

Programação genética aplicada no processo de previsão: um estudo de caso aplicado em chamadas de uma central de teleatendimento

Cláudio Lúcio do Val Lopes, Gustavo Henrique Passini Santos e
Flávio Vínicius Cruzeiro Martins

1 de Novembro de 2017

CEFET-MG

1. Introdução
2. Previsão de chamadas
3. Estudo de caso
4. Resultados
5. Comentários e observações finais

Introdução

Centrais de teleatendimento

Conceitos

Uma central de teleatendimento é um conjunto de recursos (computadores, equipamentos de telecomunicações e pessoal para atendimento) que está habilitado a prover serviços via telefone;

As centrais de teleatendimento buscam atingir uma determinada qualidade (níveis de serviço) sujeitos a recursos finitos (orçamentos, atendentes, linhas telefônicas, equipamentos);

Previsão de chamadas

A quantidade de chamadas nas centrais de teleatendimento são observadas em intervalos diários e intradiários. É um tema importante para a operação das centrais:

- programação de trabalho de atendentes com escalas e turnos;
- nível de serviço;
- contratações de mão de obra;
- realização de treinamentos dentre outros.

Programação genética para previsão de chamadas

- Verificar a aplicabilidade do paradigma da programação genética para previsão de chamadas intradiárias em centrais de teleatendimento;
- Validar a programação genética em algum caso de teste real;
- Comparar a programação genética com alguma abordagem clássica para previsão de chamadas em centrais de teleatendimento;
- Observar oportunidades e melhorias para generalizar o caso teste.

Previsão de chamadas

Modelos clássicos

Em algumas centrais de teleatendimento os níveis de serviço são monitorados de 15 em 15 minutos, de forma horária(intradiária);

Na presença de dados intradiários alguns modelos são utilizados na literatura: ARMA, alisamento exponencial e modelos Bayesianos;

Modelo ARMA

- A quantidade de chamadas é uma variável aleatória discreta com distribuição de Poisson e uma transformação na quantidade de chamadas é aplicada;
- Os modelos Gaussianos do tipo auto-regressivo e de médias móveis ARMA são adequados:

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + \dots + \phi_p(X_{t-p} - \mu) + \epsilon_t - \theta_1\epsilon_{t-1} - \dots - \theta_q\epsilon_{t-q}$$

- Em que $\mu, \phi_1 \dots \phi_p, \theta_1 \dots \theta_q$ são parâmetros reais e $\epsilon_t \approx N(0, \sigma^2)$ com tais características: temporalmente homogêneo, estacionário e sem dependência temporal. A média do processo é representado por μ .

Programação genética(PG) aplicada no processo de previsão

Trabalhos buscam criar modelos de previsão utilizando a temática de populações, através da PG, para executar previsão de series temporais;

Programas são criados utilizando a ideia de regressão simbólica em que a estrutura das funções é flexível e pode ser linear ou não linear;

Regressão simbólica

- A estrutura das funções é flexível e pode ser linear ou não linear;
- A função é flexível e obtida por algum processo de alteração de sua estrutura, seus termos e coeficientes.

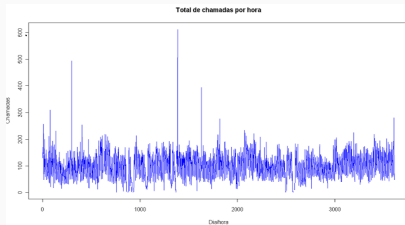
Estudo de caso

Base de dados

Base de dados pertencente a uma instituição financeira do Estado de Israel que oferece vários tipos de serviços;

A quantidade de chamadas é ,em média, de *30.000* a *40.000* chamadas por mês;

A base de dados utilizada para fazer o ajuste dos modelos de previsão apresenta *3.612* observações oriundas de um total de *356.638* chamadas da base original;



Objetivo

Prever 2 semanas de chamadas intradiárias;

Modelo clássico utilizado

O modelo ARMA proposto utiliza fatores fixos das horas e dias em conjunto com um modelo $ARMA(2,2)$;

Os valores p e q do modelo $ARMA(2,2)$ foram obtidos através da análise das funções de autocorrelação e de autocorrelação parcial:

$$\hat{y}_t = \sum_{d=1}^D \alpha_d I_d + \sum_{k=1}^K \beta_k I_k + \Theta_0 + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2}$$

Utilizou-se: N_{dk} como a quantidade de chamadas no dia $d \in \{1, \dots, D\}$, e da hora k , $k \in \{1, \dots, K\}$ e $y_{dk} = y_t$, $t \in \{1, \dots, T\}$ e $T = 3612$;

Os coeficientes α_d e β_k são relativos às variáveis indicadoras para os dias dos meses e horas do dia, respectivamente.

Modelo de programação genética

Modelar os componentes lineares e não lineares da série;

Nós terminais

- Conjunto de constantes obtidos através de um processo de extração de características da própria série: desvio padrão, média, mediana e outros (com um total de 10 argumentos);
- Constantes aleatórias com valores com distribuição de *Poisson* (λ foi a média geral da série - 32 constantes aleatórias);
- Argumentos oriundos da base de dados, representados pela defasagem da série e valores lógicos que indicavam a hora do dia em questão (25 argumentos);

Nós não terminais

- Funções de ponto flutuante como: soma, subtração, multiplicação, cosseno e seno;
- Funções lógicas como: 'se', 'e', 'ou', negação e igualdade;

Modelo de programação genética - Resumo

Detalhes do modelo de programação genética adotado:

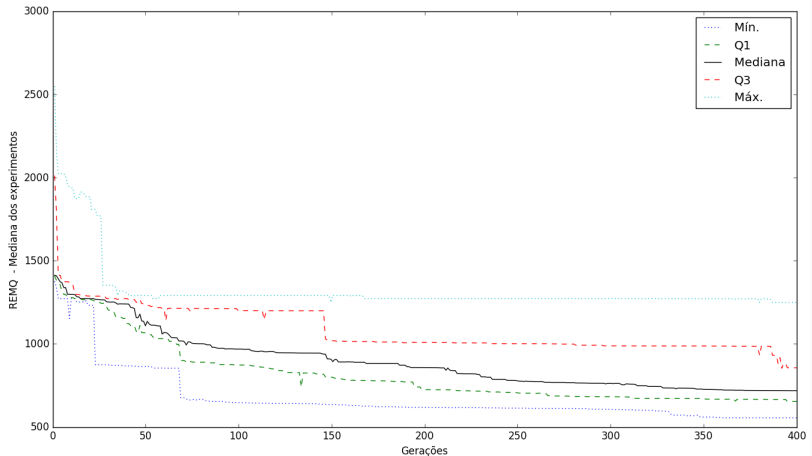
Característica do algoritmo	Descrição
<i>Fitness</i>	Raiz da média do erro quadrático - RMEQ
Nós terminais	Constantes e argumentos
Nós não terminais (funções)	Algumas funções de ponto flutuante e funções lógicas
Tamanho máximo da Arvore	De 3 até 30 níveis
Seleção	Torneio com 3 elementos
Cruzamento	Ponto único com taxa de 0.8
Mutação	Uniforme com taxa de 0.1
Condição de parada	400 gerações
Tamanho da população	100 indivíduos
Elitismo	Não, mas o DEAP mantém o "Hall of fame"

O modelo foi implementado utilizando um *framework* para algoritmos evolucionários chamado *Distributed Evolutionary Algorithm Python - DEAP*;

Modelo de programação genética - Curva de convergência

Foram realizados um total de 10 experimentos;

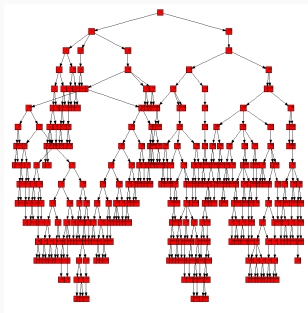
RMEQ - Mediana do erro médio quadrático dos experimentos: diminuição com o passar das gerações, satisfazendo o objetivo do estudo de caso:



Modelo de programação genética - árvore sintática

Averiguou-se boa diversidade nos valores que caracterizam os nós terminais e não terminais;

Melhor solução global gerou uma função, representada em árvore sinática com 16 níveis e 388 nós;



Para acessar a função gerada pela solução veja o link em:

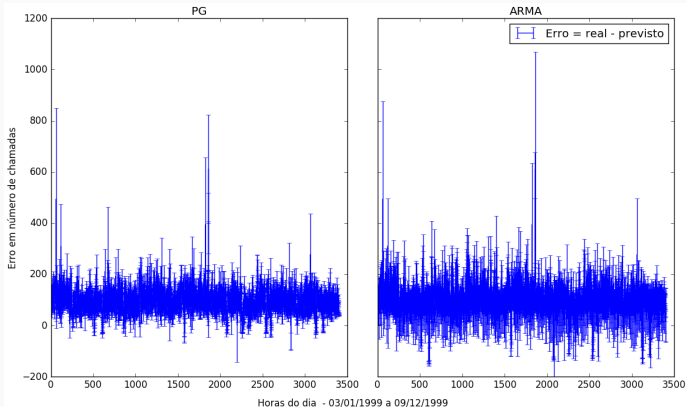
<https://1drv.ms/t/s!AiAix-tWooVygYI0Wxe5iLPSgt0ZGA>

Resultados

Validação e comparação dos modelos - base de treinamento

Averiguar a acurácia da previsão realizada pelos modelos utilizando as previsões geradas pela melhor solução da PG e os valores do ajuste do modelo ARMA *versus* a base de dados real:

O gráfico mostra o erro em número de chamadas quando comparado com o valor real da base de treinamento;



Validação e comparação dos modelos - base de treinamento

Resumo das estatísticas de comparação da base de treinamento:

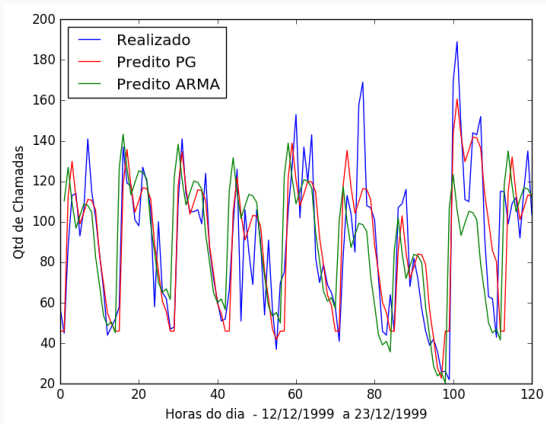
	Modelo ARMA	Modelo PG melhor solução
Soma dos quadrados dos resíduos	6678236	1883564
<i>RMEQ</i> Média	44.30	23.57
<i>RMEQ</i> Desvio padrão	86.86	54.66
<i>RMEQ</i> Quartil 1	10.44	5.30
<i>RMEQ</i> Mediana	22.17	11.81
<i>RMEQ</i> Quartil 3	40.25	21.03

Validação e comparação dos modelos - base de teste

Utilizar os modelos para gerar os valores previstos para as próximas duas semanas (12/12 a 23/12);

Ressalva que nenhum dos dois modelos utilizou os dados deste período;

O gráfico mostra os valores reais e valores gerados pelos dois modelos;



Validação e comparação dos modelos - base de teste

Resumo das estatísticas de comparação da base de teste:

	Modelo ARMA	Modelo PG melhor solução
Soma dos quadrados dos resíduos	167500	49257
<i>RMEQ Média</i>	29.11	19.61
<i>RMEQ Desvio padrão</i>	18.73	13.33
<i>RMEQ Quartil 1</i>	15.51	4.26
<i>RMEQ Mediana</i>	26.26	10.97
<i>RMEQ Quartil 3</i>	40.32	21.51

Comentários e observações finais

Comentários e observações finais

Comentários

- A abordagem da programação genética para previsão de chamadas de centrais de teleatendimento é válida;
- Resultado apresentado para PG é superior (caso de teste) comparado com método utilizado na literatura;
- No modelo da PG não existem premissas para a validade do modelo (lida com características lineares e não lineares dos dados);
- Tempo gasto no processo de evolução: necessidade de arquitetura de computação paralela e distribuída.

Observações finais

- Expandir o atual trabalho para outras séries discretas acoplando algum mecanismo de extração automática de características;
- Melhoria no processo de ajuste para pontos discrepantes;
- Testar outras estratégias de algoritmos genéticos.

Comentários e dúvidas ???