

Modelo de previsão de partida de ônibus utilizando cadeias de Markov de alcance variável

Maria das Vitórias Alexandre Serafim ¹

Manuel Rivelino Gomes de Oliveira ²

Divanilda Maia Esteves ³

Paulo José Duarte-Neto ⁴

1 Introdução

O estudo relacionado com a chegada de ônibus em seu ponto de partida é de grande importância para fornecer informações confiáveis aos usuários sobre os horários do mesmo, e com isso poder reduzir o tempo de espera dos passageiros.

Diante disto, alguns modelos são utilizados para procurar prever a chegada dos ônibus em suas paradas, mas isto é difícil de estimar devido ao fato de existir diversos fatores que podem causar o atraso do mesmo, e com isso, aumentar o tempo médio de espera dos usuários, prejudicando a qualidade do serviço (Patnaik et al., 2004).

Em relação a modelagem da chegada de ônibus a suas paradas, alguns modelos como o de regressão (Patnaik et al., 2004) e os modelos de RNA e RNN (rede neural recorrente) (Raut et al., 2012) foram desenvolvidos para estimar os horários de chegada de ônibus a seus pontos de partida. No entanto, este fenômeno também pode ser explicado por um modelo de cadeias de Markov de Alcance Variável.

Uma cadeia de Markov é um tipo especial de processo estocástico com espaço de estados discretos, onde a probabilidade de uma variável em um intervalo de tempo futuro, dado um estado presente, não é influenciado pelos estados visitados em períodos passados. O conceito de cadeia de Markov pode ser estendido para o de cadeia de Markov de ordem k , onde agora o estado futuro do sistema será influenciado pela informação dos k momentos passados mais recentes.

Uma vez que se decida pelo uso de uma cadeia de Markov, o primeiro passo é estimar as probabilidades de transição entre os estados, ou seja, os parâmetros do modelo. Para se conhecer o comportamento de uma cadeia de Markov é preciso saber como a cadeia começa (distribuição inicial) e como ela evolui (probabilidades de transição). Aqui surge um problema: o número de

¹Mestranda em Biometria e Estatística Aplicada - DEINFO/UFRPE. e-mail: mv.barros@hotmail.com

²Doutorando em Biometria e Estatística Aplicada - DEINFO/UFRPE

³Professora Adjunta - DE/UEPB

⁴Professor Adjunto - DEINFO/UFRPE

parâmetros do modelo cresce muito rápido à medida que aumentamos o tamanho da cadeia. Isto é um problema porque quanto mais parâmetros envolvidos no modelo, maior deverá ser a amostra para que possamos estimá-los.

Assim, as cadeias de Markov de Alcance Variável são um tipo mais flexível de cadeias de Markov (Rissanen, 1983). Neste caso, o presente é influenciado por uma parte variável do passado. Isso leva a uma variedade bem maior de modelos possíveis. Essas cadeias foram bem trabalhadas por (Bühlmann e Wyner, 99) e desde então surgiram aplicações em diversas áreas, como por exemplo genética e linguística.

O objetivo deste trabalho é modelar a sequência de horários diários de chegada de uma determinada linha de ônibus, para tentar prever se o ônibus sairá ou não no horário certo em um dado dia, com base no que ocorreu nos dias anteriores.

2 Material e Métodos

Os dados utilizados neste trabalho são referentes a um experimento realizado na área de transporte público, observados por um cliente da empresa em questão. As observações foram feitas na Agência de Passagens da Empresa Viação São José, em Campina Grande, PB. Para realização do estudo admitiu-se horário de partida às 18:00 horas, sendo que, para que o ônibus esteja antecipado é preciso que ele chegue na agência antes de 17:55 horas e para que esteja atrasado chegue depois de 18:05 horas, dessa forma ele só estará na hora certa se chegar entre 17:55 horas e 18:05 horas. A amostragem foi realizada pela observação direta de chegada do ônibus. As observações iniciaram-se no dia 20 de Novembro de 2013 e terminaram no dia 20 de Dezembro de 2013, sendo contados só os dias úteis, totalizando assim 23 dias observados. Para avaliação da pontualidade do ônibus admitiu-se que o processo gerador de uma sequência aleatória de atrasos, antecipações e horário certo em um determinado dia deve ser influenciado pelo que aconteceu nos dias anteriores e, portanto, pode ser descrito por uma cadeia de Markov.

Após decidir pelo uso das cadeias de Markov, o próximo passo será encontrar as probabilidades de transição dos estados e a distribuição inicial da cadeia. Para isto, foi utilizado o algoritmo do contexto (Bühlmann e Wyner, 1999) que está implementado na biblioteca VLML do pacote estatístico R. Além da estimação, também foram utilizadas no trabalho algumas amostras simuladas de cadeias de Markov. O algoritmo para simulação de tais amostras foi implementado utilizando o pacote estatístico R.

3 Resultados e Discussões

Para avaliação da pontualidade do ônibus considerou-se que o processo gerador da sequência aleatória de atrasos, antecipações e horário certo pode ser descrito por uma cadeia de Markov, com as seguintes características:

1) X_n é o estado do sistema no n-ésimo dia de observação;

2) Será usada a seguinte notação:

P = Ônibus antecipado; C = Ônibus no horário certo; A = Ônibus atrasado

3) Desta maneira a modelagem será:

- $X_n = 0$ se ocorre A no dia n;
- $X_n = 1$ se ocorre P no dia n;
- $X_n = 2$ se ocorre C no dia n.

Após modelar a cadeia, conforme descrito anteriormente, calculou-se a matriz de probabilidades de transição e a distribuição inicial da mesma considerando uma cadeia de ordem 1, sendo assim, a matriz de transição da cadeia foi:

$$P = \begin{bmatrix} 0.64 & 0 & 0.36 \\ 0 & 0.33 & 0.67 \\ 0.50 & 0.25 & 0.25 \end{bmatrix}$$

E a distribuição inicial foi: $\pi_0 = (0.50 \ 0.14 \ 0.36)$,

Também é possível construir um grafo das transições, pois se conhece todas as probabilidades de transição dos estados.

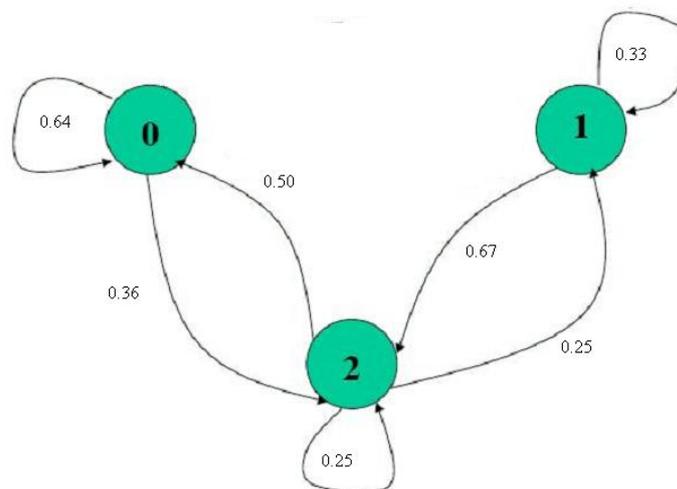


Figura 1: Grafo das Transições entre os estados da cadeia

Verifica-se através da Figura 1 que todos os estados se comunicam, portanto a cadeia é irreduzível. Como a cadeia é irreduzível e o espaço de estados finito, então todos os estados são recorrentes.

Neste trabalho foi utilizado alguns conceitos de cadeia de Markov de alcance variável para verificar o comportamento da variável em estudo (situação de pontualidade do ônibus). A idéia era observar se era viável o uso de uma VLMC para modelar os dados obtidos. Para verificar tal comportamento foi usado o Algoritmo Contexto, este algoritmo está implementado no ambiente de programação estatística R, através de um pacote chamado VLMC. O algoritmo foi aplicado nos dados reais do trabalho, mas como se tem uma amostra pequena, decidiu-se usar também o algoritmo em dados simulados no R. A simulação dos dados em questão deu-se através de um Algoritmo que gera cadeias de Markov com a mesma distribuição inicial e a mesma matriz de transição da cadeia original, o que irá diferenciar da cadeia original é só o tamanho da amostra, ou seja, qualquer amostra gerada através deste algoritmo terá, teoricamente, a mesma lei de probabilidade da cadeia primitiva. Optou-se por tal simulação para observar o comportamento da variável quando se tem uma amostra de tamanho grande, já que a amostra original em questão é pequena. Verificou-se que para as amostras de tamanho até 50, não há informação suficiente para que se detecte uma estrutura de dependência. Para amostras maiores que 50 tal estrutura começa a aparecer.

4 Conclusões

No presente trabalho, evidenciou-se a praticidade da utilização de modelagem de uma variável através de cadeias de Markov. O modelo baseado em cadeias de Markov de Alcance Variável é capaz de prever a chegada do ônibus em seu ponto de partida, utilizando uma pequena amostra da sequência de horários diários de partida do mesmo, sendo possível fazer previsões precisas sobre a pontualidade do ônibus utilizando modelagem com cadeias de Markov de Alcance Variável. Verificou-se que nem sempre é possível detectar a dependência que um estado tem com relação a outro quando se trabalha com amostras pequenas, por isto a necessidade de simular cadeias de Markov com amostras maiores. Após fazer algumas simulações no R, verificou-se que a amostra coletada foi pequena e que no momento que se tem uma amostra maior fica mais fácil visualizar a ordem da cadeia e a função contexto, ou seja, quanto do passado influencia no presente de uma variável.

Referências

- [1] ANDERSON, T. W.; GOODMAN, L. A. **Statistical inference about Markov chains.** Ann. Math. Stat., v. 28, p. 89 - 109, 1957.

- [2] BÜHLMANN, P.; WYNER, A. J. **Variable length Markov chains**. The Annals of Statistics, v. 27, p. 480 - 513, 1999.
- [3] CSISZ _ AR, I.; TALATA, Z. **Context tree estimation for not necessarily _ nite memory processes, via BIC and MDL**. IEEE Transactions on Information Theory, v. 52, p. 1007 - 1016, 2006.
- [4] FERRARI, F.; WYNER, A. **Estimation of general stationary processes by variable length Markov chains**. Scandinavian Journal of Statistics, v. 30, n. 3, p. 459 - 480, 2003.
- [5] PATNAIK, J.; CHIEN, S. and BLADIKAS, A. **Estimation of Bus Arrival Times Using APC Data**. Journal of Public Transportation, v. 7, n. 1, p. 1 - 20, 2004.
- [6] RAUT, R. D.; GOYAL, V. K. **Public transport Bus Arrival Time Prediction with Seasonal and Special Emphasis on Weather Compensation changes using RNN**. International Journal of Advanced Research in Computer and Communication Engineering, v. 1, p.378 - 382, 2012.
- [7] RISSANEN, J. **A universal data compression system**. IEEE Transactions on Information Theory, IT-29, p.656-664, 1983.