

А. С. Климова, И. З. Батыршин, Н. К. Шайдуллина

ПРИМЕНЕНИЕ МЕТОДОВ ГИБРИДНОЙ РЕЛЯЦИОННОЙ КЛАСТЕРИЗАЦИИ К АНАЛИЗУ НЕФТЯНЫХ СКВАЖИН

Ключевые слова: гибридная реляционная кластеризация, визуализация данных, инвариантная иерархическая кластеризация.

В работе приведен метод гибридной реляционной инвариантной кластеризации с визуализацией сильных связей между объектами и применение этого метода для анализа взаимодействия нефтяных скважин на примере одного из Мексиканских месторождений.

Keywords: hybrid relational clustering, data visualization, invariant hierarchical clustering.

The paper presents a method hybrid relational invariant clustering with visualization of strong links between objects and application of this method to analyze the interaction of oil wells on the example of one of the Mexican oil fields.

Введение

В эксплуатации нефтяных месторождений большую роль играет анализ взаимодействия скважин. Наличие или отсутствие взаимодействия скважин влияет на оценку продуктивности нефтяных коллекторов, определение возможного расположения новых добывающих и нагнетающих скважин, разработку специальных мер по изменению свойств пластов в желательном направлении. Существуют различные методы анализа взаимодействия скважин. Например, закачка радиоактивных индикаторов в нагнетающую скважину и фиксация наличия индикаторов в добывающих скважинах. Это дает прямой метод анализа взаимодействия скважин. Этот метод достаточно дорог и требует наличия специального оборудования и квалифицированных специалистов. Метод тестирования скважин на интерференцию требует измерения скачка давления в одной скважине в ответ на закрытие другой скважины либо изменение давления в ней. Это также дорого и не всегда дает надежные результаты из-за возможных ошибок измерения давления и слабого изменения давления в тестируемой скважине. Математические модели направлений потоков в пластах требуют знания распределения проницаемости или давления в пласте, которые часто известны лишь в небольшом числе точек. Кроме этого, эти модели основаны на допущениях об однородности пластов, которые в реальности редко выполняются. Двухфазная фильтрация описана в [2], трехфазная фильтрация описана в [1].

Для анализа взаимодействия скважин в качестве альтернативы существующим методам предлагается использовать методы, основанные на анализе связей между временными рядами объемов добычи нефти и сопутствующей воды в скважинах исследуемого месторождения. Эти методы не требуют проведения экспериментов и измерений на скважинах.

Общая схема гибридной реляционной кластеризации

Рассмотрим общую схему гибридной реляционной кластеризации и визуализации. Пусть множество RC означает множество четких или нечетких отношений на M , а PC – множество

описаний элементов из M в пространствах различной размерности. Гибридной реляционной кластеризацией и визуализацией M будет называться преобразование

$$HC(S) = \langle RC; PC \rangle. \quad (1)$$

В общем случае преобразование HC должно удовлетворять некоторым условиям. Отношения из RC или расстояния между элементами множества M , определяемые представлением PC , должны быть согласованы с заданным отношением сходства S . Дополнительно могут рассматриваться условия согласованности между отношениями из RC , а также условия согласованности между отношениями из RC и расстояниями между объектами, определяемыми представлениями из PC . Рассмотрим частные случаи схемы (1).

Если $RC = \emptyset$, а PC задает описание элементов из M в пространстве размерности 2 или 3, то HC задает визуализацию данных в пространстве малой размерности и может определяться, например, методами многомерного шкалирования или главных компонент.

Если $PC = \emptyset$ и $RC = \{E\}$, где E – четкое или взвешенное отношение эквивалентности на M , то (1) определяет, соответственно, обычную или иерархическую кластеризацию множества M . Иерархическая кластеризация может быть задана также множеством четких отношений эквивалентности $RC = \{E_1, E_2, \dots, E_m\}$ таких, что $E_1 \subset E_2 \subset \dots \subset E_m$. Часто критерием кластеризации служит максимизация суммарного сходства $S(x, y)$ между объектами, объединяемыми в общий кластер отношения E , и минимизация сходства между объектами, попадающими в разные кластеры отношения E .

Инвариантная иерархическая кластеризация

Неинвариантная оптимальная кластеризация может быть результатом наличия равных значений сходства у функции сходства. На практике равные значения сходства часто появляются как результат округления вещественных чисел или использования шкал сходства с предельным значением, или как результат топологической структуры анализируемых данных. В этом случае должны применяться методы, которые дают инвариантные кластеризации с

маленьким, но не минимальным расстоянием между начальным и конечным отношением. Инвариантное решение может быть найдено среди не оптимальных решений или мы должны искать оптимальное решение среди всех инвариантных решений.

Общая схема иерархических кластерных процедур, удовлетворяющих условиям инвариантности относительно исходной нумерации объектов и инвариантности относительно монотонных преобразований значений сходства исследовалась в работах [3, 4, 5, 6]. Схема (1) имеет вид:

$$HC(S) = \langle E; \emptyset \rangle,$$

где E – взвешенное отношение эквивалентности. Преобразование $HC(S)$ задается процедурой кластеризации $Q(S)=E$, которая определяется так:

$$E = Q(S) = TC(F(S)) = F^{\wedge}(S),$$

где F это некоторая “коррекция” данного отношения сходства S , такая что $F(S) \subseteq S$, и TC обозначает процедуру транзитивного замыкания значений отношения сходства. Процедура транзитивного замыкания определена и может быть реализована алгоритмом “ближний сосед” [5, 6, 7, 8]. Эта процедура обладает обоими типами инвариантности. Если возможно представить некоторую процедуру коррекции F также обладающую обоими типами инвариантности, то мы построим кластерную процедуру Q , удовлетворяющую желаемым свойствам инвариантности.

Для построения подходящей процедуры коррекции, желательно решить для каких пар объектов (x, y) значения сходства $S(x, y)$ должны быть скорректированы и как эти значения должны быть уменьшены. Для этих целей может быть использована следующая оценка неразличимости. Будем говорить, что два объекта x и y неразличимы с согласия объекта z , когда $S(x, z) \geq S(x, y)$ тогда и только тогда, когда $S(y, z) \geq S(x, y)$. В этом случае мы будем говорить, что объект z “поддерживает” значение сходства $S(x, y)$. Чем больше объектов в X поддерживают значение сходства $S(x, y)$, тем больше степень неразличимости x и y . Наша задача изменить значение $S(x, y)$ так, чтобы число объектов поддерживающих сходство между x и y и, следовательно, степень неразличимости увеличились. В результате мы можем сказать, что если объекты x и y неразличимы только с согласия маленькой части объектов и, следовательно, они показывают разное поведение на большей части объектов, то значение сходства $S(x, y)$ не подтверждается или не поддерживается объектами множества X и в результате значение сходства $S(x, y)$ может быть скорректировано (уменьшено).

Рассматриваемая процедура коррекции зависит от следующих множеств и параметров:

$$V_y(x) = \{z \in X \setminus \{x, y\} | S(x, z) \geq f_1(S(x, y))\},$$

$$V_x(y) = \{z \in X \setminus \{x, y\} | S(y, z) \geq f_1(S(x, y))\},$$

где $f_1: R \rightarrow R$ – монотонная неотрицательная неубывающая функция. Множества $V_y(x)$ и $V_x(y)$ обозначают множества объектов “похожих” на x и на y соответственно, где значение $f_1(S(x, y))$ служит как критерий этого сходства. Множество

$$V(x, y) = \{z \in X \setminus \{x, y\} | \max\{S(x, z), S(y, z)\} \geq f_2(S(x, y))\},$$

содержит объекты из X , которые “похожи” по крайней мере на один из объектов x и y . Когда $f_1 \equiv f_2$ мы имеем $V(x, y) = V_y(x) \cup V_x(y)$. Это множество будет представлено как множество “соседей” x и y . “Голоса” объектов из $V(x, y)$ будут приниматься во внимание, когда будет приниматься решение о коррекции значения $S(x, y)$. Следующее множество

$$W(x, y) = \{z \in X \setminus \{x, y\} | \min\{S(x, z), S(y, z)\} \geq f_3(S(x, y))\}$$

обозначает множество “сильных” или “общих” соседей, т. е. объекты, которые “похожи” на оба объекта x и y . Объекты из $W(x, y)$ будут “поддерживать” значение $S(x, y)$. Когда $f_1 \equiv f_3$ мы имеем $W(x, y) = V_y(x) \cap V_x(y)$.

Два объекта называются идентичными в S , если $S(x, y) = 1$ и $S(x, z) = S(y, z)$ для всех z из $M \setminus \{x, y\}$. Два объекта x и y называются неразличимыми на уровне $a \in L$, если $S(x, y) \geq a$ и для всех $z \in M$ выполняется $S(x, z) \geq a$ тогда и только, когда $S(y, z) \geq a$. Ясно, что два объекта неразличимые на некотором уровне a будут идентичными в отношении сходства S_a . Ясно также, что все объекты неразличимы на минимально возможном уровне 0 , и максимально возможный уровень неразличимости двух объектов x и y в отношении S равен $a = S(x, y)$. Два объекта x и y неразличимые на уровне $a = S(x, y)$ будут называться неразличимыми в S .

Решение о коррекции значения $S(x, y)$ будет зависеть от соответствующей части объектов “поддерживающих” значение сходства $S(x, y)$, которое может быть представлено как мера неразличимости объектов x и y . Мы можем представить следующие методы вычисления для каждой пары объектов x и y этой соответствующей части обозначенной h_i :

$$h_1 = \frac{|W(x, y)|}{\min\{|V_y(x)|, |V_x(y)|\}}, \quad h_2 = \frac{|W(x, y)|}{\max\{|V_y(x)|, |V_x(y)|\}},$$

$$h_3 = \frac{|W(x, y)|}{|V(x, y)|}, \quad h_4 = \frac{|X \setminus V| + |W| - 2}{|X| - 2},$$

где мы положим $h_i = 1$, если знаменатель равен 0 .

Пусть $L_V(x, y)$ обозначает упорядоченный в убывающем порядке список всех значений $S(x, z)$, $S(y, z)$, ($z \in V$), которые имеют значение меньше чем $S(x, y)$. Обозначим количество элементов в $L_V(x, y)$ как $m = |L_V(x, y)|$ и элементы $L_V(x, y)$ как l_k ($k = 1, \dots, m$). Если $n_L > 1$, то $l_k \geq l_{k+1}$ для всех $k = 1, \dots, m-1$. Порядковое обобщение процедур коррекции $F_j(x, y)$ получено Батыршиным, Шустером [9] и Батыршиным и Хабибуллиным в [10].

Возможные коррекции, когда $m > 1$ определяются параметром j :

$j=1$: $F_j(x, y) = l_m$, т. е. $F_j(x, y)$ минимальное значение в $L_V(x, y)$;

$j=2$: $F_j(x, y) = l_1$, т. е. $F_j(x, y)$ максимальное значение в $L_V(x, y)$;

$j=3$: $F_j(x, y) = (\sum l_k)/m$, т. е. $F_j(x, y)$ есть среднее всех значений из $L_V(x, y)$;

$j=4$: $F_j(x, y) = l_k$, где $k \in \{1, \dots, m\}$ – параметр, F_1 и F_2 это частные случаи F_4 .

$$j=5: F_j(x, y) = \text{median}(L_v(x, y)).$$

Процедура коррекции $F(S) = S_c$ кластерной процедуры Q может быть определена следующим образом:

$$S_c(x, y) = \begin{cases} S(x, y) & \text{если } h_i \geq p, \\ F_j(x, y) & \text{в противном случае.} \end{cases},$$

где $p \in [0, 1]$ - параметр и $F_j(x, y)$ - скорректированное значение, такое что $F_j(x, y) \subseteq S(x, y)$. Мы будем предполагать, что $F_j(x, y)$ зависит от значений $S(x, z)$, $S(y, z)$ объекта z , принадлежащего множествам соседей x и y : $V_y(x)$, $V_x(y)$ и $V(x, y)$.

Все процедуры коррекции при $j=1, 2, 3, 4, 5$ инвариантны относительно нумерации объектов и все $F_j(x, y)$ для $j=1, 2, 4, 5$ инвариантны относительно монотонного преобразования значений сходства.

Когда $p=0$, из $h_i \geq 0$ следует, что $F(S(x, y)) = S_c(x, y) = S(x, y)$, т. е. для всех x, y из X значения $S(x, y)$ не будут корректироваться и $Q(S) = TC(F(S)) = TC(S)$, т. е. кластерный алгоритм будет совпадать с алгоритмом "ближний сосед".

Вместо соответствующей части поддерживающих соседей h_i можно подставить число поддерживающих соседей, которое может быть вычислено как: $g_1 = |W(x, y)|$ или $g_2 = |W(x, y)| + |X \setminus V| - 2$. В этом случае процедура коррекции может быть определена следующим образом

$$S_c(x, y) = \begin{cases} S(x, y) & \text{если } g_i \geq t, \\ F_j(x, y) & \text{в противном случае.} \end{cases},$$

с параметром $t \in \{0, 1, \dots, n-2\}$, $n = |X|$.

Процедура транзитивного замыкания постоянна в представленной схеме кластерных процедур. Эта процедура может быть реализована алгоритмом "ближний сосед" или алгоритмами транзитивного замыкания, представленными в теории графов [8, 6, 7, 11, 12]. Так как процедура транзитивного замыкания инвариантна относительно нумерации объектов и монотонных преобразований функций сходства, представленные кластерные процедуры будут обладать обоими этими свойствами, если процедура коррекции будет обладать этими свойствами. Чтобы быть инвариантной процедура коррекции должна быть применена ко всем парам объектов (x, y) независимо от их нумерации или порядка рассмотрения и параметры, использованные в процедуре коррекции, должны быть фиксированы или не зависеть от нумерации объектов. Свойство инвариантности относительно монотонных преобразований значений сходства является дополнительным к свойству инвариантности относительно нумерации объектов.

Применение метода гибридной кластеризации с визуализацией сильных связей к анализу взаимодействия нефтяных скважин

Рассмотрим метод гибридной кластеризации, который сочетает кластеризацию с визуализацией ассоциативных связей между объектами с помощью ассоциативного графа [13]. Этот метод исходит из того, что структура данных может плохо

представляться в виде классов сходных объектов, формируемого кластерным алгоритмом, и визуализация сильных связей с помощью ассоциативного графа может быть полезна для анализа отклонения полученной кластеризации от структуры данных.

Этот подход применялся к анализу взаимодействия нефтяных скважин одного из Мексиканских месторождений. Рассматривались 9 скважин, расположенных на общей территории, из которых 2 были нагнетающими. Анализ проводился на основе временных рядов объемов помесечной добычи нефти и сопутствующей воды, а также объемов нагнетания воды за 32 месяца, начиная с января 2004 года. Объемы месячной добычи получены в результате суммирования посуточной добычи, поэтому данные не нуждаются в сглаживании, а для анализа ассоциаций применялись скользящие окна размером 2 и 3. В качестве меры ассоциаций временных рядов использовалась мера

$$AM(y, x) = \max_{k \in K} \{AF_k(y, x)\} = \max_{k \in K} (\cos s_k(y, x));$$

где $K = \{2, 3\}$ задает размер скользящих окон.

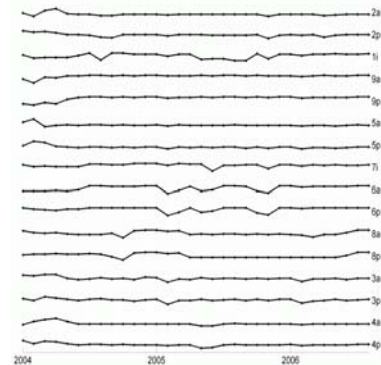


Рис. 1

На рис.1 приведены графики этих временных рядов, где 1i и 7i обозначают временные ряды, соответствующие нагнетающим скважинам, кр и ка означают, соответственно, временные ряды добычи нефти и сопутствующей воды в скважине к.

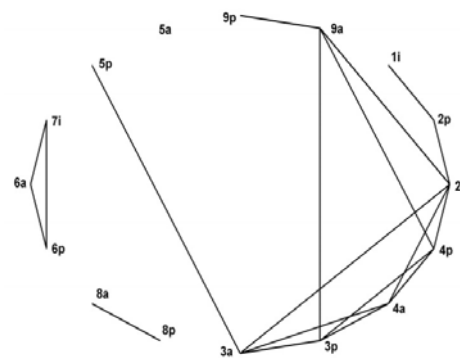


Рис. 2

Рис. 2 дает графическое представление найденного ассоциативного графа между временными рядами, для значения порога 0,468. Это значение равно величине ассоциации между нагнетающей скважиной 1i и добывающей скважиной 2p. Выбор такого порога для представления ассоциативного графа дает

нетривиальное разбиение графа на связные компоненты и позволяет выявить ассоциации между нагнетающими и добывающими скважинами. Вершины ассоциативной сети расположены так, чтобы минимизировать пересечения дуг. В соответствии с этим порядком приведены также временные ряды на рис.1, что позволяет визуально увидеть сходство между формой временных рядов с высоким значением ассоциации локальных трендов (например между 6р и 6а, 3р и 3а, 7i и 6а).

На рис. 2 представлены результаты гибридной кластеризации скважин на основе меры ассоциаций локальных трендов. Две скважины X и Y считаются взаимосвязанными, если существует значимая ассоциативная связь между хотя бы одной парой временных рядов (X_a, Y_a) , (X_a, Y_p) , (X_p, Y_a) , (X_p, Y_p) . Получено следующее разбиение скважин на 5 кластеров: $\{1i\}$, $\{2, 3, 4, 9\}$, $\{5\}$, $\{6, 7i\}$, $\{8\}$. На рис. 3 одиночные кластеры помечены кружком, а скважины из кластеров $\{2, 3, 4, 9\}$ и $\{6, 7i\}$ квадратом и ромбом соответственно. Значимые ассоциативные связи между скважинами (5, 3 и 1i, 2), принадлежащим разным кластерам, показаны пунктиром.

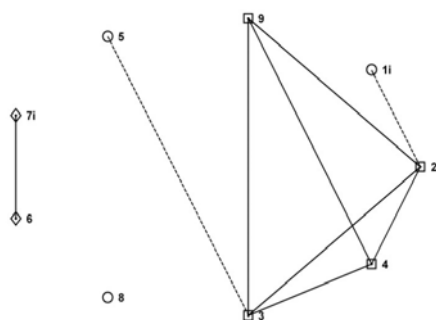


Рис. 3

Интерпретация полученных результатов позволяет судить о характере взаимодействия между скважинами рассматриваемого месторождения. В частности, установлено хорошее взаимодействие между нагнетающей скважиной 7 и добывающей скважиной 6, объединенными в один кластер, что может объясняться высокой проницаемостью пластов в окрестности этих скважин. Отсутствие значимого взаимодействия между скважиной 8, попавшей в одиночный кластер, и другими скважинами, может характеризоваться разными причинами, и ввиду этого служить основой для дополнительного анализа свойств пород вокруг этих скважин. Объединение скважин 2, 3, 4, 9 в один кластер также может служить основой для выдвижения гипотез о причинах значимого взаимодействия между скважинами: высокой проницаемости пород в окрестности этих скважин, наличия трещин или сдвигов пластов в окрестности этих скважин, и др. Результаты гибридной кластеризации дают основы для выдвижения различных гипотез о характере взаимодействия между скважинами и в результате этого служить основанием для проведения дополнительных специфических исследований на скважинах и в

месторождении, а также для изменения режимов закачки воды в скважины.

Заключение

В работе предложен метод анализа взаимодействия между нефтедобывающими и нагнетающими скважинами. Этот метод основан на анализе временных рядов объемов добычи и нагнетания флюидов в скважинах исследуемого месторождения. Предложенный в работе метод не требует проведения специальных экспериментов или измерений на скважинах, а использует данные, имеющиеся в наличии у нефтедобывающих компаний. Для анализа ассоциаций между временными рядами используется метод ассоциации локальных трендов на основе преобразования скользящих аппроксимаций [14, 15]. Этот метод является инвариантным отношению к линейным преобразованиям временных рядов, позволяет измерять положительные и отрицательные ассоциации между временными рядами, позволяет исследовать ассоциации для различных временных интервалов, например исследовать краткосрочные и долгосрочные ассоциации. Метод сочетает кластерный анализ с дополнительной информацией о структуре данных и их возможной визуализации. В основе лежат реляционные процедуры кластеризации [3, 16], обладающие рядом свойств инвариантности к преобразованию исходных данных.

Литература

1. Плохотников С.П. Осреднение моделей трехфазной фильтрации в неоднородных слоях, подчиняющихся равномерному распределению / С.П. Плохотников, В. А. Богомолов, Белова Е.Н., Богомолова О.И., Плохотников Д.С. Булгакова О.Р.//Вестник Казанского технологического университета. – Казань, 2012. - Т. 1. - С.63–68.
2. Плохотников С.П. Модифицированные проницаемости, основанные на модели схемы струй для бета-распределения и его частных случаев / С.П. Плохотников, В. А. Богомолов, Плохотников Д.С. Булгакова О.Р.//Вестник Казанского технологического университета. – Казань, 2010. - Т. 4. - С.171–175.
3. Batyrshin I.Z. On general scheme of invariant clustering procedures based on fuzzy similarity relation/Batyrshin I.Z., Rudas T., Klimova A.S.//Proceedings International Conference on Fuzzy Sets and Soft Computing in Economics and Finance. - Saint-Petersburg, Russia, 2004. - P. 122 – 129.
4. Батыршин И. З. Инвариантные иерархические алгоритмы кластеризации, основанные на коррекции значений сходства/ И. З. Батыршин, А. С. Климова//Материалы III-го Международного научно-практического семинара “Интегрированные модели и мягкие вычисления в искусственном интеллекте” – 2005. - С. 119 – 125.
5. Батыршин И. З. Структура семантического пространства словесных оценок поступков/ И. З. Батыршин, В. А. Шустер//Принципиальные вопросы теории знаний. Труды по искусственному интеллекту. Ученые записки Тартуского гос. ун-та. Вып. 688 – Тарту, 1984. – С. 20 – 38.
6. Батыршин И. З. Методы анализа систем на основе взвешенных отношений//Диссертация на соискание ученой степени кандидата технических наук - Москва: МЭИ, 1982 – 218 С.

7. S. Tamura Pattern classification based on fuzzy relations/ S. Tamura, S. Higuchi, K. Tanaka//IEEE Trans. SMC. – 1971, vol. SMC - 1, P. 61 – 66.
8. J. C. Dunn A graph-theoretic analysis of pattern classification via Tamura's fuzzy relation//IEEE Trans. SMC. – 1974, vol. SMC – 4, P. 310 – 313.
9. Батыршин И. З. Структура семантического пространства словесных оценок поступков/Батыршин И. З., Шустер В. А.//Принципиальные вопросы теории знанийТруды по искусственному интеллекту. Ученые записки Тартуского гос. ун-та. Вып. 688 – Тарту, 1984. – С. 20 – 38.
10. J. C. Hubert Monotone invariant clustering procedures//Psychometrika. – 1973, vol. 38, №1, P. 47 – 62.
11. C.J. Jardine The structure and construction of taxonomic hierarchies/C.J. Jardine, N. Jardine, R. Sibson//Math. Biosci. – 1967. – 1 - P. 173 - 179.
12. I. Bатыршин. On invariance of clustering algorithms./ I. Bатыршин I., Khabibulin R.- The Journal of Fuzzy Mathematics, Vol. 6, No. 3, 1998, 721-733.
13. А. Ахо Построение и анализ вычислительных алгоритмов/ А. Ахо, Дж. Хопкрофт, Дж. Ульман – М.: Мир, 1979. – 536 С.
14. Батыршин И.З. Анализ взаимодействия нефтяных скважин на основе гибридной кластеризации временных рядов продуктивности скважин/ Батыршин И.З., Кошульски А., Шереметов Л.Б., Климова А.С., Панова А.М.//Нечёткие системы и мягкие вычисления, том 2, №4 – 2007. - С. 63-73.
15. Bатыршин I. Moving Approximation Transform and Local Trend Associations in Time Series Data Bases/ Bатыршин I., Herrera-Avelar R., Sheremetov L., Panova A.//In: Perception-Based Data Mining and Decision Making in Economics and Finance. Studies in Computational Intelligence, Springer, vol. 36 – 2007. –С. 55-83.
16. Bатыршин I. Moving approximations in time series data mining/ Bатыршин I., Herrera-Avelar R., Sheremetov L., Suarez R.//In.: Proc. Intern. Conf. on Fuzzy Sets and Soft Computing in Economics and Finance, FSSCEF 2004.- St. Petersburg, Russia, vol. I, 2004. – С.62-72.
17. Bатыршин I.Z. On general scheme of invariant clustering procedures based on fuzzy similarity relation/ Bатыршин I.Z., Rudas T. , Klimova A.//Proc. Intern. Conf. Fuzzy Sets and Soft Computing in Economics and Finance FSSCEF 2004. - St. Petersburg, Russia, vol. I, 2004. – P. 122-129.

© **А. С. Климова** - ст. препод. каф. ИПМ КНИТУ, klangelica@yandex.ru; **И. З. Батыршин** – д-р физ.-мат. наук, проф., Research Program in Applied Mathematics and Computing Mexican Petroleum Institute, batyr1@yandex.ru; **Н. К. Шайдуллина** – асс. каф. ИПМ КНИТУ, fenakon@yandex.ru