

УДК 004.691.54

Исследование влияния параметров модели оптимизации распределенной корпоративной информационной системы на критерий эффективности

С.Ю.Землянская

Донецкий национальный технический университет
zsaa@ya.ru

Землянская С.Ю. Исследование влияния параметров модели оптимизации распределенной корпоративной информационной системы на критерий эффективности. Предложен подход к оптимизации распределенных корпоративных информационных систем, заключающийся в совместном использовании объектной модели и аппарата генетических алгоритмов. Проведено исследование влияния параметров генетического алгоритма на точность вычисления критерия эффективности

Введение

Основная цель создания распределённых корпоративных информационных систем (РКИС) – обеспечение возможности оперативного принятия управляющих решений, на основании данных о состоянии дел в корпорации (с учётом текущей финансовой и производственной ситуации), доступа пользователей к необходимым данным и предоставление возможности быстрого и безопасного обмена данными в рамках некоторой корпоративной организации [1,2]. Таким образом, одно из главных требований к РКИС – обеспечение наиболее быстрой реакции на запрос пользователя [3]. Следовательно, критерий эффективности функционирования РКИС можно представить как время реакции системы на запросы пользователей [4].

Рассмотрим, как определяется время реакции системы. Запросы, как правило, формируются пользователями на рабочих станциях и затем передаются по каналам связи узлам обработки, которые могут включать сервер баз данных и сервер приложений для обработки сложных запросов, после чего ответ по каналам связи передаётся пользователю. Тогда время выполнения запроса пользователя $Tresp$ определяется следующим образом:

$$Tresp = Tt + Tp + Tw + \max_{i=1, Nsq} \{Tsq_i\} \quad (1)$$

где, Tt – время передачи запроса, Tp – время обработки запроса, Tw – время задержки при передаче и обработке запроса, Tsq_i – время выполнения i -го подзапроса, Nsq – количество подзапросов данного запроса.

При оценке производительности РКИС используем пороговый критерий Tr_L , который отражает наихудшее время реакции системы на

запрос пользователя и гарантирует всем пользователям некоторый удовлетворительный уровень реакции системы на их запросы:

$$Tr_L = \max_{i=1, Nq} \{Tresp_i\} \quad (2)$$

где $Tresp_i$ – время реакции системы на i -ый запрос пользователя, Nq – количество пользовательских запросов, выполненных в системе за время исследования.

Распределённая корпоративная информационная система включает в себя компоненты физической и логической архитектуры [5]. Физическую архитектуру составляют множество узлов, на которых размещаются серверы баз данных, серверы приложений и рабочие станции, подключенные к сетевым устройствам; узлы РКИС соединяются между собой каналами связи. Логическую архитектуру определяют таблицы базы данных и приложения, используемые в РКИС, а также запросы, порождаемые рабочими станциями.

Одну из возможных конфигураций РКИС можно получить, если каждому из серверов поставить в соответствие определенный тип сервера, каждому сетевому устройству – определенный тип сетевого устройства и каждому каналу связи – определенный тип канала связи. Кроме того, необходимо задать вариант размещения фрагментов наборов данных на серверах баз данных и приложений на серверах приложений.

Задача оптимизации РКИС формулируется следующим образом [6]: пусть $\{DCIS_p, p=1, NP\}$ – множество всех возможных вариантов конфигураций РКИС, где NP – количество вариантов построения РКИС. Необходимо найти такой вариант построения РКИС $DCIS_x$ среди всех возможных вариантов,

для которого критерий эффективности (2) имеет минимальное значение:

$$Tr_{opt} = \min_{\{DCIS_p, p=1, NP\}} \{\max_{i=1, Nq} \{Tresp_i\}\} \quad (3)$$

где Nq – количество пользовательских запросов на чтение, обработанных в системе.

При этом необходимо учитывать ряд ограничений, накладываемых используемыми в системе данными и приложениями, а также финансовыми возможностями организации, в рамках которой функционирует или проектируется РКИС.

Сформулированная задача построения оптимальной конфигурации РКИС относится к NP-полным задачам перебора [7]. Данная задача не может быть решена классическими аналитическими методами в связи с большим количеством и разнородностью входных параметров, а также невозможностью точного определения значения целевой функции в связи со случайностью процессов, происходящих в системе. Поэтому для оптимизации параметров РКИС предлагается использовать новый подход [6], основанный на совместном использовании объектной модели РКИС [8] и аппарата генетических алгоритмов (ГА) [9], суть которого состоит в следующем.

Начальная популяция РКИС формируется на основе информации о её структуре. Параметры отдельных компонентов РКИС (типы серверов, рабочих станций, каналов связи, а также размещение наборов данных на серверах баз данных и приложений на серверах приложений) кодируются в виде хромосом ГА [6, 11]. В процессе оптимизации с помощью операторов ГА генерируются хромосомы, то есть различные варианты построения системы. Полученные схемы являются исходной информацией для объектной модели. Имитационное моделирование работы РКИС на основе объектной модели позволяет получить оценку критерия эффективности РКИС (3). Эта оценка, в свою очередь, является значением фитнес-функции ГА для данного варианта решения. Описанный процесс повторяется до тех пор, пока не будет достигнут критерий останова генетического алгоритма [6].

Оценка полученного с помощью генетического алгоритма субоптимального решения может быть проведена путем сравнения с решением, полученным с использованием метода полного перебора.

Для проведения экспериментальных исследований с использованием разработанной модели и алгоритма оптимизации с полученными в результате экспериментов параметрами в качестве объекта выбрана компьютерная информационная система ООО

«Донецкуглеобогатение».

Из-за территориальной распределенности структуры компании «Донецкуглеобогатение» ее компьютерная информационная система также имеет распределенную архитектуру, в состав которой входит 10 узлов: центральный узел (корпоративный сервер), по одному узлу на каждой фабрике и шахте. Для функционирования информационной системы используется единая база данных, включающая 9 таблиц. Таблицы базы данных горизонтально фрагментированы. С данными работают три приложения, инициирующие распределенные запросы.

Данная распределенная корпоративная информационная система имеет размер и характеристики, позволяющие использовать ее в качестве объекта экспериментальных исследований с помощью разработанных объектной модели РКИС и оптимизационного алгоритма

Для функционирования информационной системы используется единая база данных в рамках всего предприятия, включая филиалы. Для описанной выше компании с помощью разработанной объектной модели РКИС были проведены машинные эксперименты с целью выявления «узких мест» системы. Анализ загруженности каналов связи позволяет определить, какие из каналов являются наиболее загруженными и представляют из себя «узкие места» системы.

Для вычисления критерия эффективности каждый вариант конфигурации моделировался с помощью объектной модели в течение 10 дней модельного времени. Данный период времени был выбран, так как из результатов проведенных экспериментов видно, что значение среднего времени выполнения запросов стабилизируется после 9-10 дней моделирования и по прошествии такого периода времени значение критерия эффективности практически не подвержено влиянию отдельных редко возникающих запросов. Также увеличение времени моделирования приводит к значительному росту временных затрат на оптимизацию.

Была проведена оптимизация с помощью разработанного алгоритма оптимизации и с помощью процедуры полного перебора (с учетом ограничений).

В ходе процедуры полного перебора происходило моделирование каждого из возможных вариантов распределения данных и состава технических средств в течение десяти дней модельного времени.

В результате процедуры полного перебора были получены результаты, представленные на рис. 1. Полученные в ходе

моделирования данные о времени выполнения запросов позволяют сделать вывод, что в исследуемой РКИС время выполнения запросов в целом распределено равномерно, и нет запросов, время выполнения которых значительно больше или меньше времени выполнения других запросов. Таким образом, все запросы влияют на значение критерия эффективности в равной степени.

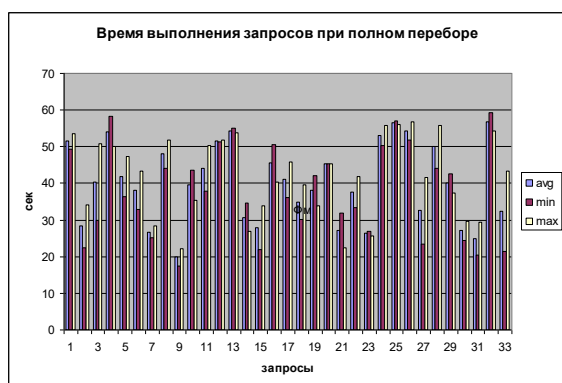


Рисунок 1 – Время выполнения запросов

Было получено оптимальное решение, при котором максимальное значение времени выполнения запроса на обновление, выбранного в качестве критерия эффективности, равно 56,83 сек. При этом прирост эффективности составляет 23%, а сокращение среднего времени выполнения запросов на 17 секунд, по сравнению с исходной системой с неоптимизированными параметрами.

Время поиска оптимального решения с использованием процедуры полного перебора с использованием ПЭВМ с процессором AMD Intel Core 2 Duo 3,2 GHz составило приблизительно 12 суток.

Используемый для оптимизации генетический алгоритм осуществляет случайный направленный поиск, а решения, получаемые с его помощью, являются субоптимальными. Отклонение решений от глобального оптимума, получаемых с помощью ГА зависит от следующих параметров [10]: размер популяции ($K_{поп}$); количество поколений ($K_{пок}$); вероятность мутации ($P_{мут}$); вероятность рекомбинации ($P_{рек}$); вероятность скрещивания ($P_{скр}$). Сущность оператора отбора состоит в выборе пар хромосом для скрещивания в зависимости от значения приспособленности каждой из особей. Оператор скрещивания реализует обмен группами генов между хромосомами. Как правило, выбирается одна либо несколько точек скрещивания в хромосомах и происходит обмен частями хромосом с разрывами в точках скрещивания. Оператор мутации представляет собой случайное изменение части хромосомы при переходе между поколениями

Таким образом, необходимо провести ряд вычислительных экспериментов, исследующих зависимость значений критерия эффективности работы РКИС, получаемых с помощью генетического алгоритма, от перечисленных параметров и выбрать такие параметры, при которых субоптимальное значение критерия эффективности, полученное при помощи генетического алгоритма, будет наиболее близким к глобальному оптимуму.

Было проведено исследование влияния размера популяции и количества поколений на значение критерия эффективности. Для этого были выбраны вероятности применения операторов мутации $P_{мут} = 0,1$, рекомбинации $P_{рек} = 0,8$ и скрещивания $P_{скр} = 0,9$.

В результате проведения данного эксперимента были получены зависимости критерия эффективности \bar{T}_Q от размера популяции и количества поколений. Данные зависимости представлены на рис. 2.

Из рисунка видно, что полученные зависимости не являются линейными. Чем больше количество особей в популяции, тем лучшее решение можно получить за определенное количество итераций. Однако чрезмерное увеличение количества поколений приводит к увеличению вычислительной сложности задачи.

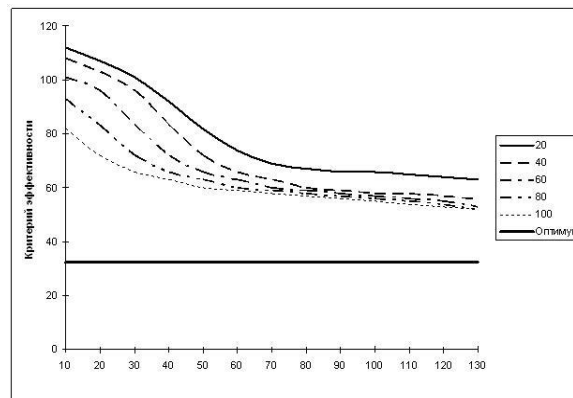


Рисунок 2 – Зависимость критерия эффективности РКИС от размера популяции ($K_{поп}$) и количества поколений ($K_{пок}$).

Из графиков видно, что увеличение количества поколений вначале приводит к значительному повышению качества поиска субоптимального решения, однако после достижения количества поколений 70, повышение эффективности существенно замедляется. Поэтому для проведения дальнейших экспериментов выбрано количество поколений $K_{пок} = 70$.

Увеличение количества особей в популяции также положительно сказывается на повышении эффективности работы

оптимизационного алгоритма, однако с увеличением числа особей в популяции растет и вычислительная сложность ГА. Из графиков (рис. 2) видно, что увеличение количества особей в популяции выше значения 40 приводит лишь к незначительному повышению эффективности работы алгоритма, поэтому для дальнейших экспериментов выбран размер популяции $K_{pop} = 40$.

При данных параметрах генетического алгоритма произведен поиск рациональных вероятностей генетических операторов: скрещивания $P_{скр}$, мутации $P_{мут}$ и рекомбинации $P_{рек}$. Зависимость критерия эффективности \bar{T}_Q от вероятностей применения операторов скрещивания $P_{скр}$ и рекомбинации $P_{рек}$ представлена на рис. 3.

Из зависимостей (рис. 3) видно, что рациональные значения генетических операторов, при которых наблюдается наибольшее приближение критерия эффективности \bar{T}_Q к глобальному оптимуму следующие: $P_{рек}=0,35$, $P_{скр}=0,7$.

Было проведено исследование зависимости критерия эффективности работы РКИС \bar{T}_Q от вероятности выполнения оператора мутации $P_{мут}$ и вероятности выполнения оператора рекомбинации $P_{рек}$, в результате чего были получены зависимости, представленные на рис. 4.

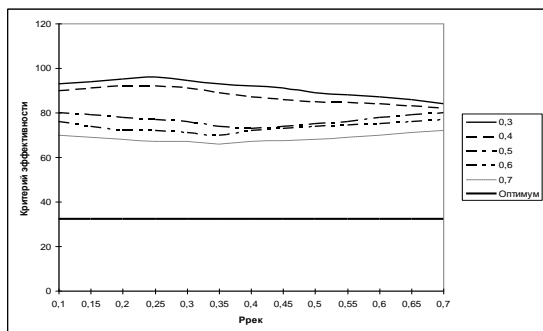


Рисунок 3 – Зависимость критерия эффективности от вероятности скрещивания $P_{скр}$ и рекомбинации $P_{рек}$

Из зависимостей (рис. 4) видно, что рациональные значения генетических операторов, при которых наблюдается наибольшее приближение критерия эффективности к глобальному оптимуму следующие: $P_{рек}=0,35$, $P_{мут}=0,05$.

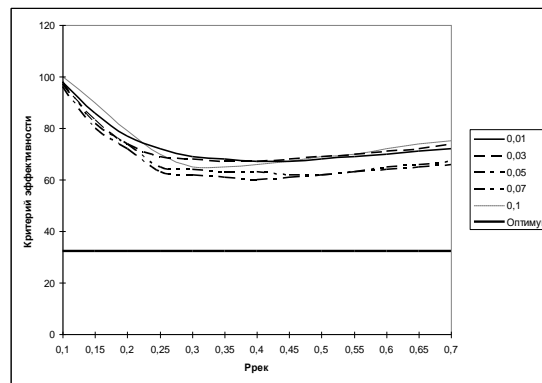


Рисунок 4 – Зависимость критерия эффективности РКИС от вероятности рекомбинации $P_{рек}$ и вероятности мутации $P_{мут}$

Было проведено исследование зависимости критерия эффективности работы \bar{T}_Q от вероятности выполнения оператора мутации $P_{мут}$ и вероятности выполнения оператора скрещивания $P_{скр}$, в результате чего были получены зависимости, представленные на рис. 5.

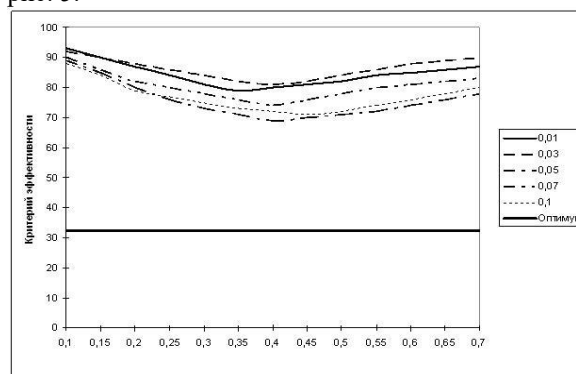


Рисунок 5 – Зависимость критерия эффективности \bar{T}_Q от вероятности скрещивания $P_{скр}$ и вероятности мутации $P_{мут}$

Из зависимостей (рис. 5) видно, что рациональные значения генетических операторов, при которых наблюдается наибольшее приближение критерия эффективности к глобальному оптимуму, следующие: $P_{скр}=0,45$, $P_{мут}=0,05$.

Таким образом, на основании проведенных экспериментов были определены следующие параметры генетического алгоритма: $P_{мут} = 0,05$, $P_{скр} = 0,5$, $P_{рек} = 0,35$, размер популяции 40 и количество поколений 70.

Разработанный программный комплекс позволяет провести частичную оптимизацию, то есть можно произвести оптимизацию распределения таблиц БД или приложений, или только параметров каналов связи, или только параметров узлов (серверов и сетевых устройств) [12]. В зависимости от этого изменяется значение достигнутого критерия эффективности.

На рис. 6 приведены значения критерия эффективности в зависимости от полноты проведенной оптимизации

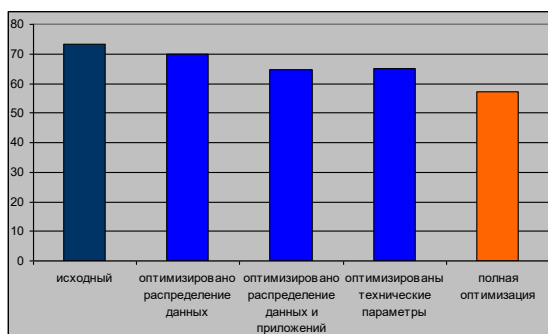


Рисунок 6 – Значения критерия эффективности в зависимости от полноты проведенной оптимизации

Наилучшее субоптимальное решение, полученное с помощью генетического алгоритма, составляет 57,33 секунд, что на 0,5 секунды больше глобального оптимума полученного с помощью алгоритма полного перебора. Относительное отклонение составило 0,88%. При этом нахождение глобального оптимума с помощью алгоритма полного перебора заняло 12 суток, в то время как поиск субоптимального решения с помощью генетического алгоритма длится 10-15 минут.

В результате проведенных вычислительных экспериментов с разработанным алгоритмом оптимизации РКИС определены рациональные значения параметров генетического алгоритма: вероятность мутации = 0,05, вероятность скрещивания = 0,7, вероятность рекомбинации = 0,23, количество популяций = 50 и количество поколений = 30. Проведенные вычислительные эксперименты показали, что применение предложенного подхода к оптимизации параметров распределенной корпоративной информационной системы позволяет повысить эффективность ее работы. Полученное при данных параметрах оптимизационной модели время выполнения самого длительного запроса на обновление, выбранного в качестве критерия эффективности работы РКИС, составило 57,33 сек, что отличается от оптимального решения, полученного методом полного перебора (56,83 сек) всего на 0,88% , то есть субоптимальное решение достаточно близко к глобальному оптимуму. Оптимизация распределения таблиц базы данных и приложений, а также параметров серверов, каналов связи и сетевых устройств обеспечивает повышение эффективности работы РКИС за счет уменьшения максимального времени выполнения пользовательских запросов

и модернизации оборудования при минимальных затратах со стороны корпорации

Литература

1. Lynda M. Applegate, James L. McKenney, F. Warren McFarlan. Corporate information systems management: The challenges of management in an information age. [Текст] / Lynda M. Applegate, James L. McKenney, F. Warren McFarlan // McGraw-Hill Higher education, 1999. – 312 p.
2. Киселев А.Г. Корпоративная и комплексная система управления промышленного предприятия (КИС). / А.Г.Киселев. – Новосибирск, 2010. – 408 с.
3. Lee Y. W. AIMQ: A Methodology for Information Quality Assessment [Электронный ресурс] / Y. W. Lee, D. M. Strong, B. K. Kahn, R. Y. Wang // Information & Management, 2002. № 40, pp 133-146. – Режим доступа: <http://www.sciencedirect.com/science/article/pii/S0378720602000435>
4. Tervonen I. Towards deeper co-understanding of software quality [Электронный ресурс] / I. Tervonen, P. Kerola // Information and Software Technology, 1999. – vol. 39, No 14-15. – Режим доступа: <http://www.sciencedirect.com/science/article/pii/S0950584997000608>
5. Агибалов Г.П., Скутин А.А. Математическая модель и технология разработки безопасных корпоративных информационных систем. [Электронный ресурс] / Г.П. Агибалов, А.А.Скутин // Электронный журнал «Исследовано в России»: Томский государственный университет, 2001. с. 1739-1750. Режим доступа: <http://zhurnal.ape.relarn.ru/articles/2001.151.pdf>
6. Землянская С.Ю. Оптимизация распределённых корпоративных информационных систем с использованием генетических алгоритмов и объектного моделирования [Текст] /С.Ю. Землянская, С.В. Лаздынь // Наукові праці ДонНТУ. Серія “Обчислювальна техніка та автоматизація” – Випуск 16(147). – Донецьк: ДонНТУ, 2009. С. 83-95.
7. Овчинников В.А. Алгоритмизация комбинаторно-оптимизационных задач при проектировании ЭВМ и систем [Текст]: Монография/ В.А. Овчинников – Москва: МГТУ им. Н.Э. Баумана, 2001. – 288 с.
8. Землянская С.Ю. Объектная модель распределённой компьютерной информационной системы. [Текст] /С.Ю. Землянская, С.Ю. Лаздынь // Наукові праці ДонНТУ. Серія “Обчислювальна техніка та автоматизація” –

Выпуск 14(129). - Донецк: ДонНТУ, 2008. – С. 107-116.

9. Скобцов Ю.А. Основы эволюционных вычислений [Текст]: Монография/ Ю.А. Скобцов – Донецк: ДонНТУ, 2008. – 326с.

10. Kalyanmou Deb. Understanding interactions among genetic algorithm parameters [Электронный ресурс] / Deb. Kalyanmou, Agrawal Samir // Foundations of Genetic Algorithms, 1999. pp. 265-286. – Режим доступа: <http://www.iitk.ac.in/kangal/papers/k99003.pdf>

11. Божич В.И. Методы генетического поиска для решений представимых мультихромосомами / В.И. Божич, В.Б. Лебедев // Перспективные информационные технологии и

интеллектуальные системы. – 2002. – № 3. – С. 38 – 44.

12. Моделирование и оптимизация распределенных информационных систем. / [Скобцов Ю.А., Лаздынь С.В., Землянская С.Ю. и др.] Донецк: Ноулидж. Донецкое отделение, 2012. – 300 с.

Землянська С. Ю. Дослідження впливу параметрів моделі оптимізації розподіленої корпоративної інформаційної системи на критерій ефективності. Запропоновано підхід до оптимізації розподілених корпоративних інформаційних систем, який полягає у спільному використанні об'єктної моделі та апарату генетичних алгоритмів. Проведено дослідження впливу параметрів генетичного алгоритму на точність обчислення критерію ефективності
Ключові слова: розподілена система, об'єктна модель, оптимізація, параметри, генетичний алгоритм, критерій ефективності

Zemlyansky S. Y. Study of the effect of Distributed corporate information system optimizing model parameters influence on the criterion of efficiency. The approach to optimization of distributed corporate information systems, which consists in the joint use of the object model and the genetic algorithms is proposed. The effect of genetic algorithm parameters on the accuracy of calculation of criteria of efficiency is investigated

Keywords: distributed system, object model, optimization, parameters, genetic algorithm, criterion of efficiency

Статья поступила в редакцию 20.09.2015
Рекомендована к публикации д-ром техн. наук В.Н. Павлышом