

УДК 519.254

ОБ ИСПОЛЬЗОВАНИИ ПАРАЛЛЕЛЬНЫХ ВЫЧИСЛЕНИЙ ДЛЯ АНАЛИЗА РАСТУЩИХ СИСТЕМ

П.Ф. Тарасенко

Томский научный центр СО РАН
Томский государственный университет
E-mail: ptara@mail.ru

Приводятся математические модели растущих систем, для которых рассматриваются возможности применения параллельных вычислений с использованием стандарта MPI при организации отбора вариантов возможного состава растущих систем. Предложен метод, адаптивно обеспечивающий равномерность загрузки параллельных процессов.

Ключевые слова:

Растущие системы, квантильная регрессия, параллельные вычисления, множество Парето.

Введение

В общем смысле под растущей системой понимается природная совокупность объектов, между которыми в течение длительного времени в стабильных условиях происходит распределение некоторого ресурса. В результате объемы этого ресурса, накопленные в объектах, подчиняются вероятностным закономерностям [1]. Впервые наличие закономерностей подобного рода подметил Ф.Н. Алексеев [2].

В качестве примеров растущих систем рассматриваются процессы накопления углеводов в совокупности месторождений и залежей, динамику перераспределения населения по административным территориям, капитала по отраслям, массы вещества по планетам, биомассы дерева по его частям и многое другое.

Вероятностные закономерности, присущие растущим системам, позволяют по наблюдаемым объемам накопленных ресурсов отвечать на ряд вопросов. Например, судить о механизме накопления, выделять совокупности объектов, принадлежащих одной растущей системе и даже прогнозировать ненаблюдаемые объекты растущей системы и объемы запасов ресурсов в них.

Решение последней задачи при применении закономерностей накопления к совокупности залежей углеводов общего генезиса позволяет [3] по уже разведанным запасам оценивать перспективность данной природной совокупности и строить прогнозы о количестве неоткрытых залежей и величинах их запасов.

Методы решения задачи прогнозирования [4] основаны на переборе и оценке большого количества вариантов прогнозов, что сопряжено с большими вычислительными затратами. Несмотря на использование способов сокращения перебора, построение прогнозов остается неприемлемо трудоемкой задачей, если речь идет о большом количестве объектов совокупности.

В то же время важной особенностью методики прогнозирования является участие эксперта по предметной области, который должен не только оценивать полученные варианты прогнозов, сопоставлять их с другими данными, но и выдвигать предположения о составе природной совокупности. Для проверки этих предположений необходимо строить новые прогнозы, анализ которых в свою очередь приводит к коррекции знаний о составе природной совокупности, к изменению критериев качества прогнозов или к изменению других условий. Здесь мы имеем дело с типичной схемой поддержки принятия решений, когда скорость получения результата расчетов играет важную роль в процессе выдвижения новых предположений, пока эксперт помнит все нюансы, заставившие его проверить предыдущее предположение.

В данной работе рассматриваются возможности использования параллельных вычислений для ускорения процесса построения прогнозов на основе математических моделей растущих систем.

1. Модель накопления ресурсов в растущих системах

Если природная совокупность состоит из N объектов, то согласно модели накопления [1] величины запасов ресурсов Q_1, \dots, Q_N каждого объекта есть квантили равноотстоящих уровней, которые соответствуют некоторой функции распределения: $F((Q_k - M)/\sigma) = P_k$, $k=1, \dots, N$ при некоторых неизвестных параметрах M и σ . Из этих соотношений получаем уравнения

$$Q_k = F^{-1}(P_k)\sigma + M, \quad k=1, \dots, N. \quad (1)$$

Функции распределения F в [5] предложено называть стандарт-функциями. В зависимости от механизма накопления ресурсов в качестве стандарт-функции могут выступать [1] функции распределения нормального, логнормального, двойного показательного или равномерного закона. Для обозначения типа стандарт-функции введем индекс I_p , который может принимать одно из четырех значений.

В качестве равноотстоящих уровней P_k могут использоваться уровни Гумбеля $P_k = k/(N+1)$ с шагом $1/(N+1)$, уровни Блома $P_k = (8k-3)/(8N+2)$ с шагом $1/(N+1/4)$ или уровни Крамера $P_k = (2k-1)/(2N)$ с шагом $1/N$. Для обозначения типа уровней будем использовать индекс I_p , который может принимать одно из трех значений.

Учитывая, что величины ресурсов могут изменяться с ошибками, приходим к уравнениям:

$$y_k = \sigma x_k + M + \varepsilon_k, \quad k=1, \dots, N, \quad (2)$$

где $x_k = F^{-1}(P_k)$ – значения независимой переменной (квантили равноотстоящих уровней стандарт-функции F); y_k – значения зависимой переменной (измеренные величины запасов); ε_k – случайные погрешности, порождаемые неточностью измерений. При $E\varepsilon_k = 0$ уравнения (2) трактуются как модель регрессии. При $Med\{\varepsilon_k\} = 0$ уравнения (2) трактуются [6] как квантильная регрессия уровня $1/2$. В зависимости от этого параметры модели (2) могут оцениваться разными способами – методом наименьших квадратов, методом наименьших модулей или знаковым методом. Для обозначения этих вариантов оценивания введем индекс I_E , который может принимать столько различных значений, сколько различных методов оценивания используется.

2. Построение прогнозов

Ситуация, в которой необходимо строить прогноз, характеризуется тем, что неизвестно, все ли N объектов природной совокупности наблюдаются, т. е. величина N неизвестна. В связи с этим при идентификации модели (2) приходится иметь дело с моделью наблюдений, которая отличается от обычной модели регрессии. Это проявляется в следующих существенных моментах. *Во-первых*, значения x_1, \dots, x_N из набора независимых переменных модели регрессии зависят от неизвестного объема N . *Во-вторых*, число n наблюдаемых зависимых переменных меньше или равно числу N независимых переменных. *В-третьих*, неизвестно парное соответствие между наблюдениями зависимых (Y_1, \dots, Y_n) и независимых (x_1, \dots, x_N) переменных. Таким образом, когда мы имеем дело с наблюдениями Y_1, \dots, Y_n , задача сводится к нахождению числа N , набора чисел $i(1), \dots, i(n)$, которые ставят в соответствие Y_k и $x_{i(k)}$, образуя таблицу парных наблюдений $\{(X_k = x_{i(k)}, Y_k = y_{i(k)}), k=1, \dots, n\}$, по которой оцениваются параметры регрессии σ и M тем или иным методом в соответствии с моделью

$$Y_k = \sigma X_k + M + \varepsilon_{i(k)}, \quad k=1, \dots, n. \quad (3)$$

Варианты решения поставленной задачи отличаются друг от друга набором условий $C_n = \{N, i(1), \dots, i(n), I_F, I_P, I_E\}$. Каждому из этих вариантов соответствуют свои оценки $\hat{\sigma}$, и прогнозы ресурсов $\hat{Y}_k = \hat{\sigma} x_{i(k)} + \hat{M}$, $k=1, \dots, N$, среди которых первые n значений относятся к наблюдаемым объектам, а остальные – к ненаблюдаемым. Набор $A_n = \{C_n, \hat{\sigma}, \hat{M}, Y_1, \dots, Y_n\}$ можно отождествить с одним из альтернативных вариантов прогноза, который для краткости будем далее называть просто *альтернативой*. Набор условий C_n будем называть *альтернативными условиями*.

Для оценивания неизвестных параметров регрессии σ и M могут быть использованы различные методы. Например, метод наименьших квадратов (МНК). Однако в нашей задаче применение МНК

не всегда можно считать оправданным. Дело в том, что оценки МНК не являются устойчивыми к так называемым выбросам наблюдений, которым часто подвержены погрешности измерения величин ресурсов. В связи с этим наряду с МНК используется знаковый метод оценивания [6], который обеспечивает устойчивость к отклонениям от гауссовской модели, в том числе и к выбросам. В простейшем варианте оценки параметров регрессии ищутся из условия минимума следующей знаковой статистики по σ и M :

$$\left[\sum_{k=1}^n X_k \text{sign}(Y_k - \sigma X_k - M) \right]^2 + \left[\sum_{k=1}^n \text{sign}(Y_k - \sigma X_k - M) \right]^2. \quad (4)$$

Данная целевая функция является ступенчатой, поэтому ее оптимизация представляет собой довольно сложную вычислительную задачу, требующую использования специальных методов, которые в данной работе не рассматриваются. Еще одним вариантом устойчивого оценивания является метод наименьших модулей [7], который предполагает минимизацию статистики

$$\sum_{k=1}^n |Y_k - \sigma X_k - M| \rightarrow \min_{\sigma, M}. \quad (5)$$

Оценки по методу наименьших модулей также устойчивы к выбросам, но по сравнению со знаковыми оценками все же более подвержены влиянию аномальных отклонений при малых объемах наблюдений. С вычислительной точки зрения задача минимизации (5) сводится к задаче линейного программирования, поэтому является менее трудоемкой по сравнению с минимизацией (4).

3. Организация перебора альтернативных условий

Сложность поставленной задачи прогнозирования состоит в том, что здесь в рамках конкретной альтернативы по-разному взаимодействуют погрешности, порожденные несколькими качественно различающимися причинами. Это погрешности измерений ε_k , погрешность определения числа N , ошибки при выборе набора $i(1), \dots, i(n)$, типа стандарт-функции и типа квантилей равноотстоящих уровней.

В этих условиях один из подходов к прогнозированию заключается в том, чтобы перебирать возможные альтернативы, выявляя при этом наилучшие варианты прогнозов с помощью некоторого набора критериев. При фиксированных I_F, I_P, I_E порядок перебора вариантов следующий. Для каждого нового значения N перебираются возможные наборы индексов $i(1), \dots, i(n)$, в соответствии с которыми оцениваются параметры регрессии σ , M и вычисляются прогнозные значения Y_k , $k=1, \dots, N$. Наилучшие варианты прогноза отбираются исходя из критериев качества, которые выражают степень соответствия восстановленной зависимости с теоретической закономерностью (1).

Будем считать (без ограничения общности), что набор наблюдаемых характеристик упорядочен по возрастанию, т. е. $Y_1 < \dots < Y_n$. Кроме этого, согласно определению, $x_1 < \dots < x_N$. Мы будем рассматривать только монотонные соответствия между ними, для которых $i(1) < \dots < i(n)$, поскольку практически все

критерии, используемые для определения качества восстановления зависимости (3), будут заведомо давать худшие результаты для соответствий, которые отличаются от монотонных. Использование только монотонных соответствий приводит к тому, что при фиксированных N и n число различных вариантов соответствий, подлежащих рассмотрению, будет равно числу сочетаний C_N^n . При большом количестве ненаблюдаемых объектов это число сочетаний достаточно велико, поэтому разумно прибегать к методам сокращенного перебора [4].

К сожалению, при работе с растущими системами, которые в своем составе имеют большое количество объектов, даже при сокращении перебора не удастся добиться приемлемого времени построения прогнозов, чтобы удовлетворить требованиям, предъявляемым к системам поддержки принятия решений. Это и вызывает необходимость использования высокопроизводительных вычислений.

4. Критерии качества альтернатив

Критерий качества каждой альтернативы призван характеризовать степень близости наблюдаемой и теоретической закономерностей, показывать, насколько хорошо наблюдаемые точки (X_k, Y_k) приближаются к линии восстановленной регрессии. Количественно измерять степень этой близости можно различными способами, поэтому в [4] предложено и обсуждено несколько различных групп критериев.

Первая группа – критерии, основанные на характеристиках зависимости, даваемых методом наименьших квадратов. К ним относятся: выборочный коэффициент корреляции; выборочный коэффициент детерминации; достигнутый уровень значимости в критерии Фишера (значимость зависимости).

Вторая группа критериев основана на абсолютных $\Delta_k = |Y_k - \hat{Y}_k|$ и относительных $\delta_k = \Delta_k / Y_k$ ошибках предсказания наблюдаемых значений Y_k с помощью модели (2). Такими критериями являются: средняя относительная или абсолютная ошибка предсказания ($\sum_{k=1}^n \delta_k / n$ или $\sum_{k=1}^n \Delta_k / n$); максимальная относительная или абсолютная ошибка предсказания ($\max_{k \leq n} \delta_k$ или $\max_{k \leq n} \Delta_k$); среднеквадратическая относительная или абсолютная ошибка предсказания ($\sqrt{\sum_{k=1}^n \delta_k^2 / n}$ или $\sqrt{\sum_{k=1}^n \Delta_k^2 / n}$).

Третья группа критериев основана на характеристиках зависимости, полученных с использованием знаковых методов восстановления регрессии: площадь доверительной области для параметров регрессии; размах этой доверительной области по направлению параметра наклона σ ; размах доверительной области по направлению параметра смещения M . Следует отметить, что вычислительные затраты для получения этих критериев выше, чем при вычислении критериев первых двух групп, но именно их использование позволяет выявлять истинные альтернативы при наличии выбросов.

Результатом оценки альтернативы является некоторый вектор $Q(A_n)$, который содержит значения вычисленных критериев.

5. Отбор альтернатив

Рассмотрим задачу выбора лучших вариантов прогноза при фиксированном гипотетическом объеме природной совокупности. При наличии многих критериев процесс отбора наилучшей альтернативы усложняется, т. к. по разным критериям наилучшими могут быть различные альтернативы. В этом случае самым распространенным подходом является построение множества Парето. Метод Парето позволяет исключить из всего множества альтернатив заведомо худшие по всем показателям (критериям) и выбрать для дальнейшего рассмотрения конкурирующие между собой альтернативы. В основе метода Парето лежит понятие недоминируемой альтернативы. Говорят, что одна альтернатива доминирует над другой, если она лучше нее хотя бы по одному критерию и не хуже по остальным. Множество недоминируемых альтернатив называют множеством Парето.

Для построения множества Парето не обязательно сравнивать каждую альтернативу с каждой. Для формального описания процедуры обозначим через $D_1(Y|x, Q)$ сужение множества Парето Y , в которое входят альтернативы, доминируемые альтернативой x по набору критериев Q . Пусть также $D_2(Y|x, Q)$ – множество, состоящее из альтернативы x , если она доминируется хотя бы одной альтернативой из Y . Очевидно, что $D_1(Y|x, Q)$ и $D_2(Y|x, Q)$ не могут быть непустыми одновременно. Пусть $S_1(X|Q)$ – множество Парето, полученное из множества альтернатив $X = \{x_1, \dots, x_K\}$ с применением набора критериев Q . В этих обозначениях пошаговая процедура вычисления $P_k = S_1(X|Q)$ запишется в виде

$$P_0 = \emptyset; P_k = [P_{k-1} \setminus D_1(P_{k-1}|x_k, Q)] \cup [\{x_k\} \setminus D_2(P_{k-1}|x_k, Q)], \\ k=1, \dots, K, \quad (6)$$

при этом порядок предъявления альтернатив не сказывается на конечном результате.

Работа с множествами Парето на реальных и модельных данных в задаче прогнозирования растущих систем показала, что часто в силу искажения исходных данных истинные альтернативы не входят во множество Парето, но лежат близко к нему в пространстве критериев. В связи с этим в [4] предложена концепция многослойного множества Парето. Эта концепция базируется на следующем простом соображении. Если из конечного множества всех предъявляемых альтернатив исключить недоминируемые, то среди оставшихся вновь можно искать множество Парето. И так можно поступать столько раз, сколько потребуются для отбора нужных (заслуживающих внимания эксперта) альтернатив. В результате появляется несколько слоев недоминируемых альтернатив.

Следуя терминологии теории множеств, будем называть многослойное множество *классом Парето*

то. Под *объемом* класса Парето P будем понимать число его слоев, и обозначать его через $M(P)$. Операцию построения класса Парето объема M для множества альтернатив X обозначим через $S_M(X|Q)$, где Q – использованный набор критериев. Пусть слои класса Парето упорядочены и пронумерованы по отношению доминирования входящих в них альтернатив. Через $L_m(P)$ обозначим m -й слой класса Парето P , $m=1, \dots, M(P)$.

По определению, если $P=S_M(X|Q)$, то $L_m(P)=S_1(X \setminus \bigcup_{k=1, \dots, m-1} L_k(P)|Q)$. Однако для того, чтобы получить новый слой класса Парето, не обязательно перебирать заново все оставшиеся альтернативы. Достаточно использовать рекурсивную процедуру, когда при предъявлении очередной альтернативы делается попытка поместить в следующий слой те альтернативы, которые вытеснены из предыдущего слоя. Для формализации введем множество альтернатив $G_1(Y|x, Q)=D_1(Y|x, Q)+D_2(Y|x, Q)$, которые «отсеиваются» из множества Парето Y при попытке добавить в него альтернативу x . Через $G_2(Y|x, Q)=[Y \setminus D_1(Y|x, Q)]+[x] \setminus D_2(Y|x, Q)$ обозначим оставшиеся после этого альтернативы. Пусть $G_1(Y|x, Q)$ и $G_2(Y|x, Q)$ – «отсеянные» и оставшиеся в Y альтернативы после последовательного добавления альтернатив из множества X . Тогда, если $X=\{x_1, \dots, x_K\}$, то класс Парето $P_K=S_M(X|Q)$ дает следующая процедура, которая при $M=1$ совпадает с (6):

$$L_m(P_0)=\emptyset, L_m(P_k)=G_2(L_m(P_{k-1})|X_{m,k}, Q), m=1, \dots, M, \\ k=1, \dots, K; \quad (7) \\ X_{1,k}=\{x_k\}, X_{m+1,k}=G_1(L_m(P_{k-1})|X_{m,k}, Q).$$

В любом случае отбор каждой альтернативы в классе Парето остается переборной задачей, которая при большом количестве отобранных альтернатив является достаточно трудоемкой. Результатом отбора альтернатив в задаче прогнозирования растущих систем является набор классов Парето, каждый из которых соответствует одному из возможных значений $N \in \{n, n+1, \dots, n+K\}$, где K выбирается как максимально возможное число ненаблюдаемых объектов. Альтернативы из этих множеств подлежат анализу со стороны эксперта.

6. Информационная структура алгоритма прогнозирования

Для представленного выше алгоритма прогнозирования на самом верхнем уровне его описания можно выделить последовательные этапы, которые выполняются для каждой альтернативы.

1. Генерирование альтернативных условий $C_n=\{N, i(1), \dots, i(n), I_p, I_p, I_E\}$ (см. раздел 3).
2. Вычисление альтернативы $A_n=\{C_n, \hat{\sigma}, \hat{M}, \hat{Y}_1, \dots, \hat{Y}_N\}$ (см. раздел 2).
3. Критериальная оценка альтернативы $Q(A_n)$ (см. раздел 4).
4. Отбор альтернатив в класс Парето (см. раздел 5).

Можно говорить о распараллеливании внутри каждого этапа, однако и без этого у данной задачи

имеется ряд особенностей, которые позволяют за счет блочно-циклического распараллеливания обеспечить основное требование к параллельным алгоритмам – равномерность загрузки большого количества процессов:

- 1) Главная особенность состоит в том, что при заданном значении N трудоемкость 2 и 3 этапов значительна, но практически одинакова для всех альтернатив. Поэтому вычисление и критериальная оценка разных альтернатив могут быть поручены всем освобождающимся процессам.
- 2) Трудоемкость первого этапа тоже одинакова для всех альтернатив, но она в сотни (а при больших n – до тысячи) раз меньше трудоемкости 2 и 3 этапов. Поэтому первый этап может быть поручен только одному из процессов, даже если общее число процессов, занятых 2 и 3 этапами, достигает нескольких сотен.
- 3) Основная трудность состоит в том, что трудоемкость четвертого этапа зависит не только от количества критериев и числа слоев множества Парето, но и от качества самой альтернативы, а также от числа и качества альтернатив, ранее отобранных в текущий класс Парето. Поэтому по альтернативам эта трудоемкость распределена очень неравномерно. При большом количестве процессов, занятых 2 и 3 этапами, эта неравномерность сглаживается, но из-за того, что альтернативы плохого качества чаще встречаются в начале и в конце перебора, а объем текущего множества Парето меняется немонотонно, выполнение четвертого этапа может стать узким местом даже при организации очередей сообщений от процессов, оценивающих альтернативы. Поэтому необходимо предусмотреть возможности распараллеливания четвертого этапа алгоритма. При этом желательно, чтобы такое распараллеливание происходило адаптивно, в зависимости от сложившейся ситуации.

7. Параллельный вариант алгоритма прогнозирования

Описание параллельного варианта алгоритма будем вести в терминах стандарта MPI [8, 9] с использованием понятий процесса, индивидуальных синхронных и асинхронных сообщений. Прежде всего, определимся с типами процессов по выполняемым задачам:

- 1) Единственный процесс первого типа генерирует альтернативные условия $C_n=\{N, i(1), \dots, i(n), I_p, I_p, I_E\}$ и отправляет об этом сообщения процессам второго типа. Для ускорения отбора альтернатив, к альтернативным условиям добавляется их идентификатор, который может служить хэш-кодом.
- 2) Процессы второго типа принимают сообщение о новых альтернативных условиях C_n , занимаются вычислением альтернатив $A_n=\{C_n, \sigma, M, Y_1, \dots, Y_N\}$ и критериев $Q(A_n)$, а также отправляют сообщения с альтернативами и критериями процессам третьего типа. После освобождения от обработки

предыдущей альтернативы они принимают сообщение о новых альтернативных условиях. Предполагается, что исходные данные Y_1, \dots, Y_n были получены этими процессами на этапе инициализации.

- 3) Процессы третьего типа отвечают за отбор альтернатив в класс Парето, выполняя процедуру (7). Среди них обязательно присутствует один процесс, который делает окончательный отбор. С целью распараллеливания отбора по мере необходимости в эту категорию могут временно переходить процессы второго типа, чтобы стать вспомогательными и заниматься промежуточным отбором альтернатив.

Распараллеливание отбора альтернатив основано на следующих свойствах классов Парето:

$$S_M(X_1 + X_2 | Q) = S_M(S_M(X_1 | Q) + X_2 | Q); \quad (8)$$

если $L(x, S_M(X_1 | Q)) = m$,

$$\text{то } L(x, S_M(S_M(X_1 | Q) + X_2 | Q)) \geq m, \quad (9)$$

где $L(x, P)$ – номер слоя в классе Парето P , которому принадлежит альтернатива x . Если x не входит в P , то $L(x, P) = M(P) + 1$. Благодаря (8) взаимодействие между процессами третьего типа можно организовать в виде дерева, в корне которого находится основной процесс, производящий окончательный отбор (припишем ему тип 3с), в терминальных узлах находятся процессы, принимающие сообщения от процессов второго типа (припишем им тип 3а). Промежуточные места в этом дереве занимают процессы, которые строят классы Парето от классов Парето (таким процессам припишем тип 3б). Они сообщают о новых или о перемещенных альтернативах (в своих классах Парето) на следующий уровень дерева. При этом сообщение о перемещении должно содержать только идентификатор альтернативы и номер ее нового уровня. Свойство (9) позволяет процессам типа 3б и 3с использовать этот номер, чтобы сократить процедуру (7).

В вычислительной системе с однородными по производительности процессорами сообщения от процесса типа 1 процессам типа 2 могут отправляться циклически по порядку номеров и даже быть синхронными, т. к. трудоемкость задач, решаемых одновременно процессами 2 типа, приближенно одинакова. Остальные сообщения для обеспечения адаптивности должны быть асинхронными, что позволяет периодически контролировать, как изменяются длины очередей к процессам 3 типа и за счет этого управлять использованием процессов. При превышении определенного верхнего порога длины очереди один из процессов 2 типа должен завершить свое задание и занять соседнее место в дереве процессов 3 типа, разделив очередь, класс Парето и отправителей перегруженного процесса. Если длина очереди становится меньше некоторого нижнего порога, то процесс должен разделить свою очередь и отправителей со своими соседями и получателем, разделить свой класс Парето с соседями и перейти обратно во 2 тип.

Для поддержания обмена индивидуальными сообщениями необходимо, чтобы процессы хранили номера получателей сообщений. Поэтому до и после реорганизации дерева процессов 3 типа, должны быть разосланы синхронные сообщения об изменениях в топологии обмена, а между этими сообщениями процессы, меняющие свой тип, должны завершить старые обязанности и приступить к новым. Причем сопутствующие этому сообщения могут быть асинхронными.

Разделение нагрузки между процессами достигается за счет того, что доминируемые альтернативы отсеиваются на предыдущих уровнях дерева процессов третьего типа, а недоминируемые альтернативы сопровождаются дополнительной информацией о своем уровне в классе Парето. В связи с этим при изменениях топологии обмена должны оптимизироваться определенные критерии качества затронутого поддерева, связанные с эффективностью распараллеливания (с учетом сбалансированности проектируемого дерева, а также равномерностью распределения проектируемых длин очередей). Поэтому для управления этими действиями целесообразно выделить один процесс, который будет выполнять диспетчерские функции.

При реализации такой адаптивной процедуры возникают дополнительные накладные расходы. К ним относятся расходы на взаимодействие между процессами третьего типа, на реорганизацию дерева процессов третьего типа, на контроль длин очередей, затраты оперативной памяти на повторное хранение альтернатив на разных уровнях этого дерева, а также на хранение идентификаторов асинхронных сообщений, отправленных процессам 3 типа. Однако расходы времени на реорганизацию могут контролироваться за счет выбора пороговых значений длин очередей, а расход времени на контроль длин очередей регулироваться выбором периодичности проверок.

8. Заключение

В работе рассмотрены методы анализа математических моделей растущих систем, которые позволяют использовать мегакластеры высокопроизводительных вычислений для ускорения отбора вариантов возможного состава растущих систем. Приведена информационная структура алгоритма прогнозирования и рассмотрен вариант его параллельной организации. Результаты могут быть использованы при создании и совершенствовании автоматизированных систем поддержки принятия решений в геологическом прогнозировании. Это дает возможность эксперту по предметной области в реальном масштабе времени анализировать варианты прогнозов, сопоставлять их с другими данными, выдвигать и проверять предположения о составе природных совокупностей месторождений полезных ископаемых.

СПИСОК ЛИТЕРАТУРЫ

1. Дмитриев Ю.Г., Устинов Ю.К. Математические модели растущих систем // Вычислительные технологии. – 2007. – Т. 12. – Спец. вып. № 1. – С. 68–75.
2. Алексеев Ф.Н. О распределении запасов рудных месторождений // Доклады АН СССР. – 1978. – Т. 240. – № 2. – С. 387–388.
3. Алексеев Ф.Н., Ростовцев В.Н. Теория образования месторождений полезных ископаемых и практика ее применения. – Томск: STT, 2004. – 315 с.
4. Дмитриев Ю.Г., Тарасенко П.Ф. Автоматизированная система «Октава» для геологического прогнозирования // Вычислительные технологии. – 2003. – Т. 8. – Спец. вып. – С. 74–91.
5. Алексеев Ф.Н. Теория накопления и прогнозирования запасов полезных ископаемых. – Томск: Изд-во Том. ун-та, 1996. – 175 с.
6. Тарасенко П.Ф. Знаковые процедуры анализа растущих систем // Вычислительные технологии. – 2007. – Т. 12. – Спец. вып. № 1. – С. 76–85.
7. Koenker R., Bassett G. Regression quantiles // *Econometrica*. – 1978. – V. 46. – № 1. – P. 33–50.
8. Корнеев В.Д. Параллельное программирование в MPI. – Новосибирск: Изд-во СО РАН, 2002. – 214 с.
9. Антонов А.С. Введение в параллельные вычисления. – М.: НИВЦ МГУ, ТЕИС, 2002. – 70 с.

Поступила 27.10.2008 г.

УДК 550.053.510.2+550.053.681.3(571.16)

АНАЛИЗ ФУНКЦИИ КАЧЕСТВА АЛГОРИТМОВ ФАЗОЧАСТОТНОГО ПРОСЛЕЖИВАНИЯ СЕЙСМИЧЕСКИХ ВОЛН

В.П. Иванченков, А.И. Кочегуров, О.В. Орлов

Томский политехнический университет
E-mail: kai@cc.tpu.edu.ru

Рассмотрено обобщенное определение функции качества при фазочастотном прослеживании сейсмических волн. Показана высокая разрешающая способность алгоритма фазочастотного прослеживания, а также наличие связи функции качества с петрофизическими параметрами сейсмических сред.

Ключевые слова:

Сейсмическая трасса, фазочастотное прослеживание, функция качества.

1. Определение функции качества

Прослеживание сейсмических волн является одной из центральных задач структурной сейсмологии. Не менее важное значение оно имеет и для решения задач прогноза геологического разреза, в том числе прогноза залежей углеводородов.

В настоящее время для анализа волновых сейсмических полей широко применяются динамические параметры волн, связанные с амплитудой и энергией отражений. В значительно меньшей степени используются фазочастотные характеристики (ФЧХ) отраженных волн. Между тем, в фазовых спектрах сейсмических сигналов заложена важная информация о местоположении отражающих границ, поглощающих и дисперсионных свойствах слоистых сред [1–3]. На этой основе построены фазочастотные алгоритмы обработки сейсмических данных, позволяющие в условиях априорной неопределенности относительно формы сейсмических волн обнаруживать и разрешать сигналы на фоне интенсивных помех [4–6].

В [4] был предложен оптимальный фазочастотный алгоритм прослеживания сейсмических волн на фоне гауссовых помех, который реализуется как

процедура поиска положения максимума функции правдоподобия вида:

$$G(\tau) = \sum_{i=1}^n \gamma(f_i) \cos[\Delta\phi(f_i) - 2\pi f_i \tau]. \quad (1)$$

В данном выражении $\gamma(f_i)$ определяет отношение сигнал к шуму, а $\Delta\phi(f_i)$ – отклонение фазового спектра сигнала от фазового спектра смеси сигнала и шума на частоте f_i .

При практическом применении оптимального фазочастотного алгоритма прослеживания возникают определенные трудности, связанные, в частности, с оценкой функции $\gamma(f_i)$ в выбранном диапазоне частот. Поэтому в [7] был предложен фазочастотный алгоритм, полученный из оптимального алгоритма (1) путем замены в нем весовой функции $\gamma(f_i)$ на другие, специально подобранные функции. В общем случае, функция качества (критерий оценки местоположения сигналов) для таких алгоритмов фазочастотного прослеживания (ФЧП) может быть представлена в виде:

$$L(\tau) = \sum_{i=1}^n W(f_i) \cos[\phi(f_i) - 2\pi f_i \tau], \quad (2)$$

где $W(f)$ – некоторая весовая функция, $\phi(f)$ – текущий фазовый спектр, вычисленный в скользящем