S, M) = 100 \cdot N \cdot S \cdot M.$$

Заключение

В настоящей статье рассмотрена проблема построения упорядоченного множества вершин как одного из этапов эвристического алгоритма построения БС. Для решения поставленной задачи предложено использование разных методик оценки взаимосвязей между вершинами.

Описанные методы оценки взаимосвязей можно разделить на две группы: исключительно эвристические метрики, основанные на статистике χ^2 -квадрат, и информационные метрики связи.

Из результатов проведенных вычислительных экспериментов видно, что использование метрик обоюдной информации, Пирсона, Чупрова и Крамера дают фактически идентичные результаты. Использование коэффициента Гудмана ведет к ухудшению результата, в сравнении с вышеупомянутыми. Таким образом, в качестве оценки связи в эвристическом алгоритме рекомендуется использовать значение обоюдной информации, коэффициенты Пирсона, Чупрова или Крамера.

Авторами предложена функция для определения размера обучающей выборки, необходимой для построения топологии БС.

В дальнейших исследованиях планируется разработать методы улучшения качества алгоритмов построения БС, в частности, на основании полученных результатов.

П.І. Бідюк, В.І. Давиденко, Д.В. Трофименко, О.М. Терентьєв

ПОРІВНЯЛЬНИЙ АНАЛІЗ МЕТОДІВ ОЦІНКИ ВЗАЄМОЗВ'ЯЗКУ МІЖ ВЕРШИНАМИ ПРИ ПОБУДОВІ МЕРЕЖ БАЙЕСА

Проаналізовано методи оцінки взаємозв'язку між вершинами при побудові мереж Байеса з використанням евристичного алгоритму. Наведено теоретичне обґрунтування методів, розглянуто результати їх практичного використання при побудові класичних мереж, а також описано алгоритм проведення експериментів на основі псевдовипадкового генерування байесових мереж. Отримані для кожного методу результати порівнювались. Зроблено висновки про можливість застосування розглянутих методів оцінки взаємозв'язку між вершинами при побудові байесових мереж.

COMPARATIVE ANALYSIS OF ESTIMATION METHODS OF NODES CORRELATION IN BAYESIAN NETWORKS CONSTRUCTION

The estimation methods of nodes interrelation in Bayesian networks construction using heuristic algorithm are analysed. Theoretical justification of methods is given, results of practical use in construction of classic networks are considered and experiments implementation algorithm based on pseudorandom generation of Bayesian networks is described. Comparison of received results for each method is performed and conclusions are drawn about possibility of using the considered estimation methods of nodes correlation in Bayesian networks construction.

1. Терентьев А.Н., Бидюк П.И. Эвристический метод построения байесовских сетей // Математические машины и системы. — Киев : ТОВ «РВПК ЕКСЛИБРИС», 2006. — № 3. — С. 12–23.
2. Елисеева И.И., Рукавишников В.О. Группировка, корреляция, распознавание образов (Статистические методы классификации и измерения связей). — М. : Статистика, 1977. — 144 с.
3. Cramer H. Mathematical methods of statistics. — Princeton. — NJ : Princeton U-t Press, 1946. — 575 p.
4. Liebetrau A.M. Measures of association. — Newbury Park, CA : Sage Publications // Quantitative Appl. in the Social Sci., 1983, Ser. 32.
5. Goodman L.A., Kruskal W.H. Measures of association for cross-classifications // J. Amer. Statist. Assoc. — 1954. — 49. — P. 732–764.
6. Chow C.K., Liu C.N. Approximating discrete probability distributions with dependence trees // IEE Trans. on Inform. Theory. — 1968. — IT-14, N 3. — 6 p.
7. Zheng Y., Kwoh C.K. Improved MDL score for learning of Bayesian networks // Proc. of the Intern. Conf. on Artificial Intell. in Sci. and Technology (AISAT 2004). — 2004. — P. 98–103.
8. Heckerman D., Geiger D., Chickering D.M. Learning Bayesian networks: the combination of knowledge and statistical data // Machine Learning. — 1995. — 20, N 3. — P. 197–243.
9. Spiegel M.R., Stephens L.J. Schaum's outline of theory and problems of statistics. — McGraw-Hill Professional, 1999. — P. 345–349.
10. Терентьев А.Н., Бидюк П.И., Миронова А.В., Медин Н.Ю. Сравнение методов интеллектуального анализа данных при оценивании кредитоспособности физических лиц // Проблемы управления и информатики. — 2009. — № 5. — С. 141–149.
11. Терентьев А.Н., Бидюк П.И., Коршевнюк Л.А. Байесовская сеть — инструмент интеллектуального анализа данных // Там же. — 2007. — № 4. — С. 83–92.
12. Афифи А., Эйзен С. Статистический анализ: Подход с использованием ЭВМ: Пер. с англ. — М. : Мир, 1982. — 488 с.

Получено 13.05.2010

ВНИМАНИЮ АВТОРОВ ЖУРНАЛА!

ТРЕБОВАНИЯ К РУКОПИСЯМ СТАТЕЙ:

Материал статьи должен быть оригинальным, т.е. не быть опубликованным ранее или представленным к опубликованию в другом издании.

Рукопись представляется на белой бумаге в двух экземплярах, *аннотация* (5–6 строк с указанием УДК, ф.и.о. автора, названия работы) — на украинском, русском и английском языках, экспертное заключение о возможности опубликования. На отдельной странице приводятся сведения об авторе: ф.и.о., почтовый и электронный адреса, телефон, факс, место работы, научное звание, ученая степень, должность. Необходимо указать, с кем из авторов вести переписку. На последней странице статьи каждый автор ставит свою подпись.

К рукописи статьи обязательно прилагается согласие авторов на публикацию (см. образец в журнале).

На первой странице рукописи желательно указать *раздел журнала* (см. четвертую страницу обложки), в котором должна быть опубликована статья.

Текст следует печатать через полтора интервала. *Поля* должны быть достаточно широкими, страницы пронумерованы. Материал публикации может быть разбит на подразделы не более двух уровней.

Таблицы и рисунки представляются на отдельных листах.

Ссылки на литературные источники нумеруются последовательно (в квадратных скобках) по тексту. *Список литературы* приводится в конце статьи.

На отдельном листе прилагается *перечень всех специальных символов*, греческих и готических букв, используемых в статье, например: греческие буквы γ — гамма, Ω — омега; готические \mathfrak{Z} — йот; специальные символы \int — интеграл, \cup — объединить, \cap — включить.

Все статьи проходят закрытое рецензирование и, в случае необходимости, могут быть возвращены автору на доработку.

После утверждения редколлегией статьи к печати автор представляет в редакцию текст статьи и аннотаций в электронном виде (файл формата *.doc, *.rtf). Получив верстку статьи, согласовывает редакторскую правку и срочно отправляет верстку в редакцию. В случае значительного изменения автором отредактированного текста статья будет снята с рассмотрения на предмет публикации.

Редколлегия не сообщает мотивов отказа в опубликовании статей и оставляет за собой право не возвращать рукописи.

Статьи, присланные в электронном виде, не рассматриваются.

Рукописи следует направлять в редакцию журнала по адресу:

03680 ГСП Киев 187, проспект Академика Глушкова, 40, корпус 4/1,

Институт космических исследований НАНУ и НКАУ, редакция журнала (комн. 106).

Справки по телефону 526 22 29. E-mail: red@nonnared.kiev.ua.

ТРЕБОВАНИЯ К ТЕКСТОВОМУ ФАЙЛУ:

Текст статьи и аннотации набираются с помощью текстового редактора Microsoft Word.

Используемые стили: шрифт Times New Roman, высота 12 пт., межстрочное расстояние 1,5. Формат бумаги А4, поля (слева, справа, сверху, снизу) 2 см.

Все переменные и формулы набираются в редакторе формул Microsoft Equation Editor 3.0 в масштабе 100×100. Опции редактора формул — по умолчанию. Ширина формул не более 12 см. Скобки из шаблона редактора формул применяются для знаков:

$\left(\frac{a}{b}, \Pi, \cup, \Sigma, \int\right)$. Для переменных, в том числе с индексами, применяются скобки из

клавиатурного набора, например (x_i^j, l_i^b) . Знак транспонирования — Т.

В *формулах и рисунках* все переменные и индексы обозначаются латинскими буквами.

Нумерация формул должна быть *сквозная*, т.е. не следует отдельно нумеровать формулы каждого раздела статьи или вводить двойную нумерацию.

Рисунки (не более 10) создаются встроенным редактором рисунков Word Picture либо иными Windows-приложениями (в этом случае рисунки необходимо представлять отдельными файлами формата tiff с разрешением 300–600 dpi или eps).

Таблицы выполняются стандартным встроенным в Word инструментарием «Таблица».

Помните, что никакие расхождения текста файла с текстом статьи не допускаются.

Редакция

**ЖУРНАЛ ПРИНИМАЕТ ДЛЯ ПУБЛИКАЦИИ РАБОТЫ
ПО СЛЕДУЮЩИМ НАУЧНЫМ НАПРАВЛЕНИЯМ:**

проблемы динамики управляемых систем;
методы идентификации и адаптивного управления;
оптимальное управление и методы оптимизации;
математическое моделирование и исследование сложных
управляемых систем;
общие проблемы исследования космоса;
управление физическими объектами и техническими системами;
управление и оптимизация систем с распределенными параметрами;
методы управления и оценивания в условиях неопределенности;
качественные методы в теории управляемых систем;
методы обработки информации;
космические информационные технологии и системы;
технические средства для измерений и управления;
космический мониторинг;
экономические и управленческие системы;
управление в биологических и природных системах;
роботы и системы искусственного интеллекта;
проблемы защиты информации.

Журнал в полном объеме переиздается на английском языке издательской
фирмой BEGELL HOUSE, INC. (США).

Рукописи следует направлять в редакцию журнала по адресу:

03680, Киев 187, пр. Академика Глушкова, 40, корпус 4/1,
Институт космических исследований НАН Украины и НКА Украины.

Справки по телефону 526 22 29.