

УДК 519.6

ПРИМЕНЕНИЕ ИНСТРУМЕНТОВ БАЙЕСОВСКИХ СЕТЕЙ ДЛЯ ПРОГНОЗИРОВАНИЯ ТЕНДЕНЦИЙ ИЗМЕНЕНИЯ СТОИМОСТИ АКЦИЙ

Азарнова Татьяна Васильевна, д-р техн. наук, проф.
Медведев Олег Николаевич, маг.

Воронежский государственный университет, Университетская пл., 1, Воронеж,
Россия, 394006; e-mail: ivdas92@mail.ru

Цель: разработка и анализ вероятностной модели прогнозирования тенденций изменения стоимости акций. *Обсуждение:* рассматриваются возможности представления модели в виде динамической сети Байеса. Инструменты динамических сетей Байеса хорошо апробированы в задачах моделирования динамических процессов, протекающих в условиях риска и неопределенности. В работе исследованы вопросы синтеза и семантики анализируемой динамической сети Байеса. *Результаты:* разработан алгоритм опроса сети, базирующийся на построении дерева сочленений. Предложенный в рамках исследования математический инструментарий алгоритмически и программно подготовлен к практическому внедрению.

Ключевые слова: байесовская сеть, алгоритм опроса сети, дерево сочленений, приближенные алгоритмы вероятностного вывода, фильтрации и прогнозирования байесовских сетей.

DOI: 10.17308/meps.2015.4/1224

Введение

Стоимость акций компаний формируется под воздействием большого количества различных экономических, политических, технологических факторов. Влияние факторов на формирование стоимости носит недетерминированный случайный характер, практически невозможно выявить точные закономерности влияния, в лучшем случае, удастся оценить вероятность, с которой изменение фактора приведет к определенному изменению стоимости акций [2-4]. Эффективным инструментом моделирования таких вероятностных закономерностей являются статические и динамические байесовские сети [1].

Байесовская сеть моделируется как ориентированный граф без ци-

клов [7]. Вершинами данного графа являются случайные величины (дискретные и (или) непрерывные). Наличие дуги от вершины x к вершине o означает, что $x \in Parents(o)$, т.е. x является родительской для o и оказывает непосредственное влияние на o . Сеть, как правило, изображается так, что родительские вершины находятся на более высоком уровне, чем их потомки. Каждая вершина x байесовской сети характеризуется распределением условных вероятностей $P(x|Parents(x))$, предполагается, что после задания родительских вершин x становится условно независимой от других вершин, лежащих на более высоких уровнях, чем родительские. С помощью условных вероятностей задается полное совместное распределение:

$$P(x_1, x_2, \dots, x_n) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \prod_{i=1}^n P(x_i | Parents(x_i)).$$

Байесовские сети позволяют реализовать задачу вероятностного вывода [1], заключающуюся в вычислении распределения вероятностей для переменных запроса $X = \{X_1, X_2, \dots, X_n\}$, если в результате события e переменные свидетельства $E = \{E_1, E_2, \dots, E_m\}$ приняли значения $E_1 = e_1, E_2 = e_2, \dots, E_m = e_m$ (множество всех вершин сети представляется в виде $X \cup E \cup Y$, где $Y = \{Y_1, Y_2, \dots, Y_p\}$ – скрытые переменные). Для вычисления вероятностей $P(X|E) = \alpha \sum_Y P(X, E, Y)$ (α – нормирующий множитель) используются точные и приближенные алгоритмы вероятностного вывода. Среди точных алгоритмов можно отметить алгоритмы: перебора, устранения переменной и кластеризации, среди приближенных – алгоритмы: непосредственной выборки с исключением выборок, несоответствующих свидетельству $E = \{E_1, E_2, \dots, E_m\}$; оценки веса выборки с учетом правдоподобия того, насколько она согласована со свидетельством; Монте-Карло с применением цепи Маркова.

Вероятностный вывод на основе динамических байесовских сетей

Динамические байесовские сети представляют последовательность байесовских сетей – состояний во времени (временных срезов), начиная с момента $t = 0$. Для каждого момента времени t есть свои множества X_t, E_t ненаблюдаемых и наблюдаемых переменных, причем считается, что свидетельства начинают появляться с $t = 1$. Через $X_{t_1:t_2}, E_{t_1:t_2}$ обозначается последовательность наборов вершин временных срезов с момента t_1 до момента t_2 . Внутри всех временных срезов таблицы условных вероятностей остаются неизменными, они задаются для одного репрезентативного среза, например, для момента времени $t = 0$ ($P(X_0)$). Для динамических байесовских сетей каждая переменная может быть связана с неограниченным количеством переменных на предшествующих временных срезах, для решения данной проблемы вводится предположение марковости, в соответствии данным предположением любое состояние зависит от конечного числа предыдущих состояний. Для Марковских процессов первого порядка текущее состояние переменной состояния зависит только от непосредственно предшествующего состояния

$$P(X_t | X_{0:t-1}) = P(X_t | X_{t-1}),$$

а состояние переменной свидетельства зависит только от текущего состояния

$$P(E_t | X_{0:t}, E_{0:t-1}) = P(E_t | X_t).$$

Вероятностное распределение $P(X_t | X_{t-1})$ называется моделью перехода, а вероятностное распределение $P(E_t | X_t)$ – моделью восприятия.

Полное совместное распределение вероятностей динамической байесовской сети задается

$$P(X_0, X_1, \dots, X_t, E_1, \dots, E_t) = P(X_0) \prod_{i=1}^t P(X_i | X_{i-1}) P(E_i | X_i).$$

В таблице приведены основные задачи, которые можно решать с помощью динамических байесовских сетей.

Таблица

Задачи вероятностного вывода [1]

Задача	Характеристика
Фильтрация	<p>Вычисление вероятностей $P(X_t E_{1:t})$ для текущего состояния при наличии свидетельств $E_{1:t}$, начиная с момента $t = 1$ и до момента t. Задача решается рекурсивным способом, проецированием от t к $t+1$, далее, при новых свидетельствах для момента $t+1$, распределение обновляется</p> $P(X_{t+1} E_{1:t+1}) = P(X_{t+1} E_{1:t}, E_{t+1}) = \alpha P(E_{t+1} X_{t+1}) P(X_{t+1} E_{t:t}) = \alpha P(E_{t+1} X_{t+1}) \sum_{X_t} P(X_{t+1} X_t) P(X_t E_{1:t})$
Предсказание	<p>Вычисление апостериорных вероятностей $P(X_{t+k} E_{1:t})$ будущего состояния относительно всех свидетельств, полученные к данному моменту времени. Задача решается путем рекурсивного вычисления вероятностного распределения в момент времени $t+k+1$ на основании предсказания для $t+k$</p> $P(X_{t+k+1} E_{1:t}) = \sum_{X_{t+k}} P(X_{t+k+1} X_{t+k}) P(X_{t+k} E_{1:t})$
Сглаживание (ретроспективный анализ)	<p>Вычисление апостериорных вероятностей значений переменных $P(X_k E_{1:t})$ в прошлом, если известны все свидетельства вплоть до нынешнего. Вычисление осуществляется по формулам:</p> $P(X_k E_{1:t}) = \alpha P(X_k E_{1:k}) P(E_{k+1:t} X_k)$ $P(E_{k+1:t} X_k) = \sum_{X_{k+1}} P(E_{k+1:t} X_k, X_{k+1}) P(X_{k+1} X_k) = \sum_{X_{k+1}} P(E_{k+1} X_{k+1}) P(E_{k+2:t} X_{k+1}) P(X_{k+1} X_k)$

Временной срез байесовской сети для прогнозирования тенденций изменения стоимости акций нефтяной компании приведен на рис. 1. В динамической сети каждый срез имеет аналогичную структуру, связи между слоями (не отражены на рисунке) удовлетворяют требованию Марковских процессов.

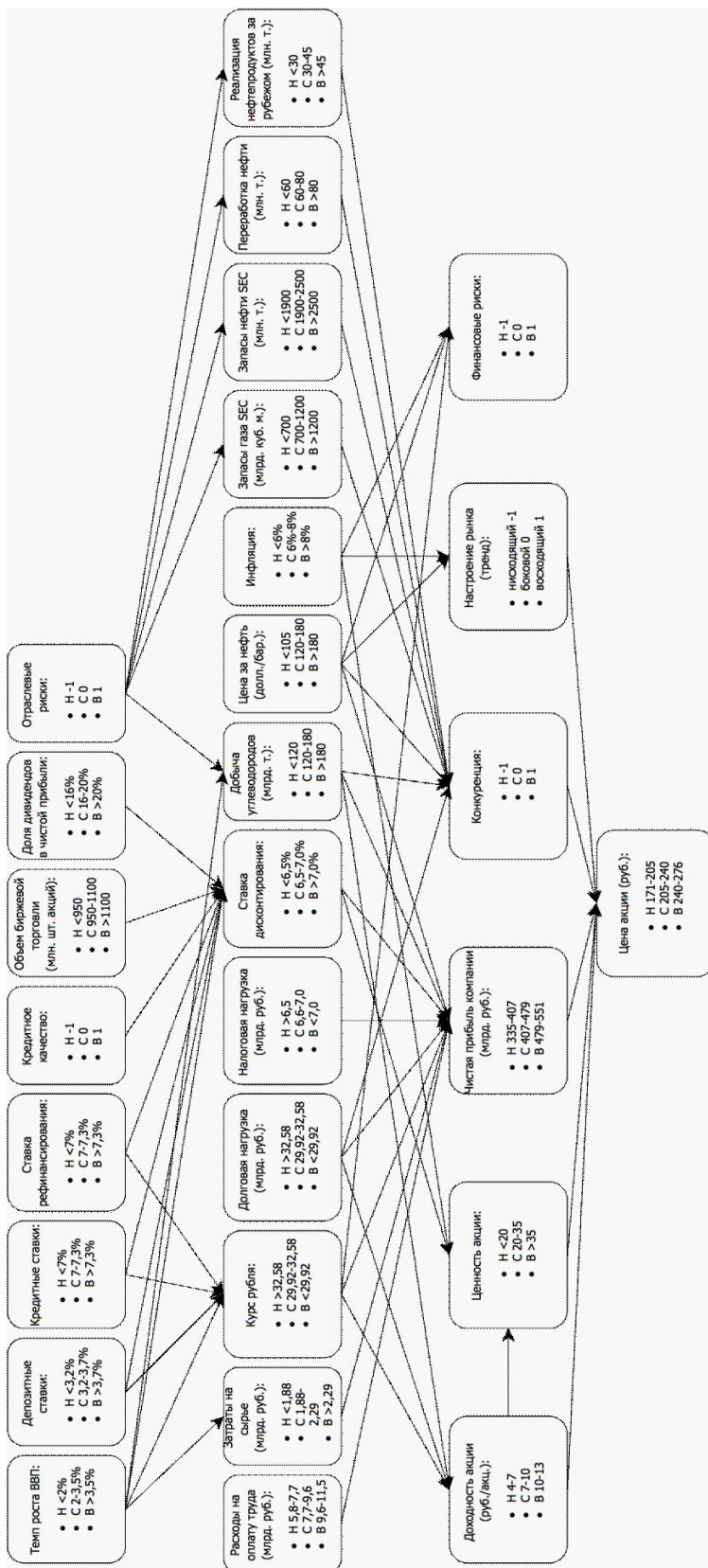


Рис. 1. Байесовская сеть для прогнозирования тенденции изменения стоимости акций, где Н – низкий, С – средний, В – высокий уровень

Детерминанты стоимости акции в задаче вероятностного вывода

Для представленной сети задача вероятностного вывода (опроса) является задачей предсказания вероятностей различных тенденций изменения стоимости акций при известных свидетельствах – значениях факторов, влияющих на стоимость акций. Задача фильтрации заключается в вычислении апостериорных вероятностей различных тенденций изменения стоимости акций $P(X_t|E_{1:t})$ при условии наличия всех свидетельств $E_{1:t}$ по значениям факторов, начиная с начального момента $t = 1$ и до текущего момента времени t . Аналогично можно проинтерпретировать другие задачи вероятностного вывода.

Опишем кратко основные факторы, которые являются вершинами сети. Важнейшим фактором стоимости акций является прибыль, которую зарабатывает компания. В свою очередь, ключевыми показателями деятельности предприятия, которые определяют уровень прибыли, являются выручка и себестоимость. На выручку и себестоимость оказывает влияние целый ряд факторов, структура которых приведена на рис. 2.

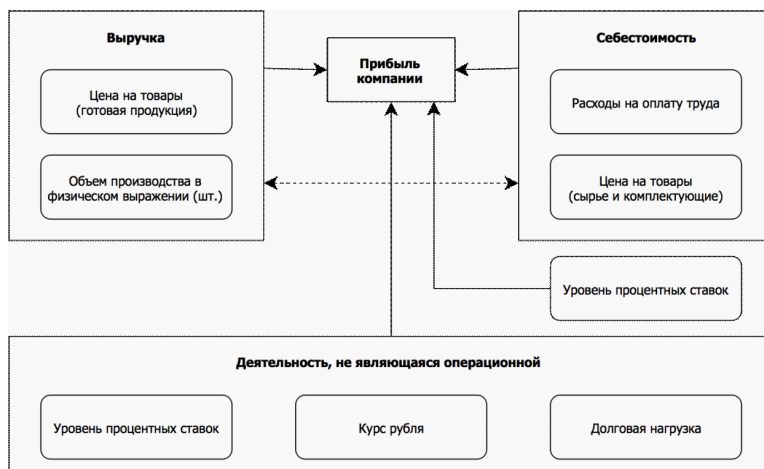


Рис. 2. Структура факторов, влияющих на выручку и себестоимость

Цена акций также существенно зависит от ставки дисконтирования, отражающей минимально допустимую отдачу на вложенный капитал, при которой инвестор не предпочтет участие в альтернативном проекте с сопоставимой степенью риска. Структура факторов формирования ставки дисконтирования приведена на рис. 3.

Важнейшим макроэкономическим показателем, от которого зависит стоимость всего бизнеса в стране, является общий уровень процентных ставок. Он зависит от уровня кредитных и депозитных ставок, уровня инфляции, ставки рефинансирования, кредитного рейтинга государства и еще целого ряда факторов. Кредитное качество – ключевая индивидуальная характеристика компании, определяющая ставку дисконтирования, это способность компании брать деньги в долг. Уровень ликвидности определяется объемом биржевой торговли акциями.



Рис. 3. Структура факторов, формирования ставки дисконтирования

Важным показателем является ценность акции. Ценность акции (Price / Earnings Ratio, P/E) рассчитывается как частное от деления рыночной цены акции на доход. Коэффициент P/E служит индикатором спроса на акции данной фирмы, поскольку показывает, как много согласны платить инвесторы в данный момент за один рубль прибыли на акцию. Рост этого показателя в динамике указывает на то, что инвесторы ожидают более быстрый рост прибыли компании по сравнению с другими.

Алгоритм опроса сети на основе дерева сочленений

В данной работе для осуществления вероятностного вывода и обучения сети предлагается использовать алгоритм кластеризации, основанный на построении дерева сочленений.

Построение дерева сочленений состоит из трех этапов.

1) Построение доменного графа. Строится граф, вершины которого соответствуют вершинам БСД, а ребрами соединяются те вершины, которые в сети были зависимы друг от друга. Доменный граф не содержит таблиц условных вероятностей.

2) Построение морального графа. На этом этапе происходит морализация доменного графа – добавление ребер для узлов, имеющих общих потомков.

3) Построение триангулярного графа.

Триангулярным графом называется граф, не имеющий циклов, состоящих из четырех и более вершин. Для триангуляции морализованного доменного графа G можно использовать следующий алгоритм: создается копия графа $G - G_1$, пока граф G_1 не пуст, выбирается вершина V , имеющая наименьшее число соседних вершин. Если таких вершин несколько, то выбирается вершина с минимальным весом. Вес вершины определяется как произведение количества значений, принимаемых данной переменной и ее соседями. Если несколько вершин имеют наименьший вес – берется любая из них. Соединяются все вершины из числа соседей V . Аналогичные дуги добавляются в граф G . Вершина V удаляется из графа G_1 . Шаг повторяется, пока граф G_1 не пуст.

Построение дерева объединений. Дерево объединений – древовидный граф, каждый узел которого представляет собой такое подмножество

вершин триангулярного графа, что каждые две вершины этого подмножества соединены ребром графа. Такое множество называется кликой графа. Пусть есть триангулярный граф G . Создадим переменную-счетчик i . Изначально $i=0$. Пока граф G не пуст, выбираем из него вершину V , имеющую наименьшее число соседних вершин. Если таких вершин несколько, то выбираем вершину с наименьшим весом. Множество, состоящее из вершины V и её соседей, помечаем как C_i . Множество вершин из C_i , имеющих соседей, не принадлежащих C_i , помечаем как S_i (множество сепаратор). Удаляем из G тех соседей V , которые не вошли в сепаратор S_i . Удаляем V из G . Увеличиваем счетчик i на 1. Повторяем шаг. Каждый узел C_i соединяем с сепаратором S_i . При этом узел, который был создан последним, на данном шаге алгоритма остается несоединенным из-за того, что сепараторов на один меньше числа узлов дерева. Каждый сепаратор S_i соединяем с таким узлом C_j , что $j > i$, C_j содержит все элементы S_i . Полученная структура из узлов в виде клик триангулярного графа, соединенных между собой через сепараторы, является деревом объединений.

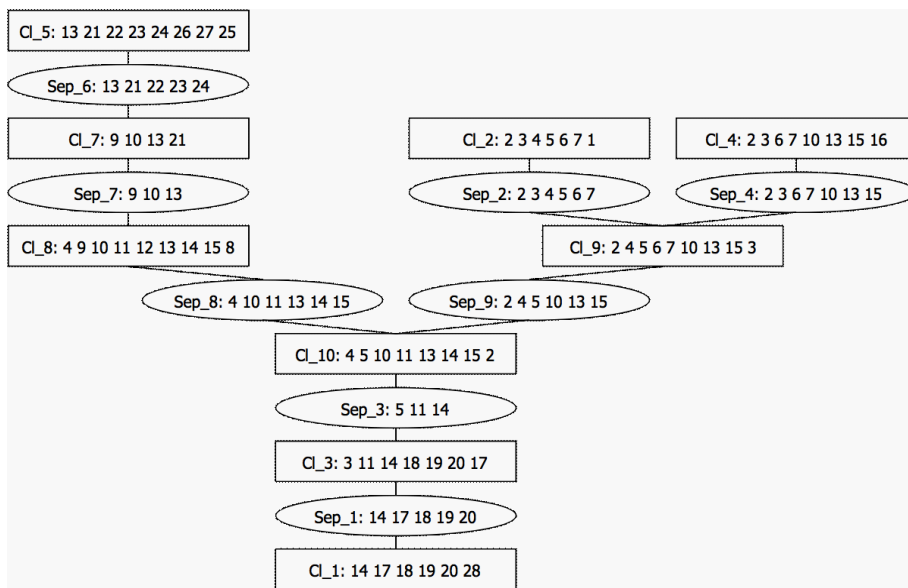


Рис. 4. Структура дерева объединений

Построение дерева сочленений. Деревом сочленений рассматриваемой БСД называется ее дерево объединений, в котором каждой вершине сопоставлено некоторое множество потенциалов условных вероятностей сети по правилу, что каждый потенциал сопоставляется любой вершине дерева, полностью содержащей его область определения. Потенциалы являются дискретными функциями нескольких переменных и представляют собой иную форму записи условных вероятностей:

$$P(x | y_1, y_2, \dots) = \varphi(x, y_1, y_2, \dots).$$

В итоге получим структуру из узлов, соединенных между собой через сепараторы, каждый узел содержит множество переменных, представляю-

щих собой клику триангулированного доменного графа, и набор потенциалов сети, области определения которых полностью содержатся во множестве переменных, принадлежащих этой клике. В процессе работы с деревом сепараторы также будут содержать в себе потенциалы, областью определения которых будет являться множество переменных сепаратора – на данном этапе потенциалы содержатся лишь в узлах дерева.

Построив дерево сочленений для рассматриваемой БСД, можно перейти непосредственно к процедуре опроса сети с использованием дерева.

Процесс вычисления конкретных вероятностей исходов рассматривается как распространение свидетельства в дереве сочленений – считая одну из вершин «корнем» дерева, потенциалы узлов (и сепараторов) обновляются с учетом поступившего свидетельства последовательно от «листьев» к «корню». Приведенный ниже алгоритм решает задачу вычисления вероятностей исходов некоторой переменной исследуемой БСД при заданном свидетельстве.

1. Пусть требуется вычислить вероятности различных исходов переменной V сети. Выбираем «корень» дерева, т.е. любую вершину дерева сочленений, содержащую переменную V (обозначим эту вершину через T). На данном этапе предполагается, что все сепараторы дерева не содержат потенциалов.

2. Распространение начать с вершин, с которыми соединен только один сепаратор. Отметим, что на данном и следующем этапах мы не рассматриваем вершину T . Распространение осуществляется следующим образом: все потенциалы, принадлежащие данной вершине, перемножаются между собой, затем полученный потенциал проецируется на множество переменных сепаратора, с которым соединена вершина, и помещается в сепаратор.

3. Далее распространение можно продолжить через остальные вершины по следующему правилу. Ищем вершину, у которой только один соседний сепаратор не содержит потенциала; перемножаем между собой все потенциалы этой вершины и полученный потенциал перемножаем с потенциалами соседних сепараторов, затем проецируем на множество переменных «пустого» соседнего сепаратора; повторяем шаг 3 алгоритма, если это возможно.

4. Теперь все соседние сепараторы вершины T содержат потенциалы. Перемножим все потенциалы вершины T и соседних сепараторов. Теперь спроецируем полученный потенциал на множество $\{V\}$. Полученный потенциал, зависящий только от одной переменной V , после нормировки на 1 будет содержать в себе искомые вероятности.

Заключение

Предложенный в рамках исследования математический инструментарий алгоритмически и программно подготовлен к практическому внедрению. Данный инструментарий можно использовать как средство поддержки принятия решений при управлении портфелями ценных бумаг. Апробация

и сравнительный анализ данного инструментария с другими разработками [5, 6] в рассматриваемой прикладной области показывают его конструктивный характер и хорошие прогностические способности.

Список источников

1. Азарнова Т.В., Полухин П.В. Применение динамических Байесовских сетей для повышения эффективности процесса фаззинга SQL –инъекций ВЕБ-приложений // *Системы управления и информационные технологии*, 2014, по. 1.1 (55), с. 106-111.
2. Давнис В.В., Коротких В.В. Адаптивное трендовое разложение финансовых временных рядов // *Современная экономика: проблемы и решения*, 2014, по. 10 (58), с. 8-24.
3. Давнис В.В., Коротких В.В. Модель альтернативных ожиданий и одно из ее приложений в портфельном анализе // *Современная экономика: проблемы и решения*, 2014, по. 5 (53), с. 31-46.
4. Давнис В.В., Коротких В.В. Об использовании двух гипотез при эконометрическом моделировании стохастических процессов // *Современная экономика: проблемы и решения*, 2014, по. 7 (55), с. 30-43.
5. Каширина И.Л. Управление портфелем ценных бумаг на основе методов прогнозирования достижения граничных состояний в дуальной вычислительной среде // *Экономика и менеджмент систем управления*, 2014, по. 1, с. 32-39.
6. Каширина И.Л., Азарнова Т.В., Косенко Д.О. Управление портфелем ценных бумаг с помощью генетического алгоритма // *Экономика и менеджмент систем управления*, 2014, по. 4.1, с. 177-184.
7. Тулупьев А.Л., Николенко С.И., Сироткин А.В. *Байесовские сети, логико-вероятностный подход*. Санкт-Петербург, Наука, 2006.

APPLICATION OF BAYESIAN NETWORK TO STOCK PRICE TREND PREDICTION

Azarnova Tatiana Vasilievna, Dr. Sc. (Eng.), Prof.
Medvedev Oleg Nikolaevich, M.Sc. student

Voronezh State University, University sq., 1, Voronezh, Russia, 394006;
e-mail: ivdas92@mail.ru

Purpose: analysis and development of the probabilistic model for stock price trend forecasting. *Discussion:* we presented the model as a dynamic Bayesian network. Dynamic Bayesian Networks tools well tested in problems of modeling of dynamic processes in the conditions of risk and uncertainty. This issue examines issues of synthesis and the semantics of dynamic Bayesian network analyzed. *Results:* we developed the network polling algorithm based on building wood joints. Proposed in the study mathematical tools algorithmically and program are ready for practical implementation.

Keywords: Bayesian Network algorithm of network polling, wood joints, approximate algorithms for probabilistic inference, filtering and prediction of Bayesian Networks.

References

1. Azarnova T.V., Polukhin P.V. Primenenie dinamicheskikh Baiesovskikh setei dlia povysheniia effektivnosti protsessa fazzinga SQL-in"ektsii VEB-prilozhenii. *Sistemy upravleniia i informatsionnye tekhnologii*, 2014, no. 1.1 (55), pp. 106-111. (In Russ.)
2. Davnis V.V., Korotkikh V.V. Adaptivnoe trendovoe razlozhenie finansovykh vremennykh riadov. *Sovremennaia ekonomika: problemy i resheniia*, 2014, no. 10 (58), pp. 8-24. (In Russ.)
3. Davnis V.V., Korotkikh V.V. Model' al'ternativnykh ozhidanii i odno iz ee prilozhenii v portfel'nom analize. *Sovremennaia ekonomika: problemy i resheniia*, 2014, no. 5 (53), pp. 31-46. (In Russ.)
4. Davnis V.V., Korotkikh V.V. Ob ispol'zovanii dvukh gipotez pri ekonometricheskom modelirovanii stokhasticheskikh protsessov. *Sovremennaia ekonomika: problemy i resheniia*, 2014, no. 7 (55), pp. 30-43. (In Russ.)
5. Kashirina I.L. Upravlenie portfelem tsennykh bumag na osnove metodov prognozirovaniia dostizheniia granichnykh sostoianii v dual'noi vychislitel'noi srede. *Ekonomika i menedzhment sistem upravleniia*, 2014, no. 1, pp. 32-39. (In Russ.)
6. Kashirina I.L., Azarnova T.V., Kosenko D.O. Upravlenie portfelem tsennykh bumag s pomoshch'iu geneticheskogo algoritma. *Ekonomika i menedzhment sistem upravleniia*, 2014, no. 4.1, pp. 177-184. (In Russ.)
7. Tulup'ev A.L., Nikolenko S.I., Sirotkin A.V. *Baiesovskie seti, logiko-veroiatnostnyi podkhod*. Saint-Petersburg, Nauka Publ., 2006.