

Моделирование стохастических процессов на языке Perl: Часть 3. Использование модулей Perl для прогнозирования котировок рынка Forex (Марковская модель)

Михаил Смирнов

14.09.2010

Консультант

Независимый разработчик

Третья статья серии "Моделирование стохастических процессов на языке Perl" посвящена созданию программ для вероятностного анализа и прогнозирования котировок рынка Forex, включая марковскую модель. В [предшествующих статьях](#) цикла представлено введение в язык данных Perl PDL (Perl Data Language), используемого в матричном исчислении, а так же рассмотрены примеры программирования прикладных задач от создания модели физического процесса до экономического прогнозирования на основе дискретных цепей Маркова.

1. Элементы вероятностного анализа рынка Forex

Как известно, имеется два основных подхода к анализу валютного рынка - фундаментальный и технический. Фундаментальный анализ определяет тенденции спроса и предложения на основе финансовых и экономических теорий, доминантой, которых является политико-экономические ситуации. Технический анализ имеет дело непосредственно с курсами валют и объемами трейдинга. Формализованность технического анализа позволяет создавать математические модели и специализированное программное обеспечение.

Одним из лидеров Forex в создании on-line технологий является MG Financial Group¹, которая выпустила первую версию on-line трейдинг платформы Deal Station. Эта программа позволяет трейдерам просматривать котировки валют, производить транзакции и отслеживать открытые позиции в режиме реального времени. Одним из конкурентов Deal Station является программный комплекс WinChart компании Straits Index². Среди достоинств этой программы можно отметить возможность on-line обучения приемам технического анализа.

Программное обеспечение технического анализа основывается на графическом представлении котировок валют во времени и направлено на оценку тенденции в будущем.

Технический анализ позволяет предсказывать движения курсов валют на основе оценки данных о прошлых курсах и объемах трейдинга. Алгоритмы и программное обеспечение технического анализа полагаются на эвристические формулы для отслеживания курсов валют и позволяют оценить возможности для продажи или покупки валюты.

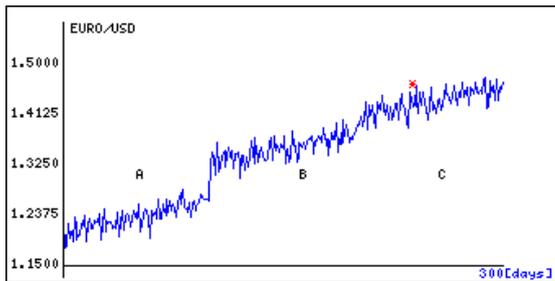
Таким образом, можно заключить, что направленность технического анализа связана с оценкой вероятности последующего события. Технический анализ оперирует важнейшими статистическими характеристиками - средним, стандартным отклонением, статистиками более высоких порядков (моментами). Однако, собственно вероятность остается не востребованной. Вместе с тем, элементы вероятностного анализа могут с успехом использоваться как для вычисления вероятностей движения курсов валют, так и в качестве графического инструмента, который знаком и привычен трейдерам.

В качестве модели изменения курсов валют будем использовать случайный процесс с распределением вероятностей, приближенным к нормальному распределению. Фрагмент программного кода, сочетающий случайный характер процесса и детерминированные тенденции (тренды) движения курсов валют представлен ниже.

```
$N = 300;
$a = 1/(7.333*$N);
for ($k=0; $k<$N/3; $k++) {
    $tr[$k] = 1.2 + $a*$k;
    $tr[$k+$N/3] = 1.3 + $a*$k;
    $tr[$k+2*$N/3] = 1.37 + $a*$k;
}
for ($k=0; $k<$N; $k++) {
    $g[$k] = $tr[$k] + 0.07*&smgauss($k);
}
sub smgauss{
    local(*J)=@_;
    $U=0;
    $AN12=12;
    for($i=$J;$i<$J+$AN12;$i++){
        $x=2*(rand(10)/10 - .5);
        $U += $x;
    }
    return $U/($AN12-1);
},
```

где массив @tr является массивом трендов. В выходном массиве @g сохранена искомая выборка отношения пары валют - EURO и USD. Подпрограмма smgauss() является генератором гауссова шума и возвращает случайные действительные числа x в диапазоне значений $-1 < x < 1$. Размерность выборки равна 300 отсчетам и сопоставима с 24 часовым интервалом. На рис.1 показан результат моделирования дневного графика пары EURO/USD:

Рис. 1. Модельное распределение суточного изменения котировок

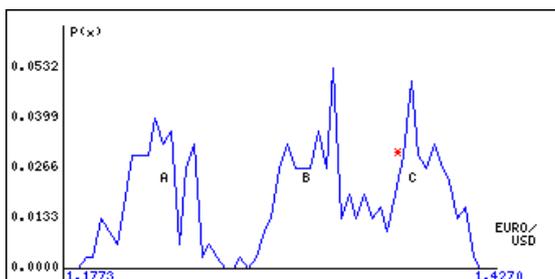


На графике рис.1 выделены три тренда: А, В, С. Все тренды линейно-возрастающие, с коэффициентом наклона $\$a$. Красной звездочкой * отмечено значение 1.3874 в области определения тренда С. Для получения вероятностных оценок требуется построение гистограмм движения курсов валют. Нормирование гистограммы, при соблюдении условия

$$\sum_{j=0}^{M-1} p_j(x) = 1. \quad (1)$$

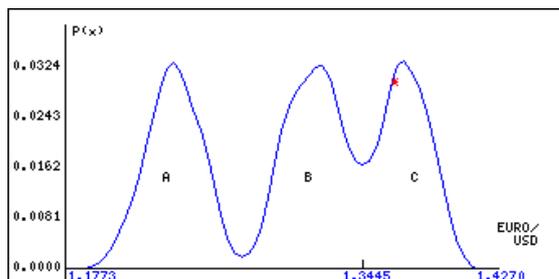
позволяет получить оценку распределения вероятностей $P(x)$ курсов валют, где M область определения случайной величины x . Более подробно, вопросы построения гистограмм рассмотрены в статье³. На рис.2 показано распределение вероятностей модельного распределения курсов валют.

Рис. 2. Распределение вероятностей модельного распределения курсов валют



Следует отметить, что тренды представляют собой систематическую ошибку и в принципе их следует удалить. Однако, в контексте решаемой задачи, они являются полезной информацией. На рис.3 представлен сглаженный вариант вероятностного распределения модельной последовательности курсов валют (сглаживанию подвергается гистограмма).

Рис. 3. Сглаженный вариант вероятностного распределения модельной последовательности курсов валют



Из анализа распределения вероятностей на рис.2 и рис.3, видно четкое разделение (локализация) трендов. В традиционном техническом анализе это не возможно. Впадины значений функции на графике рис.3, образуют "коридор" для анализируемого значения 1.3874 в интервал котировок от 1.3445 до 1.4270.

Рассмотрим теперь пример локализации трендов для реального распределения котировок EURO/USD. На рис.4 показан дневной график котировок EURO/USD за три года с 2005, по 2007 годы включительно. Всего 750 отсчетов (исключены выходные и праздничные дни).

Рис. 4. График 24 часового движения курсов валют



Красной звездочкой отмечено анализируемое значение пары EURO/USD равное 1.2959. На рис.5 и рис.6 показаны графики распределения вероятностей котировок валют без сглаживания и со сглаживанием гистограммы, соответственно.

Рис. 5. Распределение вероятностей суточного движения курсов валют

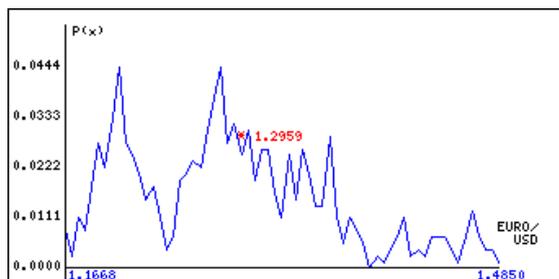
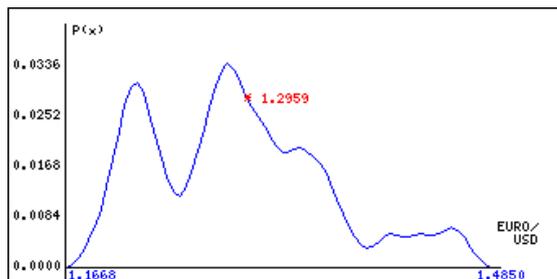
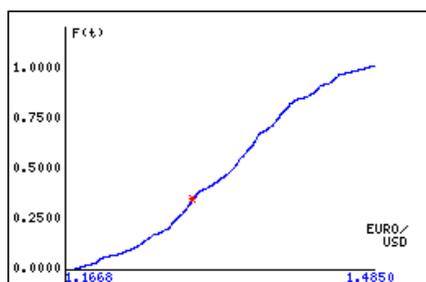


Рис. 6. Сглаженный вариант вероятностного суточного движения курсов валют



Из рассмотрения рис.5 и рис.6 видно, что анализируемое значение принадлежит к группе локально нисходящих трендов. Для графического расчета вероятностей преобразуем распределение на рис.5 в интеграл вероятностей, график которого показан на рис.7.

Рис. 7. Интеграл вероятностей движения курсов валют



Из представленного выше графика следует, что вероятность котировок EURO/USD оказаться в интервале [1.1668; 1.4850] составляет около 35%.

Для вычисления вероятностей с заданной (высокой) точностью нам все-таки потребуется удалить тренды. В традиционном техническом анализе, удаление трендов также практикуется. Оптимальное удаление трендов можно получить с помощью минимизации среднеквадратической ошибки. Программный код, реализующий приближение методом наименьших квадратов представлен ниже:

```
use Time::Local;
use PDL;
use PDL::Matrix;
$Lpp = 3;
&polyDegree(\$Lpp, \@g, \@out);
for($i=0;$i<scalar(@g);$i++){
  $x[$i]=$g[$i] - $out[$i];
}
sub polyDegree{
  local(*poly_degree,*B,*out)=@_;
  $Pn = 2*$poly_degree;
  @c=(0) x $Pn;
  @q=(0) x $poly_degree;
  for($p=0;$p<$Pn+1;$p++){
    for($i=0;$i<scalar(@B);$i++){
      $c[$p] += $i**$p;
    }
  }
  for($p=0;$p<$poly_degree+1;$p++){
    for($i=0;$i<scalar(@B);$i++){
      $q[$p] += $B[$i]*$i**$p;
    }
  }
}
```

```

}}
$k=0;
foreach(@q){
$q2[$k][0]=$_;
$k++;
}
for($m=0;$m<$poly_degree+1;$m++){
for($p=0;$p<$poly_degree+1;$p++){
$dt2[$m][$p]=$c[$p+$m];
}}
$p2 = pdl[[@dt2]];
$det2 = determinant($p2);
for($m=0;$m<$poly_degree+1;$m++){
$r=$p2->slice("$m,");
$r=-s/[\\]/g;
@cv=split(" ",$r);
for($c=0;$c<scalar(@cv);$c++){
$V[$c][$m]=$cv[$c];
}
}
for($z=0;$z<$poly_degree+1;$z++){
for($m=0;$m<$poly_degree+1;$m++){
$VV[$m][0] = $q2[$m][0];
$t=1;
for($i=0;$i<$poly_degree+1;$i++){
if($z != $i){
$VV[$m][$t]= $V[$i][$m];
$t++;
}
}
}
$t1 = pdl[@VV];
$polynoms[$z] = determinant($t1)/$det2;
}
for($z=0;$z<$poly_degree+1;$z++){
$polynoms[$z]=-s/[\\]/g;
}
for($i=0;$i<scalar(@B);$i++){
$zn=-1;
$znc=1;
for($z=0;$z<$poly_degree+1;$z++){
$xx = $i**$z;
$out[$i] += $znc*$polynoms[$z]*$xx;
$znc = $znc * $zn;
}
}
}
}

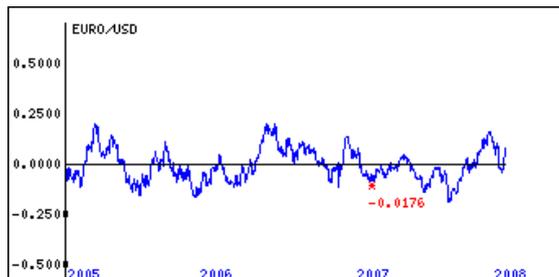
```

Массив @g содержит исходную последовательность курсов валют, показанную на рис.4. Массив @out содержит значения скользящих средних, полученных с помощью приближения методом наименьших квадратов, и является оптимальным, в смысле среднеквадратичной ошибки. Приближение методом наименьших квадратов осуществляется с помощью подпрограмма polyDegree(), в которой используются степенные полиномы произвольной степени. Для вычисления искомым полиномов применяются матричные алгоритмы и подпрограммы Perl библиотеки PDL, что обеспечивает высокую скорость вычислений. Вычисление степенных полиномов @polynoms выполняется на основе определителей Вандермонда⁴:

$$P_n(x) = \begin{vmatrix} |Y_{2n}| \\ |C_{p2n}| \end{vmatrix},$$

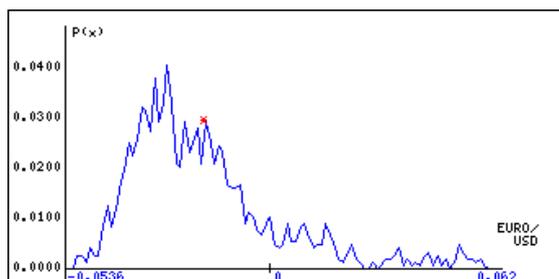
где $P_n(x)$ - полином порядка n , Y_{2n} - алгебраическое дополнение (определитель матрицы $\$t1$), Sp_{2n} - определитель системы $\$det2$. Порядок матриц задается выбором порядка полиномов $\$Lpp$, который в данном варианте равен 3. Последовательность значений в массиве @x представляет собой оптимальный результат удаления трендов и показана на рис.8.

Рис. 8. Результат удаления тренда



На рис.9 показано искомое распределение вероятностей случайного процесса изменения котировок на рис.8.

Рис. 9. Распределение вероятностей случайного процесса изменения котировок



Внешний вид распределения указывает на то, что рассматриваемый случайный процесс является квазинормальным асимметричным распределением (близкое к распределению Пирсона) и имеет смещенный центр распределения. Для оценки асимметрии служат моменты третьего порядка, знак которых указывает либо на правостороннюю или на левостороннюю асимметрию. Последнее обстоятельство может использоваться в принятии решения о движении курсов валют.

Если на рис.9 подсчитать интеграл вероятностей в промежутке от -0.0536 до -0.0176 (красная звездочка), то он будет равен 0.5936. Таким образом, можно утверждать, что изменение котировок валют в пределах приращения от -0.0536 до -0.0176 ожидается с вероятностью около 60%. Если мы подсчитаем интеграл вероятностей в симметричном интервале $[-0.0176; +0.0176]$, то можно утверждать, что приращение котировок валют в интервале $[-0.0176; +0.0176]$ следует ожидать с доверительной вероятностью 34%.

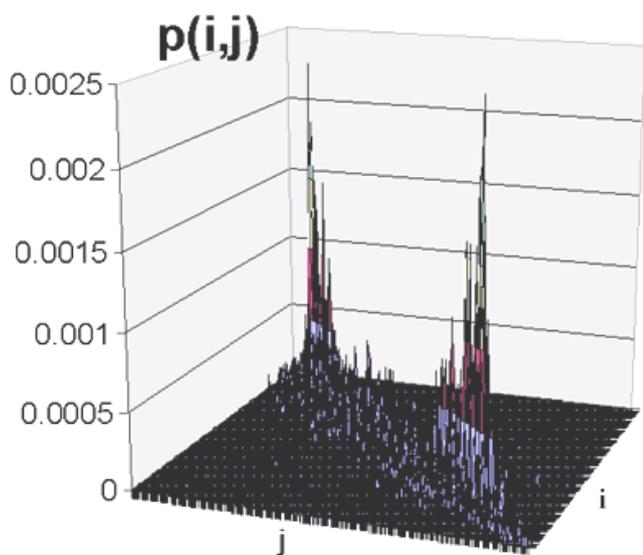
2. Применение марковских переходных вероятностей

Процесс изменения курсов валют является случайным процессом, и содержат детерминированную составляющую или, иначе говоря, тренды валют (рис.4). Такие

случайные процессы называются стохастическими процессами и могут быть описаны с помощью переходных (марковских) вероятностей. Матрицы переходных вероятностей характеризуют процесс переходов (например, во времени) случайной величины x из одного состояния x_i в другое состояние x_j .

Если говорить о процессе движения курсов валют, то переходные вероятности это условные вероятности $p_{i,j}$ того, что в момент времени t текущее значение курса валюты равно j , при условии, что в момент $t-1$ оно составляло величину i . Пример марковского распределения вероятностей $P(i,j)$ курсов валют показан на рис.10. Более подробно вопросы, связанные с программированием матричного исчисления рассмотрены в статье⁵.

Рис.10. Распределение переходных вероятностей $P(i,j)$ курсов валют



Основное свойство марковских вероятностей это память о предшествующих переходах. Это свойство в контексте рассматриваемой задачи можно сформулировать следующим образом: распределение вероятностей $P^t(i,j)$ характеризует вероятность того, что котировка валюты примет значение j , при условии, того, что отстоящая на t шагов (например, t дней) котировка, имела значение i .

Формулировка реальной ситуации может выглядеть следующим образом. В распоряжении трейдера имеются данные о дневном изменении EURO/USD за последние 1,5 - 2 года (например, курсы закрытия). Последовательность значений EURO/USD образует цепь Маркова. Ежедневное изменение курса EURO/USD будем интерпретировать как переход случайной величины x из состояния x_i в состояние x_j . Совокупность всех переходов образует гистограмму второго порядка или матрицу переходов $N(i,j)$:


```
C = 0.3225 - 0.1009 = 0.2216 >0.
```

Таким образом, с учетом точности оценки (2), можно утверждать, что с вероятностью 0.5766 состояние котировки EURO/USD останется прежним.

3. Разработка серверного приложения для предсказания курсов валют

Рассмотрим вариант программы для дневного предсказания пары EURO/USD. В реально действующей программе необходимо наличие базы данным EURO/USD за период 1,5 - 2 и текущее (ежедневное) значение котировки EURO/USD. В простейшем случае база данных может представлять собой текстовый файл с последовательностью значений EURO/USD. Для уменьшения порядка матрицы $N(i,j)$, и соответственно $P(i,j)$, без потери точности вычислений, будем использовать только мантиссу отношения EURO/USD, предполагая, что в ближайшее время отношение EURO/USD не опустится ниже единицы. Выполнение программы разбивается на несколько этапов.

На первом этапе программа считывает текстовую базу данных, разворачивает последовательность значений EURO/USD в цепь Маркова и образует матрицу частот повторения $N(i,j)$. На втором этапе выполняется вычисление матрицы переходных вероятностей $P(i,j)$. На третьем этапе импортируется текущее значение EURO/USD и образуется входной вектор вероятностей s_1 . На четвертом, последнем, этапе вычисляется выходной вектор вероятностей s_2 по формуле (3) и реализуется алгоритм принятия решения.

```
use Time::Local;
use PDL;
use PDL::Matrix;
use LWP::UserAgent;
#-----Stage I-----
open(READ,$file_pair);
@pair_ini=<READ>;
close(READ);
$leng_dec=scalar(@pair_ini);
$coeff = 1000; #--10000;
$i=0;
$min=9999;
foreach (@pair_ini){
if(!$_ || $_==0){ next; }
$dec[$i] = int($coeff*($_ - 1));
if($dec[$i] < $min){ $min = $dec[$i];}
$i++;
}
$max=0;
foreach(@dec){
$_ -= $min;
if($_ > $max){ $max = $_;}
}
$MM = $max + 1;
for($i=0;$i<$MM;$i++){
for($j=0;$j<$MM;$j++){
$np[$i][$j] = 0;
}}
for($k=1;$k<$leng_dec;$k++){
$i=$dec[$k-1];
$j=$dec[$k];
```

```

$np[$i][$j]++;
}
#-----Stage II-----
$zr = zeroes($MM,$MM);
$zr->diagonal(0,1) ++;
$Bp = pdl[[@np]];
$Bp = $Bp + $zr;
$row = $Bp->sumover;
$row=~s/[\[\]]//g;
@VERT = (0) x $MM;
@VERT=split(" ", $row);
$k=0;
foreach $xv(@VERT){
$c[$k][0]=$xv;
$s1a[$k]=0;
$k++;
}
$col = pdl[@c];    #-- 1 x N
$sp = $Bp/$col;
#-----Stage III -----
$ua = LWP::UserAgent->new;
$ua->agent("MSIE 6.0");
$req = HTTP::Request->new(GET => $url);
$res = $ua->request($req);
$buf=$res->as_string;
$current_piar=&eurousdParser($buf);
##- $current_piar=1.2345;
$fin = sprintf "%6.4f", $current_piar;
$dc = int($coeff*($fin - 1) - $min);
$s1a[$dc]=1;
$s1 = pdl[@s1a]; #-- N x 1
#-----Stage IV-----
$pn = $sp x $sp;
$s2 = $s1 x $pn;
$s2=~s/[\[\]]//g;
@LINE = (0) x $MM;
@LINE=split(" ", $s2);
$k=0;
$max2=0;
foreach $xv(@LINE){
if($xv > $max2){ $ind_max=$k; $max2 = $xv; }
}
$Jj=($ind_max + $min)/$coeff;
$sum_l=0;
$sum_r=0;
for($k=0;$k<$ind_max;$k++){
$sum_l += $LINE[$k];
}
for($k=$ind_max+1;$k<$MM;$k++){
$sum_r += $LINE[$k];
}
$excess = $sum_l - $sum_r;
$x = $MM*$excess/2;
$result = 1 + ($ind_max - $x + $min)/$coeff;
$Icorrect = sprintf "%6.4f", $result;
print "Prediction: $Icorrect\n";

```

Переменная `$file_pair` содержит путь к файлу с котировками EURO/USD. Выбор коэффициента `$coeff` определяет точность вычислений. В режиме отладки, вполне возможно задаться значением равным 1000. Наилучшую точность, соответствующую четырехзначной мантиссе, можно получить при `$coeff=10000`. Гистограмма второго порядка $N(i,j)$ накапливается в матрице `@np`. Чтобы исключить нулевые строки, диагональные значения матрицы увеличиваются на 1, что не скажется на статистике.

Для преобразования матрицы $@np$ в матрицу переходных вероятностей требуется вычислить суммы всех строк матрицы. Последняя процедура осуществляется с помощью оператора PDL `sumover`, результатом применения которого будет матрица `$col` размерностью $1 \times \$MM$. Экспорт текущего значения EURO/USD можно получить при реализации http-клиента с помощью библиотеки LWP. Входным параметром http-клиента является URL (`$url`) ресурса, содержащего текущее значение пары EURO/USD. Выделение искомого значения EURO/USD возлагается на подпрограмму парсинга `eurousdParser()`, рассмотрение которой выходит за рамки статьи. В режиме отладки, можно задаться некоторым значением EURO/USD, например `$current_piar=1.2345`.

В качестве критерия принятия решения используется максимальная вероятность `$max2` результирующего вектора `$s2`. Эта вероятность представляет собой максимальную вероятность для всех переходов курса валюты $j \rightarrow i$, при $j=0..MM-1$. Положение максимального значения `$max2` рассчитывается в переменной `$ind_max`. Движение курса валют, и шаг этого движения оцениваются при вычислении эксцесса `$excess`, который численно равен разности интегралов вероятностей справа (`$sum_r`) и слева (`$sum_l`) от максимального значения вероятностей вектора `$s2`. Таким образом, предсказываемое значение `$lcorrect` курса валют является случайной величиной, значение которой зависит от предшествующих значений EURO/USD.

Выводы

Матричные исчисления, реализованные в PDL, позволяют быстро создавать программные модули для многих типовых задач, связанных с оптимизацией ошибок в приближенных вычислениях. Примером может случить оптимальное квадратичное приближение, реализованное в этой статье. Благодаря компактности программного кода и высокой скорости вычислений при использовании модуля PDL можно достаточно просто выполнять вероятностный анализ движения курсов валют рынка Forex. Такие модули Perl как LWP и PDL позволяют создавать не сложные серверные приложения, которые решают задачи предсказания котировок валют, и в том числе, на основе модели Маркова.

Литература

¹ MG Financial Group - www.mgforex.com.

² Straits Index - www.straitsindex.com.

³ [Применение Perl для построения и преобразования гистограмм сигналов](#). - //Цикл статей "Моделирование методов цифровой обработки сигналов на языке Perl".

⁴ Теория интерполирования и приближения функций. В.А. Гончаров.

⁵ [Матричные преобразования на Perl в решении прикладных задач](#). -//Цикл статей "Моделирование стохастических процессов на языке Perl".

Об авторе

Михаил Смирнов

Михаил Смирнов — независимый web-программист и разработчик прикладного программного обеспечения (Perl, PHP, Borland Builder C++) в области информационных технологий, кандидат технических наук. Основные направления деятельности изложены на личном сайте <http://www.smirnov.sp.ru> и затрагивают области синтеза и цифровой обработки изображений, цифровой голографии, распознавания образов, анализа и моделирования случайных процессов.

© Copyright IBM Corporation 2010

(www.ibm.com/legal/copytrade.shtml)

Торговые марки

(www.ibm.com/developerworks/ru/ibm/trademarks/)