

*Рассмотрены ключевые вопросы компьютерной реализации информационной технологии вероятностного моделирования состояний сложных систем на базе нечетких сетей доверия. Конструктивно изложен математический аппарат, с помощью которого осуществляется анализ нечетких сетей доверия и прогнозирования состояний исследуемых систем в условиях неопределенной информации. Значительное внимание уделено методологическим, архитектурным и функциональным вопросам компьютеризации.*

© И.Н. Парасюк,  
Ф.В. Костукевич, 2011

УДК 681.3.06

И.Н. ПАРАСЮК, Ф.В. КОСТУКЕВИЧ

## **О РЕАЛИЗАЦИИ ИНФОРМАЦИОННОЙ ТЕХНОЛОГИИ ВЕРОЯТНОСТНОГО МОДЕЛИРОВАНИЯ СОСТОЯНИЙ СЛОЖНЫХ СИСТЕМ НА БАЗЕ НЕЧЕТКИХ СЕТЕЙ ДОВЕРИЯ**

**Введение.** Сети доверия, известные также как Байесовские сети (БС), широко используются на практике для вероятностного моделирования состояний сложных систем в условиях неопределенности причинно-следственных отношений между их компонентами [1]. Построение более адекватных моделей исследуемых систем и получение более информативных результатов моделирования возможно, если воспользоваться методами теории нечетких множеств [2]. Положительный опыт такого подхода имеется, например в [3]. Уместно отметить, что трудности, возникающие на этом пути, – пути системного учета принципиально различных видов неопределенностей в процессах моделирования, в основном, сопряжены с усложнением не только самих математических моделей, но и усложнением всего математического аппарата оценивания и механизмов вывода, простоты поиска решений и др.

В статье, обобщая понятие лингвистической вероятности [2] и нечетких чисел [4], введено понятие нечетких потенциалов и операции над ними, осуществлена имплементация этих понятий в математический аппарат вероятностного вывода, процедур оценивания и прогнозирования состояний исследуемых систем, описаны аспекты компьютерной реализации соответствующей информационной технологии.

**Математические основы анализа нечетких сетей доверия.** Введем основные понятия и обозначения. Переменные и их множества обозначим большими буквами, а их отдельные значения – малыми. Запись  $X = x$  значит, что переменная  $X$  получает значение  $x$ . Область определения (или пространство состояний) для  $X$  обозначим  $\text{dom}(X)$ ;  $\|X\|=|\text{dom}(X)|$  определяет количество возможных разных значений переменной  $X$ .

$n$ -арная переменная  $X$  – это упорядоченный набор унарных переменных  $X=(X_1, X_2, \dots, X_n)$ , таких, что пространство состояний  $\text{dom}(X)$  – суть декартово произведение  $\text{dom}(X)=\text{dom}(X_1)\times\text{dom}(X_2)\times\dots\times\text{dom}(X_n)$  соответствующих пространств, а  $\|X\|=\prod_i\|X_i\|$ . Элементы пространства состояний будем называть состояниями или конфигурациями переменной  $X$  и записывать в виде  $(x_1, \dots, x_n)$ . Важно, что пространство состояний для каждой переменной образует полную группу несовместимых взаимоисключающих событий (или альтернативных гипотез).

*Лингвистический (или нечеткий) потенциал* (в дальнейшем, коротко – потенциал) – это функция, которая каждой конфигурации  $x_i\in\text{dom}(X)$  переменной  $X$ , ставит во взаимно однозначное соответствие лингвистическую вероятность  $P_i$ , где  $i=1, \dots, n$ ,  $n$  – количество значений переменной  $X$  из  $\text{dom}(X)$ , т. е.  $\varphi:\text{dom}(X)\rightarrow\{P_i\}, i=1, \dots, n$ .  $\{P_i\}$  – множество лингвистических вероятностей.

Над потенциалами определены операции: умножение, маргинализация (проецирование), деление и нормализация [5, 6]. Так, если  $\varphi$  и  $\psi$  – потенциалы, определенные над  $\text{dom}(X)$ ,  $X\subseteq D$ , и  $\text{dom}(Y)$ ,  $Y\subseteq D$  соответственно, и  $z\in\text{dom}(X\cup Y)$ ,  $X\cup Y\subseteq D$  ( $D$  – множество переменных) есть некоторой конфигурацией, тогда умножение потенциалов  $\varphi\cdot\psi$  определяется через умножение лингвистических вероятностей для соответствующих конфигураций по формуле (1):

$$(\varphi \cdot \psi)(z) = \varphi(z_X) \cdot \psi(z_Y), \quad (1)$$

где  $z_X$  и  $z_Y$  – проекции  $z$  на  $\text{dom}(X)$  и  $\text{dom}(Y)$  соответственно. Причем, перед умножением потенциалов их области определения расширяются к  $\text{dom}(\varphi\psi) = \text{dom}(X\cup Y) = X_1\times\dots\times X_n$ , а элементами этой области определения становятся конфигурации вида  $z = (x_1, \dots, x_n)$ .

Операция *маргинализации потенциала* выполняется по одной из формул

$$M_X = \varphi^{\downarrow X}(z_X) = \sum_{z_Y} \varphi(z), \quad (2)$$

$$M_X = \varphi^{\downarrow X}(z_X) = \max_{z_Y} \varphi(z), \quad (3)$$

где  $z_X$  и  $z_Y$  – проекции  $z$  на  $\text{dom}(X)$  и  $\text{dom}(Y)$  соответственно. Операции маргинализации (2) и (3) вычисляют потенциал, который определен на множестве, являющимся подмножеством области определения исходного потенциала. Эта операция используется в алгоритмах распространения доверия для вычисления коэффициента обновления доверия. Локальные вычисления

такого коэффициента согласовывают априорные распределения вероятностей отдельных переменных так, чтобы совместно они определяли распределение, представляемое всей БС. Применение sum-маргинализации потенциалов в алгоритмах распространения доверия позволяет вычислить апостериорное распределение отдельно по каждой переменной БС. Применение max-маргинализации позволяет вычислить наиболее вероятную конфигурацию одновременно для всех переменных БС. Уместно отметить, что для выполнения max-маргинализации требуется сначала вычислить центр тяжести каждой лингвистической вероятности, а затем максимальное значение в соответствии с (3).

Деление потенциалов [5, 6] вычисляется согласно

$$(\varphi/\psi)(z) = \begin{cases} \varphi(z_X)/\psi(z_Y), & \text{если } \psi(z_Y) \neq 0, \\ 0 & \text{– в других случаях.} \end{cases} \quad (4)$$

Нормализация потенциала [8] выполняется согласно формуле

$$\eta(\varphi(X)) = \frac{\varphi(X)}{\sum_{x \in \text{dom}(X)} \varphi(x)}, \quad (5)$$

при этом  $\sum_{x \in \text{dom}(X)} \eta(\varphi(x)) = 1$  (единица четкая).

Основной моделью для представления причинно-следственных связей между переменными в условиях неопределенной информации является *нечеткая сеть доверия* (иначе – нечеткая байесовская сеть (НБС))  $N = (X, G, M)$ , которая состоит из

- 1) ациклического ориентированного графа  $G = (V, E)$  с вершинами  $V = \{v_1, \dots, v_n\}$  и дугами  $E = V \times V$ ;
- 2) множества переменных  $X = (X_{v_1}, \dots, X_{v_n})$ , элементы которого сопоставлены вершинам графа  $G$ , причем для каждой вершины сети определен нечеткий потенциал  $\varphi_{v_i}(X_{v_i})$ , где  $i = 1, \dots, n$ ;
- 3) множества нечетких потенциалов  $\{\varphi(X_{v_1} | X_{pa(v_1)}), \dots, \varphi(X_{v_n} | X_{pa(v_n)})\}$  (здесь,  $X_{pa(v)}$  – родительские переменные для переменной  $X_v$ ; если для вершины  $v$  множество родительских переменных пусто, тогда  $\varphi(X_v | X_{pa(v)}) = \varphi(X_v)$  и такую вершину называют маргиналом).

*Основной задачей* вероятностного вывода на НБС является вычисление распределения апостериорных оценок нечетких потенциалов (лингвистических вероятностей) над множеством переменных запроса при условии, что имеет место наблюдение события, состоящего в присвоении конкретных значений некото-

рому множеству переменных (свидетельств). Таким образом, НБС содержит исчерпывающее представление полного общего распределения лингвистических вероятностей. Обозначим:  $Z$  – переменная запроса;  $E$  – множество переменных-свидетельств  $E_1, E_2, \dots, E_m$ ;  $Y$  – другие переменные  $Y_1, Y_2, \dots, Y_n$ , которые не входят во множество переменных-запроса и во множество переменных-свидетельств. В этих обозначениях суть запроса можно представить выражением вида

$$\varphi(Z|E) = M_{Z,E} \varphi(Z, E, Y), \quad (6)$$

где  $\varphi(Z, E, Y)$  – произведение всех потенциалов заданных над НБС. Содержательно, операция (6) – суть вычисления распределения апостериорных оценок для переменной запроса  $Z$ , когда выполнено некоторое присваивание значений множеству переменных-свидетельств  $E$ , на основе выбранной операции маргинализации потенциалов.

Уместно отметить, что непосредственно реализовать такой запрос на практике можно лишь для достаточно простых НБС, состоящих из небольшого количества переменных (порядка 10). Известно, что размер компьютерной памяти, необходимый для вычисления (6), выполняемого методом перебора всех значений элементов множества  $Y$ , линейно зависит от количества переменных. Однако количество шагов, выполняемых во время вычисления (6) для  $n$  переменных, каждая из которых может находиться не больше, чем в  $m$  состояниях, всегда составляет  $O(m^n)$ .

Более прогрессивные методы вычисления распределения апостериорных вероятностей используют идею *элиминации* (исключения) переменной. Ее суть в следующем: поочередно выполняются операции умножения и маргинализации потенциалов таким образом, чтобы каждая последующая операция выполнялась лишь над частью общего распределения лингвистических вероятностей. Известно [1], что оценка времени выполнения алгоритма вероятностного вывода на НБС типа дерева или полидерева (дерева, которое имеет несколько корней), линейно зависит от размера сети, если количество родительских вершин ограничено некоторой константой. Относительно времени вычисления апостериорного распределения вероятностей на НБС, представленной орграфом, время выполнения алгоритмов элиминации переменной в наихудшем случае будет иметь экспоненциальную оценку [6], даже для ограниченного количества родительских переменных. То есть вероятностный вывод на произвольной НБС относится к NP-трудной задаче [5, 6, 8].

В работе для эффективного решения проблемы вероятностного вывода на НБС использован метод, предложенный Перлом [9]. Суть этого метода состоит в том, что для реализации вероятностного вывода на НБС предварительно создается вторичная структура данных в виде *узловой дерева*, позволяющая уменьшить количество шагов выполнения алгоритма вывода до  $O(n^2)$ , при условии ограничения на количество родительских переменных. Методологически этот подход предусматривает выполнение трех нетривиальных алгоритмов [5, 6]:

- трансформация НБС в узловое дерево (УД) для последующего выполнения локальных вычислений над НБС;
- инициализация нечеткими оценками УД;
- выполнение алгоритма *распространения доверия* (АРД): локальные вычисления и глобальное согласование полученных результатов.

Трансформация НБС осуществляется по схеме модифицированного алгоритма [10]:

- 1) создание морального графа – морализация;
- 2) создание хордального графа и клик (полных подграфов) – хордализация;
- 3) создание сепараторов (структур данных для сохранения информации о ребрах дерева) и их упорядочивание согласно заранее выбранной эвристики;
- 4) построение УД путем объединения клик в соответствии с упорядоченной последовательностью сепараторов.

*Морализация* НБС сводится к попарному соединению ребром всех родительских вершин для каждой вершины графа. После морализации в графе все дуги заменяются ребрами.

*Хордализация* графа заключается в выполнении такого его разбиения, чтобы любой цикл, с числом вершин свыше четырех, можно было разбить на циклы с тремя вершинами. Поскольку эта задача является NP-трудной для произвольного орграфа, поэтому для ее выполнения применяют разные эвристики. Результатом применения эвристик является хордализированный (триангулированный) граф с разбивкой на треугольные циклы, которая (разбивка) необязательно является оптимальной, но достаточно близкая к оптимальному варианту.

Создание *сепараторов* происходит на основе *триангулированного графа*, в котором каждый сепаратор – это полный подграф, являющийся пересечением двух клик. На этом же шаге сепараторы сортируются для ускорения их анализа на следующем шаге.

Последний шаг трансформации заключается в создании узлового дерева, совмещающего клики так, чтобы сепаратор, соединяющий их, имел наибольшую мощность.

*Инициализация* нечеткими оценками УД заключается в связывании потенциалов НБС с кликами УД. Потенциал НБС связывается (умножается) с потенциалом клики, если множество переменных, над которыми определен потенциал НБС, является подмножеством множества переменных НБС, входящих в клику. Потенциалы сепараторов инициализируются единицей.

На последнем этапе вероятностного вывода выполняется АРД над УД, в основе которого находится процедура *пересылки сообщений* между соседними кликами. Сообщение (коэффициент обновления доверия) вычисляется на основе маргинализации потенциалов соседних клик и умножения результата маргинализации на потенциал клики, называемой корневой кликой. Такие пересылки выполняются в двух направлениях: к выбранной, в качестве корня, клике и в обратном направлении. В результате происходит распространение свидетельств по всему узловому дереву, а следовательно и по всей НБС.

АРД выполняет вероятностный вывод за время, пропорциональное размеру сети. Выполнение промежуточных шагов, вообще, требует использования дополнительных памяти компьютера и времени для построения клик. Уменьшения указанных затрат удается достичь используя дополнительные структуры данных вместе с эвристиками во время построения клик [6, 10].

**Программные средства анализа нечетких сетей доверия.** Для автоматизации этапов моделирования на НБС предложена послойная структура программной системы (таблица). Механизм обмена данными между компонентами этой системы спроектирован согласно модельно-ориентированного подхода и реализуется путем обмена метаданными между подсистемами. Таким образом, структурно классы системы можно разделить на четыре слоя: Основа, Ресурс, Механизм вывода, Управление.

ТАБЛИЦА. Послойная структура программной системы

<i>Управление</i>	Создание, чтение, запись НБС в XML-формате		Чтение, запись НБС в базу данных	
<i>Механизм вывода</i>	Трансформация (модель, настройка)	Алгоритмы распространения доверия	Информационная визуализация	
<i>Ресурс</i>	XML-ресурс		База данных (технология ADO)	
<i>Основа</i>	Вершины НБС и узлового дерева	Дуги НБС	Сепараторы (ребра) узлового дерева	Потенциалы

Нижний слой системы – *Основа* состоит из классов, которые поддерживают спецификацию базовых структурных элементов. Все они используются классами верхних уровней. В этот слой входят классы, которые представляют вершины НБС и УД [10], дуги НБС и ребра УД, потенциалы вершин, индексы конфигураций состояний в потенциалах и векторы ключей для установления соответствия между областями определения перемножаемых потенциалов.

Второй слой системы – *Ресурс* состоит из классов, которые используются для описания информационных источников, представленных в формате стандарта PMML [2], использующим XML-формат для описания моделей данных, и базы данных, работа с которой организуется средствами технологии ADO. В процессе вызова метода записи НБС проверяются соответствующие настройки на необходимость изменения внутреннего формата данных системы в формат, используемый внешней системой. Перед изменением формата над данными выполняется ряд трансформаций: дефузификация, нормализация, отображение дискретных значений из одной шкалы в другую. В процессе считывания НБС выполняется обратное изменение форматов и проверка на необходимость трансформации данных: дискретизация, фузификация [7, 11], нормализация, отображение дискретных значений из одной шкалы в другую.

Третий слой – *Механизм вывода* состоит из следующих частей: Трансформация, Алгоритм, Информационная Визуализация (ИВ). Этот слой является основной частью бизнес-логики системы. Поэтому каждая составляющая третьего слоя имеет сложную структуру. Классы, которые принадлежат к Трансформации, концептуально объединяют две области: 1) Модель; 2) Настройка. Пакет классов. Модель состоит из классов, которые представляют метаданные, описывают набор входных атрибутов для построения графа, процедуры конструирования графов, результат соответствующего этапа трансформации, процедуры проверки корректности модели. Центральным классом пакета Алгоритм является АРД, содержащий описание процедуры распространения доверия в УД. Используя настройки, задающиеся пользователем (вид маргинализации, наличие свидетельств), выполняется алгоритм вероятностного вывода.

К ИВ принадлежат следующие классы: 1) классы, представляющие метаданные для визуализации НБС, УД, таблиц потенциалов и результатов вычислений в виде диаграмм; 2) класс описания входных параметров визуализации; 3) класс, специфицирующий процедуру визуализации изображения НБС и УД. Графический интерфейс системы позволяет пользователю создавать и вносить изменения в структуру НБС и значение размытых потенциалов вручную, или для автоматического построения соединиться с источником данных (в том числе с отдаленным), который содержит информацию о созданной ранее НБС. С помощью интерфейса системы пользователь имеет возможность пересмотреть, как изменяются значения размытых потенциалов на каждом этапе вероятностного вывода на НБС.

Четвертый слой – *Управление* состоит из классов, специфицирующих процессы контроля за загрузкой, сохранением и внесением изменений в структуру и числовые параметры НБС. В этот слой включены объекты, которые на основе технологии ADO устанавливают соединение с источником данных, получают в ответ на созданный пользователем запрос данные для построения НБС. Благодаря этим объектам система получает универсальный доступ (в том числе и отдаленный) к источникам, которые содержат информацию о НБС.

В соответствии с приведенной архитектурой программная система совмещает ряд объектно-ориентированных подсистем, каждая из которых является самостоятельной универсальной программной единицей и предназначена для автоматизации отдельных этапов трансформации и выполнения АРД (рис. 1).

Каждая подсистема унифицирована, что позволяет вносить конструктивные изменения в алгоритмы, не модифицируя при этом остальные части системы. Благодаря модельно-ориентированному подходу к обмену метаданными между подсистемами, система позволяет создать НБС с унифицированной структурой причинно-следственных связей между вершинами и разными числовыми параметрами этой сети. Такая возможность позволяет проводить одновременное исследование состояний сложных систем в условиях различной информационной нагрузки.

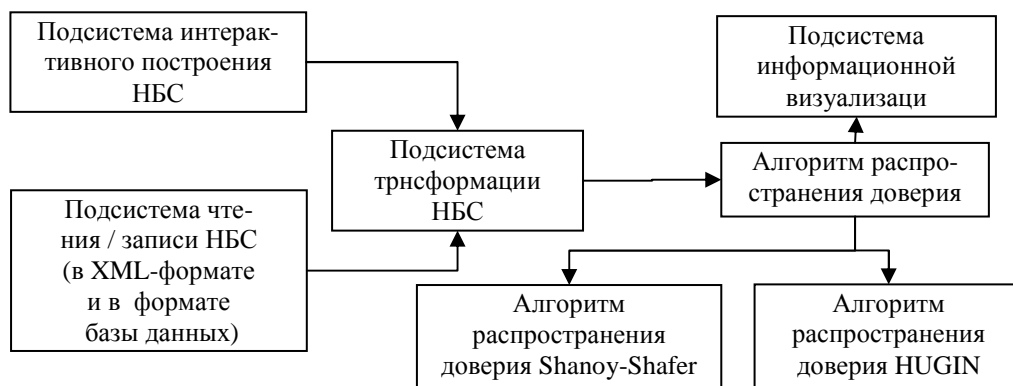


РИС. 1. Архитектура системы

Визуализация результатов выполнения АРД позволяет обнаружить возможные коллизии между свидетельствами и выполнить коррекцию структуры БС, а также ее числовых параметров. Система предоставляет возможность изменять свидетельство в вершинах БС, тем самым позволяя пользователю выполнять анализ, цель которого обнаружение для состояний целевых переменных наиболее влиятельные связки (провести так называемый анализ чувствительности [8]).

Функциональные возможности системы отображаются через графический интерфейс системы (рис. 2).

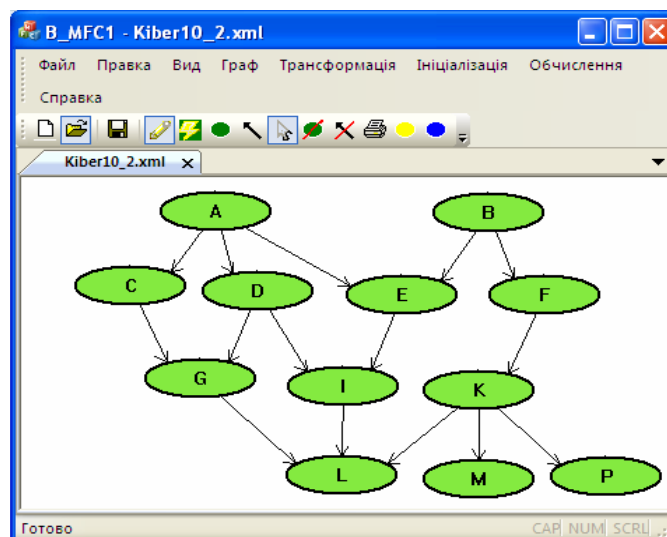


РИС. 2. Иллюстрация функций главного меню и поля графа модели



Команды Файл, Правка, Вид в соответствии с принятым стандартом интерфейса обеспечивают взаимодействие пользователя с данными, которые сохраняются или читаются из диска, позволяют пользователю интерактивно вносить изменения в НБС, используя буфер обмена, и настраивать общий вид главного окна системы.

Меню команды Граф обеспечивает интерактивное взаимодействие пользователя с системой через графический интерфейс: вставлять / удалять объекты НБС и инициализировать потенциалы каждой вершины через систему диалоговых окон. Доступ к основным инструментам системы пользователь получает через меню команд Трансформация, Инициализация, Вычисление.

**Выводы.** Таким образом, в статье освещены основные наиболее принципиальные и важные вопросы создания информационной технологии вероятностного моделирования на основе НБС. На этих результатах авторами построена экспериментальная версия информационной технологии вероятностного моделирования состояний сложных систем для треугольных и трапециевидных непрерывных функций принадлежности.

*І.М. Парасюк, Ф.В. Костукевич*

ПРО РЕАЛІЗАЦІЮ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ЙМОВІРНІСНОГО МОДЕЛЮВАННЯ СТАНІВ СКЛАДНИХ СИСТЕМ НА БАЗІ НЕЧІТКИХ МЕРЕЖ ДОВІРИ

Розглянуті ключові питання комп'ютерної реалізації інформаційної технології ймовірнісного моделювання станів складних систем на базі нечітких мереж довіри. Конструктивно викладений математичний апарат, за допомогою якого здійснюється аналіз нечітких мереж довіри та прогнозування станів досліджуваних систем в умовах невизначеної інформації. Значна увага приділена технологічним, архітектурним і функціональним питанням комп'ютеризації.

*I.M. Parasyuk, F.V. Kostukevich*

ON REALIZATION OF INFORMATION TECHNOLOGY OF PROBABILISTIC MODELING OF THE STATES OF COMPLEX SYSTEMS ON A BASE OF FUZZY BELIEF NETWORKS

The key computer realization questions of information probabilistic modeling technology of the states of complex systems are considered on the base of fuzzy belief networks. A mathematical tool by which the analysis of fuzzy belief networks and prognostication of the states of the investigated systems is carried out under the conditions of indefinite information is structurally expounded. Considerable attention is spared to the technological, architectural and functional questions of computerization.

1. *Рассел С., Норвиг П.* Искусственный интеллект: современный подход, 2-е изд.: Пер. с англ. – М.: Издательский дом «Вильямс», 2006. – 1408 с.
2. *Лотфи А. Заде.* Основы нового подхода к анализу сложных систем и процессов принятия решений // Математика сегодня (сборник статей, перевод с англ.). – М.: «Знание», 1974. – С. 5–48.

3. *Веревка О.В., Парасюк И.Н.* О распространении вероятностей в нечетких байесовских сетях с недетерминированными состояниями // Кибернетика и системный анализ. – 2008. – № 6. – С. 153–169.
4. *Мацневский С.В.* Нечеткие множества: Учебное пособие. – Калининград: Изд-во КГУ, 2004. – 176 с.
5. *Lauritzen S.L., Spiegelhalter D.J.* Local computations with probabilities on graphical structures and their application to expert systems (with discussion) // J. the Royal Statistical Society, Series B. – 1998. – **50**. – P. 157–224.
6. *Cowell R.G., Dawid A.P., Spiegelhalter D.J., Lauritzen S.L.* Probabilistic Networks and Expert Systems. – New York: Inc., Springer-Verlag, 1999. – 321 p.
7. *Рутковская Д., Пилиньский М., Рудковский Л.* Нейронные сети, генетические алгоритмы и нечеткие системы: Пер. с польск. И.Д. Рудинского. – М.: Горячая линия. – Телеком, 2006. – 452 с.
8. *Kjaerulff U., Madsen A.* Bayesian Networks and Influence Diagrams. – Springer Science+Business Media, LLC, 2008. – 318 p.
9. *Pearl J.* Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. – Morgan Kaufmann, San Mateo, CA, 1991. – 552 p.
10. *Парасюк И.Н., Костукевич Ф.В.* Методы трансформации байесовской сети для построения узлового дерева и их модификация // Компьютерная математика. – 2008. – № 1. – С. 70–80.
11. *Кофман А.* Введение в теорию нечетких множеств. – М.: Радио и связь, 1982. – 432 с.

Получено 14.12.2010

**Об авторах:**

*Парасюк Иван Николаевич,*  
член-корреспондент НАН Украины,  
заведующий отделом Института кибернетики имени В.М. Глушкова НАН Украины,  
E-Mail: [ivpar1@i.com.ua](mailto:ivpar1@i.com.ua)

*Костукевич Феликс Витальевич,*  
аспирант Института кибернетики имени В.М. Глушкова НАН Украины.  
E-Mail: [felkost@gmail.com](mailto:felkost@gmail.com)