УДК 621.372.2

Р.И. Аширбакиев, А.О. Мелкозеров, Ег.В. Лежнин

Алгоритм аппроксимации набора данных в применении к параметрам трасс печатных плат

Предложен алгоритм аппроксимации большого набора данных. Выполнено сравнение предложенного алгоритма и полиномиального приближения для случая аппроксимации погонной задержки линий передачи в широком диапазоне параметров основных структур печатных плат. Показана работоспособность предложенного алгоритма и описаны возможности его совершенствования.

Ключевые слова: аппроксимация данных, печатная плата, аналитические модели, полиномиальное приближение.

Для решения задачи аппроксимации больших наборов данных аналитическими моделями применяют различные методы. Широко используются: метод группового учета аргументов [1], алгоритм Лассо [2], метод наименьших углов LARS [3], генетическое программирование [4], регрессионный анализ [5]. Они обладают различными достоинствами и недостатками, поэтому целесообразно исследовать их для решения конкретных задач, а также рассмотреть возможность создания новых методов или алгоритмов. Для случая аппроксимации погонной задержки линий передачи в широком диапазоне параметров основных структур печатных плат все методы, перечисленные выше, показали неудовлетворительные результаты: относительная погрешность полученных моделей по отношению к расчетным данным составила 50% и более [6].

Цель работы – разработать новый алгоритм для аппроксимации большого набора данных.

Предложенный алгоритм. Набор данных состоит из числовых значений, которые были получены в результате длительных вычислений. Рассмотрим предложенный алгоритм аппроксимации на примере четырех независимых параметров. Пусть дан набор данных из n строк:

$$x_{i1}\ x_{i2}\ x_{i3}\ x_{i4}\ y_i,$$
$$\ldots$$
$$x_{n1}\ x_{n2}\ x_{n3}\ x_{n4}\ y_n,$$

где n – количество строк, которое можно определить как $n=\prod_{k=1}^{4} Cnt_k$, где Cnt_k – количество значений k-го независимого x параметра. В данной задаче значения независимых x параметров заданы в определенном диапазоне небольшим количеством различных значений с заданным шагом, так что $x_{ik} = t_k + s_k dt_k$, где t_k – начальное значение, dt_k – шаг, $s_k = 0,1, ..., Cnt_k$. Набор данных представляет все комбинации значений независимых переменных, где каждой комбинации соответствует одно значение y, которое зависит от каждого значения этой комбинации.

Необходимо синтезировать функцию $f(x_1, x_2, x_3, x_4) = y$, с помощью которой можно будет вычислить приближенное значение y для заданных x_1, x_2, x_3, x_4.

Шаг 1. Пусть $1 \le i \le n$, где $n=\prod_{k=1}^{4} Cnt_k$, и $1 \le j \le m$, где $m=\prod_{k=1}^{3} Cnt_k = \frac{n}{Cnt_4}$ – количество различных комбинаций без последнего параметра x_{i4}. Сформируем m блоков, в которых $x_{i1} = x_{j1}$, $x_{i2} = x_{j2}$, $x_{i3} = x_{j3}$,x_{i4} ,y_i, т.е. значения повторяются в первых трех столбцах, но не в последнем.

Шаг 2. Необходимо выполнить аппроксимацию каждого из m блоков по четвертому параметру. Применим простейшую линейную регрессию. Для $x = x_{i4}$ в каждом блоке необходимо найти аппроксимирующую функцию $f(x) = y = a_1x + b_1$. Таким образом, получаем два коэффициента a_1 и b_1. После этого необходимо построить новый набор данных $x_{j1}\ x_{j2}\ x_{j3}\ a_{j1}\ b_{j1}$, где $1 \le j \le m$, а a_{j1} и b_{j1} – коэффициенты для j-го блока. (Отметим, что с помощью коэффициентов a_{j1} и b_{j1} можно обратно восстановить $y_i = a_{j1}x_{i4}+b_{j1}$.) После выполнения аппроксимации по x_{i4} формируется новый набор данных $x_{j1}\ x_{j2}\ x_{j3}\ a_{j1}\ b_{j1}$, в котором отсутствуют y_i и x_{i4}, и который в Cnt_4 раз меньше.

Шаг 3. Далее необходимо заменить таким же образом данные по x_{j3}, т.е. выделить все блоки данных, в которых $x_{i1} = x_{j1}$, $x_{i2} = x_{j2}$, $x_{i3}\ a_{i1}$, b_{i1}. Однако теперь нужно получить четыре коэффициента для $f_1(x_{i3}) = a_{i1}$ и $f_2(x_{i3}) = b_{i1}$. В итоге объем данных сократится еще в Cnt_3 раз, а коэффициентов станет в два раза больше.

Шаги 4, 5. Далее остальные переменные x_{i1} и x_{i2} заменяются по схеме, аналогичной описанной выше. После замены x_{i2} останутся коэффициенты и одна переменная $x_{i1}a_{4i}b_{4i}a_{5i}b_{5i}a_{6i}b_{6i}a_{7i}b_{7i}$. После замены x_{i1} остаются только 16 коэффициентов, по которым можно восстановить все данные: $a_8b_8a_9b_9a_{10}b_{10}a_{11}b_{11}a_{12}b_{12}a_{13}b_{13}a_{14}b_{14}a_{15}b_{15}$.

Конец.

Коэффициенты используются для формирования аппроксимирующей функции

$$f(x_1,x_2,x_3,x_4) = (((a_8x_1+b_8)x_2+a_9x_1+b_9))x_3 + ((a_{10}x_1+b_{10})x_2+a_{11}x_1+b_{11})))x_4 +$$
$$+ (((a_{12}x_1+b_{12})x_2+(a_{13}x_1+b_{13}))x_3 + ((a_{14}x_1+b_{14})\, x_2 + (a_{15}x_1+b_{15}))).$$

Полиномиальное приближение. Рассмотрим аппроксимацию с помощью полинома второй степени для погонной задержки линий передачи в широком диапазоне параметров основных структур печатных плат. Аппроксимирующая функция имеет вид $f(x_1,x_2,...,x_n)=a_0+\sum_{i=1}^{n}a_ix_i+$ $+\sum_{i=1}^{n}b_ix_i^2+\sum_{i=1}^{n}\sum_{j=i+1}^{n}c_{ij}x_ix_j$, где n – количество независимых параметров; x_i – значения параметров структур (например, w, t и т.д.), a_i, b_i и c_{ij} – коэффициенты.

Сравнение аппроксимации с помощью полиномиального приближения и предложенного алгоритма. Полный набор параметров структур ПП принят следующим.

ε_r диэлектриков:

- препрег и подложка: 3,8; 3,9; 4,0; 4,1; 4,2; 4,3; 4,4; 4,5 (8 значений);
- паяльная маска: 3,0; 3,1; 3,2; 3,3; 3,4; 3,5; 3,6; 3,7; 3,8; 3,9; 4,0 (11 значений);
- влагозащитное покрытие: 2,2; 2,5; 2,6; 3,3; 3,4; 3,5; 3,6; 3,7; 3,8; 3,9; 4,0; 4,2; 4,3; 4,4; 4,5; 4,6; 4,7; 4,8; 4,9; 5,0; 5,1; 5,2 (22 значения).

Толщина диэлектриков:

- влагозащитное покрытие: 8; 12; 15; 18; 22; 25; 30; 35; 40; 45; 50; 55; 60; 65; 70; 75; 150; 155; 160; 165; 170; 175; 180; 185; 190; 195; 200; 205; 210; 215; 220; 225; 230; 235; 240; 245; 250; 255; 260; 265; 270; 275; 280; 285; 290; 295; 300 мкм (47 значений);
- паяльная маска: 20, 40, 80, 100 мкм (4 значения);
- препрег: 50, 66, 100, 105, 132, 150, 180, 198, 200, 210, 250, 264, 315, 330, 360, 420, 525, 540, 720, 900 мкм (20 значений);
- подложка: 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 710, 930, 1000, 1500, 2000 мкм (15 значений).

Параметры проводников:

- толщина: 5, 18, 35 мкм (3 значения);
- ширина: 0,05; 0,075; 0,1; 0,125; 0,15; 0,175; 0,2; 0,25; 0,3 мм (9 значений);
- расстояние между краями: 0,05–1,2 мм (2 значения);

Для каждого исходного набора значений (использованного для аппроксимации) параметров каждой структуры вычислены значения относительной погрешности по формуле $\Delta\tau=((\tau_M-\tau_T)/\tau_T)\cdot 100\%$, где τ_T – табличное (истинное) значение и τ_M – значение, вычисленное по модели. В таблице приведены средние и максимальные значения для разных структур.

Предложенный алгоритм протестирован на различных наборах данных, в которых количество независимых переменных изменяется от 4 до 11. В таблице представлены результаты практической аппроксимации данных с помощью предложенного метода. Например, набор данных для одной из структур состоит из 3428985 строк вида: hPrep ErPrep hCore ErCore t w τ, где первые 6 параметров являются независимыми. Объем данных в распакованном виде составляет 208 Мб.

Для понижения размерности в предложенном методе используется функция polyfit (из MATLAB), которая позволяет получить коэффициенты для $f(x)=\sum_{i=0}^{n}a_ix_i^i$, где n – степень полиномов, a_i – коэффициенты полинома. Из таблицы видно, что структуры 4, 5 и 8 аппроксимируются с приемлемой точностью, тогда как на остальных максимальная ошибка достигает больших значений. Однако средняя ошибка для всех структур меньше 12%. Точность аппроксимации можно повышать за счет увеличения степени аппроксимирующих полиномов. При этом количество коэффициентов будет равно $(p+1)^s$, где p – степень аппроксимирующих полиномов (в примере p = 1, что соответствует линейной регрессии), s – количество независимых параметров (в примере s = 4). Степень для каждого шага можно делать различную, в зависимости от данных на каждом из шагов.

Сравнение предложенного и полиномиального методов аппроксимации

Номер структуры	Степень полиномов		Максимальная относительная ошибка, %		Средняя относительная ошибка, %		Объем коэффициентов, Кбайт	
	Предл.	Полин.	Предл.	Полин.	Предл.	Полин.	Предл.	Полин.
4	5	4	2,0	5,5	0,2	1,0	5	3,8
5	4	4	6,0	6,1	1,0	0,5	61	3,8
6	4	4	521,0	5,5	11,2	0,9	61	3,8
7	4	7	78,0	6,0	1,2	0,4	61	93,8
8	5	4	12,3	6,2	0,9	1,0	185	3,8

Таким образом, разработан новый метод аппроксимации набора данных, у которого средняя ошибка для структур 4 и 8 меньше, чем во втором методе. Однако для данной задачи метод аппроксимации с помощью полиномиального приближения второй степени оказался более приемлем, поскольку позволил с ошибкой не более 6,1% выполнять аппроксимацию значений погонных задержек основных видов линий передачи в широком диапазоне параметров печатных плат.

Исследование выполнено за счет гранта Российского научного фонда (проект №14-19-01232) в ТУСУРе.

Литература

1. Ивахненко А.Г. Помехоустойчивость моделирования / А.Г. Ивахненко, В.С. Степашко. – Киев: Наукова думка, 1985. – 206 с.
2. Tibshirani R. Regression shrinkage and selection via the lasso / R. Tibshirani // Journal of the Royal Statistical Society. – 1996. – Vol. 32. – P. 267–288.
3. Efron B. Least angle regression / B. Efron, T. Hastie, I. Johnstone, R.Tibshirani // The Annals of Statistics. Journal of the Royal Statistical Society. – 2004. – Vol. 32. – P. 407–499.
4. Koza J.R. Genetic Programming: On the Programming of Computers by Means of Natrual Selection. – Oxford, USA: MIT Press, 2012. – 609 p.
5. Стрижневой В.В. Методы выбора регрессионных моделей / В.В. Стрижневой, Е.А. Крымова. – М.: Вычислительный центр РАН, 2010. – 60 с.
6. Разработка математических моделей меандровых линий задержки с оптимальными параметрами: отчет о НИР / рук. Т.Р. Газизов; исполн. Р.И. Аширбакиев и др. – СПб.: ООО «Эремекс», 2013. – 46 с. – №. Р-2013011.

Аширбакиев Ренат Ихсанович
МНС каф. телевидения и управления ТУСУРа
Тел.: 8 (923) 419-40-29
Эл. почта: cr4cpp.2@gmail.com

Мелкозеров Александр Олегович
НС каф. телевидения и управления ТУСУРа
Тел.: 8 (913) 855-42-48
Эл. почта: alexander.melkozerov@gmail.com

Лежнин Егор Владимирович
Лаборант каф. телевидения и управления ТУСУРа
Тел.: 8 (962) 777-00-93
Эл. почта: pavertomato@gmail.com

Ashirbakiev R.I., Melkozerov A.O., Lejnin Eg.V.

An algorithm for data set approximation using coefficients

An algorithm for approximation of a large data set has been proposed. A comparison of the new algorithm and the polynomial approximation for the case of the approximation of the delay per unit length of transmission lines in a wide range of parameters of the basic structures of printed circuit boards has been performed. The performance of the proposed algorithm and the possibilities of its improvement have been described.

Keywords: data approximation, printed circuit board, analytical models, polynomial approximation.