

Интернет-журнал «Наукovedение» ISSN 2223-5167 <http://naukovedenie.ru/>

Том 7, №6 (2015) <http://naukovedenie.ru/index.php?p=vol7-6>

URL статьи: <http://naukovedenie.ru/PDF/80TVN615.pdf>

DOI: 10.15862/80TVN615 (<http://dx.doi.org/10.15862/80TVN615>)

УДК 004.02:004.6:519.7

Карандеев Денис Юрьевич

ФГБОУ ВПО «Хакасский государственный университет им. Н.Ф. Катанова»

Россия, Абакан¹

Аспирант кафедры «Информационных технологий и систем»

E-mail: den_dr_house_1991@mail.ru

РИНЦ: http://elibrary.ru/author_profile.asp?id=806741

Голубничий Артем Александрович

ФГБОУ ВПО «Хакасский государственный университет им. Н.Ф. Катанова»

Россия, Абакан

Ассистент кафедры «Инженерной экологии и основ производства»

E-mail: artem@golubnichij.ru

РИНЦ: http://elibrary.ru/author_items.asp?authorid=683836

ORCID: <http://orcid.org/0000-0002-7559-7492>

Реализация метода ветвей и границ в статистической среде R

¹ 655017, Республика Хакасия, г. Абакан, ул. Ленина, д. 90

Аннотация. В статье анализируются пути и подходы к решению задачи энергоэффективности отраслей промышленности. В качестве основных направлений авторами выделяется: 1. Прогнозирование электропотребления с целью уменьшения затрат и 2. Оптимизация базовых объектов потребления электроэнергии. Оба направления, по мнению авторов, целесообразно решать методами оптимизации. Проведенный анализ существующих методов позволяет сделать вывод об оптимальности использования в сферах технических систем метода «ветвей и границ». Проверка метода осуществлялась на основе решения NP-полной задачи. В качестве примера была выбрана одна из классических задач оптимизации «задача коммивояжера». В качестве среды для реализации метода «ветвей и границ» выбрана статистическая среда R. Причинами для выбора послужили: 1. Возможность подключения сторонних пакетов, 2. Возможность работы с большим массивом данных и 3. Открытость среды с точки зрения возможности внесения изменений. В статье рассматривается алгоритм решения оптимизационной задачи, а также представляется часть реализуемого кода метода «ветвей и границ» для решения NP-полной задачи.

Ключевые слова: метод ветвей и границ; статистическая среда R; оптимизационные методы; задача коммивояжера; условия неопределенности; оптимальные структуры технических систем.

Ссылка для цитирования этой статьи:

Карандеев Д.Ю., Голубничий А.А. Реализация метода ветвей и границ в статистической среде R // Интернет-журнал «НАУКОВЕДЕНИЕ» Том 7, №6 (2015) <http://naukovedenie.ru/PDF/80TVN615.pdf> (доступ свободный). Загл. с экрана. Яз. рус., англ. DOI: 10.15862/80TVN615

Статья опубликована 25.11.2015.

Введение

На сегодняшний день одним из высших приоритетов государственной энергетической политики является повышение энергоэффективности промышленности. Существует достаточно большое количество способов реализации данного направления:

1. Повышение точности краткосрочного прогнозирования электропотребления путем выявления и анализа факторов, влияющих на электропотребление [1-2], а впоследствии разработка с учетом этих факторов метода, позволяющего повысить точность прогнозирования и тем самым снизить стоимость закупок электроэнергии на оптовом рынке [3].

2. Оптимизация базовых объектов потребления электроэнергии, таких, например, как асинхронные двигатели [4] или же оптимизацию структур сложных технических систем, таких как электрические или транспортные сети, с учетом их экономичности и надежности [5]. При этом стоит отметить, что данные технические системы находятся в условиях неопределенности информации, вызванным вероятностной величиной надежности объектов, из которых состоят данные системы, это также ведет к усложнению решения задачи оптимизации данных систем.

Из всего выше сказанного следует, что на сегодняшний день одной из самых перспективных областей изучения является анализ существующих методов оптимизации процессов и технологий в различных областях экономики и техники. Оптимизация – это одна из важнейших задач, решаемых в большинстве сфер деятельности, будь то программирование, определение надежности технических систем, выполнение проектов по созданию программного обеспечения и т.д. Известно достаточно много оптимизационных методов, имеющих как свои преимущества, так и недостатки. При этом стоит отметить, что большинство задач глобальной оптимизации входят в класс NP-полных задач, то есть таких задач, точные алгоритмы решения которых не могут иметь временную сложность меньше экспоненциальной. Соответственно, важной задачей является реализация методов оптимизации, способных решать задачи такого уровня, на наиболее подходящем для этой цели языке программирования. Стоит пояснить, что одной из глобальных задач оптимизаций можно выделить задачу оптимизации структур сложных технических систем. В силу того, что разработка наиболее оптимального оптимизационного метода, позволяющего строить наиболее оптимальные технические сети, это может впоследствии позволить сэкономить значительные ресурсы при проектировании сетей различного вида, а также повысить надежность данных сетей. В данной статье авторами анализируется возможность реализации одного из лучших методов оптимизации метода «ветвей и границ» в статистической среде R.

Методы оптимизации

Методов оптимизации насчитывают достаточно большое количество, если говорить лишь о тех методах оптимизации, которые можно впоследствии использовать при создании метода, позволяющего рассчитывать оптимальные структуры сложных технических сетей, то можно выделить так называемые алгоритмы поиска на графах, так как сеть чаще всего представляют в виде графа. В свою очередь данные алгоритмы делятся на неинформированные методы (поиск в ширину, поиск в глубину, поиск с возвратом, двунаправленный поиск, лучевой поиск и т.д.), информированные методы (метод ветвей и границ, поиск по первому наилучшему совпадению, альфа-бета-отсечение, RBFS и т.д.), кратчайшие пути (волновой алгоритм, алгоритм Беллмана — Форда, алгоритм Джонсона, алгоритм Дейкстры и т.д.), минимальное остовное дерево (алгоритм Прима, алгоритм Крускала и алгоритм Борувки) и прочие. В силу того, что информационные методы поиска, известные также как эвристический поиск (англ. informed search, heuristic search) [6], по той

причине, что он использует информацию, которая относится к решению той задачи, которая в данный момент исследуется, например, учитывает критерий экономичности при построении электрической сети, данный вид поиска является более целесообразным, нежели остальные.

Метод «ветвей и границ»

Среди всех информационных методов можно выделить такой метод как метод «ветвей и границ» (англ. branch and bound). Это из самых признанных и известных оптимизационных методов, способных решать задачи класса NP [7]. Метод ветвей и границ впервые предложили в 1960 году Ленд и Дойг [8] для решения задач целочисленного программирования. Это общий алгоритмических метод для нахождения оптимальных решений различных задач оптимизации, особенно дискретной и комбинаторной оптимизации, поиска оптимального решения и использующий результаты решения оценочных задач. При этом данный метод представляет собой алгоритм решения задачи, имеющей структуру в виде дерева ветвления, описывающего разбиение всего множества решений на непересекающиеся подмножества. Суть данного метода неявного перебора заключается в полном переборе и отсеивании подмножеств допустимых решений, заведомо не содержащих оптимальных решений или другими словами.

Алгоритм решения задачи коммивояжера методом ветвей и границ

Одним из лучших способов проверки качества свойств метода оптимизации является проверка его на решении NP-полной задачи. Одной из самых известных NP-полных задач комбинаторной оптимизации считается так называемая задача коммивояжера (Travelling salesman problem, TSP), суть которой заключается в поиске самого выгодного маршрута, проходящего через указанные города хотя бы по одному разу с последующим возвратом в исходный город. Алгоритм решения данной задачи посредством метода ветвей и границ был предложен в 1963 году группой авторов Дж. Литтлом, К. Мурти, Д. Суини и К. Кэролом и впоследствии получил название алгоритм Литтла [9]. Сложность данной задачи заключается в том, что при увеличении количества городов, при полном переборе всех вариантов количество возможных маршрутов будет очень большим, к примеру, для 10 городов это число уже превышает миллион, а при количестве городов количество маршрутов превышает 70 миллиардов, следовательно, решение задачи коммивояжера методом полного перебора оказывается практически неосуществимым. Данную задачу также можно решить, используя алгоритм Крускала и «деревянного» алгоритма, эти методы ускоряют разработку алгоритма по сравнению с методом полного перебора, однако не всегда дают оптимальное решение. В виду этого, применение метода ветвей и границ для решения данной задачи весьма актуально и действенно. Суть данного алгоритма заключается в поиске самого короткого оптимального гамильтонова контура в графе, путем перебора вариантов маршрута (т.е. циклов обхода графа) с отсечением, при этом отсекаются такие построенные маршруты, оценка снизу длины маршрутов которых больше или равна длине построенного ранее полного наилучшего маршрута.

Поэтапно алгоритм решения задачи коммивояжера, используя метод «ветвей и границ», выглядит следующим образом:

1. Всё количество возможных путей разбивается на 2 подмножества, содержащее непосредственный переход из вершины в вершину и не содержащее его. Для каждого из классов находится нижняя оценка.
2. Далее проводится разбиение одного из уже существующих подмножеств на два подмножества по выше описанному методу. Затем находится их нижняя оценка.

Если одно из подмножеств является уже построенным путем, то сравниваются длины этого пути с наилучшим найденным к текущему моменту значением и происходит изменение данного наилучшего пути если необходимо.

3. Проводится отсев.
4. Повторяются второй и третий этапы вплоть до нахождения самого оптимального решения задачи коммивояжера.

При этом в ходе решения ведется постоянный подсчет текущего значения нижней границы. Итоговое значение нижней границы должно совпасть с длиной результирующего пути.

Также стоит отметить, что задача коммивояжера в силу своей специфичности пребывает в условиях неопределенности информации, что приводит к ухудшению решения ее путем использования простых методов оптимизации.

Статистическая среда R

На сегодняшний день большинство качественных программных средств реализации методов оптимизации наподобие метода ветвей и границ, являются коммерческими продуктами, статистическая же среда R является свободно распространяемым продуктом, обладающим огромным количеством достоинств. Основные достоинства данной среды: бесплатность, многофункциональность, благодаря наличию большого перечня подключаемых пакетов, количество которых уже превысило 7000 и это цифра неуклонно растет, благодаря постоянно увеличивающейся популярности данного языка программирования, на данный момент данные пакеты пишутся и публикуются ведущими университетами мира, такими как Гарвард, Оксфорд, Кембридж и так далее. Помимо этого, данный язык обладает достаточно простым синтаксисом, что значительно упрощает чтения его кода. Данная среда была разработана как бесплатная альтернатива коммерческого языка S-PLUS, разработанного в 1988 году, который в свою очередь был третьей версией языка S, разработанного в 1976 году в компании AT&T Labs [10]. На сегодняшний день данный язык вырос в высокофункциональный инструмент для разработок огромного количества ученых. В нашей стране он пока не обрел должной популярности в связи с малым количеством документаций на русском языке, однако в зарубежных университетах издаются многостраничные учебники и монографии относительно как работы с самим пакетом R, так и его применения при исследовании и обработке данных в статистике, медицине, финансовом анализе, экологии, актуарной математике и так далее [11].

Реализация метода ветвей и границ в статистической среде R

Авторами было принято решение реализовать метод ветвей и границ в данной статистической среде, в частности для работы использовался графический интерфейс RStudio. Данная среда была выбрана по нескольким причинам:

Во-первых, среда R благодаря наличию подключаемых пакетов обладает огромной функциональностью, при этом можно одновременно установить неограниченное количество пакетов и включать их по мере надобности, благодаря этому производительность данной среды находится на достаточно высоком уровне.

Во-вторых, R – это в первую язык позволяющий работать с огромными массивами данных, так как в первую очередь он нацелен на работу со статистикой [12], а при анализе методов оптимизации зачастую требуется работать с таблицами, содержащими огромное количество данных.

В-третьих, благодаря тому, что это скриптовый язык, внесение необходимых изменений и дополнений производится без перекомпиляции программы.

В-четвертых, использование данной статистической среды в ходе анализа большого массива данных по электропотреблению города Абакана [1-3] показало существенные преимущества данного языка программирования и простоту работы с данной средой.

Часть кода реализации метода ветвей и границ на языке программирования R представлена на рисунке 1, как видно, благодаря достаточно легкому синтаксису данного языка программирования, код легко читабелен:

```
if(is.null(tree$edge.length)){
  ntree<-list(tree(n=6,tip.label=sample(tree$tip.label,6)))
} else {
  pr<-names(sort(colSums(eltric(tree))))[1:6]
  ntree<-list(stree(n=6,tip.label=pr))
}
class(ntree)<-"Commiv"
dop<-ntree[[1]]$tip.label; remaining<-setdiff(tree$tip.label,dop)
while(length(remaining)>0){
  otree<-ntree; ntree<-list()
  ntrees<-sample(remaining,1)
  res<-vector()
  for(i in 1:length(otree)){
    pgk<-add.everywhere(otree[[i]],ntrees)
    score<-parsimony(pgk,data)
    ntree<-c(ntree,pgk[score<=bound])
    res<-c(res,score[score<=bound])
  }
}
```

Рисунок 1. Часть исходного кода метода «ветвей и границ» в статистической среде R

Заключение

Реализованный в статистической среде R метод оптимизации под названием метод «ветвей и границ» показал хорошие результаты по функциональности и времени расчета задачи коммивояжера, что говорит о целесообразности использования данного метода в качестве инструмента при решении NP-полных задач, к примеру, задач оптимизации структур сложных технических систем (электрических, транспортных сетей и так далее), так как при решении задач такого рода очень важна высокая производительность и функциональность той среды, в которой метод реализован.

ЛИТЕРАТУРА

1. Карандеев Д.Ю. Влияние температуры окружающей среды на электропотребление большого города // Научно-практический журнал «Современные научные исследования и инновации»: 05.00.00 Технические науки: электрон. научн. журн. 2015. №2 (46).
2. Карандеев Д.Ю. Влияние типов дней на электропотребление города // Научно-практический журнал «Современная техника и технологии»: электр. научн. журн. 2015 №2 (42).
3. Карандеев Д.Ю., Валь П.В. Повышение качества краткосрочного прогнозирования электропотребления с помощью метода на основе поиска подобных суток // Энергетика Глазами молодежи: научные труды V международной научно-технической конференции, Т.2. / Министерство образования и науки РФ, Томский политехнический университет. – Томск, 2014. С. 223-227.
4. Карандеев Д.Ю., Энгель Е.А. Прямое управление моментом асинхронного двигателя с использованием адаптивного нейроконтроллера в условиях неопределенности // Интернет-журнал «НАУКОВЕДЕНИЕ» Том 7, №5 (2015).
5. В.Н. Костин: Оптимизационные задачи электроэнергетики: Учеб. пособие. – СПб.: СЗТУ, 2003 – 120 с.
6. Стюарт Рассел, Питер Норвиг. Составление допустимых эвристических функций // Искусственный интеллект: современный подход = ArtificialIntelligence: A ModernApproach. - 2-е изд. - М.: Вильямс, 2006. - С. 170 - 173. - 1408 с.
7. Томас Х. Кормен, NP-полнота // Алгоритмы: построение и анализ = INTRODUCTION TO ALGORITHMS. - 2-е изд. — М.: «Вильямс», 2006. - С. 1296.
8. A.H. Land and A.G. Doig. An automatic method of solving discrete programming problems. *Econometrica*. v28 (1960), pp. 497-520.
9. Little J.D.C., Murty K.G., Sweeney D.W., and Karel C. An algorithm for the Traveling Salesman Problem // *Operations Research*. 1963. No. 11. P. 972–989.
10. Зарядов, И.С. Введение в статистический пакет R: учебно-методическое пособие / И.С. Зарядов – Москва: Издательство Российского университета дружбы народов, 2010. – 207 с.
11. Шипунов, А.Б. Наглядная статистика. Используем R! / А.Б. Шипунов, Е.М. Балдин, П.А. Волкова и др. – М.: ДМК Пресс, 2012. – 298 с.
12. R in action. Data analysis and graphics with R, Robert I. Kabakoff, 2011.

Рецензент: Статья рецензирована членами редколлегии журнала.

Karandeev Denis Jurevich

Katanov Khakass State University
Russia, Abakan

E-mail: den_dr_house_1991@mail.ru

Golubnichiy Artem Aleksandrovich

Katanov Khakass State University
Russia, Abakan

E-mail: artem@golubnichij.ru

Implementation of the method of branch and bound in the statistical environment R

Abstract. The article analyzes the ways and approaches to the problem solution of industries energy efficiency. As the guidelines the authors mark out: 1. Forecasting of power consumption in order to reduce costs and 2. Optimization of the basic objects of electricity consumption. In the authors' opinion both guidelines should be solved by optimization techniques. The drawn analysis of existing methods allows one to conclude that the branch-and-bound method finds an optimal use in the fields of technical systems. Checking this method was carried out on the basis of solution of nondeterministic polynomial time complete problem (NP-complete problem). As an example, it was chosen one of the classic problems of optimization - "a traveling salesman problem". The statistical environment R was selected as a medium for implementing the branch-and-bound method. The reasons for this choice were the following: 1. Connectivity of third-party packages, 2. Capability of working with a large data array and 3. Openness of environment from the point of view of possible introduction of changes. The article considers an algorithm for solving optimization problem and represents a part of the instrumented code for the branch-and-bound method to solve NP-complete problem.

Keywords: branch-and-bound method, statistical environment R, optimization methods, traveling salesman problem, conditions of indeterminacy, optimal structures of technical systems

REFERENCES

1. Karandeev D.Yu. Vliyanie temperatury okruzhayushchey sredy na elektropotreblenie bol'shogo goroda // Nauchno-prakticheskiy zhurnal «Sovremennyye nauchnye issledovaniya i innovatsii»: 05.00.00 Tekhnicheskie nauki: elektron. nauchn. zhurn. 2015. №2 (46).
2. Karandeev D.Yu. Vliyanie tipov dney na elektropotreblenie goroda // Nauchno-prakticheskiy zhurnal «Sovremennaya tekhnika i tekhnologii»: elektr. nauchn. zhurn. 2015 №2 (42).
3. Karandeev D.Yu., Val' P.V. Povyshenie kachestva kratkosrochnogo prognozirovaniya elektropotrebleniya s pomoshch'yu metoda na osnove poiska podobnykh sutok // Energetika Glazami molodezhi: nauchnye trudy V mezhdunarodnoy nauchno-tekhnicheskoy konferentsii, T.2. / Ministerstvo obrazovaniya i nauki RF, Tomskiy politekhnicheskoy universitet. – Tomsk, 2014. С. 223-227.
4. Karandeev D.Yu., Engel' E.A. Pryamoe upravlenie momentom asinkhronnogo dvigatelya s ispol'zovaniem adaptivnogo neyrokontrollera v usloviyakh neopredelennosti // Internet-zhurnal «NAUKOVEDENIE» Tom 7, №5 (2015).
5. V.N. Kostin: Optimizatsionnye zadachi elektroenergetiki: Ucheb. posobie. – SPb.: SZTU, 2003 – 120 s.
6. Styuart Rassel, Piter Norvig. Sostavlenie dopustimyykh evristicheskikh funktsiy // Iskusstvennyy intellekt: sovremennyy podkhod = ArtificialIntelligence: A ModernApproach. - 2-e izd. - M.: Vil'yams, 2006. - S. 170 - 173. - 1408 s.
7. Tomas Kh. Kormen, NP-polnota // Algoritmy: postroenie i analiz = INTRODUCTION TO ALGORITHMS. - 2-e izd. — M.: «Vil'yams», 2006. - S. 1296.
8. A.H. Land and A.G. Doig. An automatic method of solving discrete programming problems. *Econometrica*. v28 (1960), pp. 497-520.
9. Little J.D.C., Murty K.G., Sweeney D.W., and Karel C. An algorithm for the Traveling Salesman Problem // *Operations Research*. 1963. No. 11. P. 972–989.
10. Zaryadov, I.S. Vvedenie v statisticheskiy paket R: uchebno-metodicheskoe posobie / I.S. Zaryadov – Moskva: Izdatel'stvo Rossiyskogo universiteta druzhby narodov, 2010. – 207 s.
11. Shipunov, A.B. Naglyadnaya statistika. Ispol'zuem R! / A.B. Shipunov, E.M. Baldin, P.A. Volkova i dr. – M.: DMK Press, 2012. – 298 s.
12. R in action. Data analysis and graphics with R, Robert I. Kabakoff, 2011.