

Proposta de aprimoramento da relação Z-R de estimativa de precipitação utilizando aprendizagem de máquina

Fernanda F. Verdelho^{1,2}, Marco A. Z. Alves¹, Luiz E. S. Oliveira¹, Cesar Beneti²

¹Depto. de Informática – Universidade Federal do Paraná (UFPR) - Curitiba/PR

²Serviço Meteorológico do Paraná (SIMEPAR) - Curitiba/PR

{fernandaverdelho@gmail.com, mazalves@inf.ufpr.br}

Resumo. *Apresentamos neste artigo uma proposta de uso e otimização de algoritmos de aprendizagem de máquina, tais como Random Forest e Gradient Boosting, para melhorar o desempenho e acurácia da Estimativa Quantitativa de Precipitação (QPE), utilizando dados provenientes de radar meteorológico e estações pluviométricas.*

1. Proposta de Pesquisa

A precipitação é a principal variável nos processos hidrológicos e sua medição é de extrema importância para aplicações socioeconômicas, incluindo a agricultura, geração de energia, abastecimento de água e gestão de riscos associados a eventos extremos.

O ato de obter dados de precipitação requer vários sistemas e redes de sensoriamento remoto que englobam medidores pluviômetros, radares meteorológico e satélites. Dentre esses medidores, o mais utilizado para gerar a Estimativa Quantitativa de Precipitação (QPE) é a baseada em radar meteorológico, ela é obtida por uma Relação (Z-R) apurada através dos dados de intensidade da precipitação (R) (i.e. adquiridas por pluviômetros) e as variáveis polarimétricas do radar de dupla-polarização, tais como refletividade (Z), (ZDR) e (KDP) (i.e. variáveis que definem como radar observa a gota de chuva).

Um QPE de alta resolução e alta precisão é um componente chave para muitas aplicações - como na agricultura, gerenciamento de recursos hídricos e alerta da defesa civil. Ademais, essencial para geração de energia, principalmente no estado do Paraná, um dos maiores geradores de energia hidrelétrica do país.

Mesmo sendo o método tradicional, a relação Z-R acaba sofrendo com incertezas geradas pelo sensor, condições sinóticas, hidrológicas, geográficas e distribuição do tamanho da gota. Alguns centros veem aplicando esta técnica para obter essa relação usando métodos estatísticos ao longo de uma escala de tempo climática e então aplicada à estimativa quantitativa da precipitação.

Com os diversos avanços de poder computacional e algorítmico a aprendizagem de máquina recentemente explodiu em popularidade em muitos campos, incluindo em aplicações em meteorologia [McGovern et al. 2019].

Os trabalhos presentes na literatura não utilizam de todas as variáveis polarimétricas disponíveis no radar. Contudo, resultados obtidos por [Shin et al. 2019], mostrou a eficiência do uso de aprendizado de máquina, com resultados que superaram a acurácia da relação Z-R tradicional, em alguns fenômenos de precipitantes.

Para atingir o objetivo proposto, se faz necessária a execução de experimentos para identificar *features* com maior impacto no desenvolvimento de uma relação Z-R. Empregando algoritmos de aprendizagem máquina, tais como algoritmos de *ensemble* como *Random Forest* [Breiman 2001] e *Gradient Boosting* [Friedman 2001]. Algoritmos como alternativa para implementar um modelo de QPE, para melhorar o método tradicional da obtenção da relação Z-R e conseguir uma resolução espacial e temporal significativamente melhorada.

Ambos algoritmos não executam diversas árvores em paralelo. Com um *dataset* compreendido entre os anos 2018 e 2021 e volume grande de informações (i.e. dados de pluviômetros e radar), surgiu a necessidade de otimizar o processamento. Para o algoritmo *random forest*, se propõem usar instruções predicadas [Dulong 1998], que permite a um programa executar instruções e apenas gravar seus resultados dependendo de uma condicional (predicado).

Com o algoritmo *gradient boosting*, foi considerado o uso do algoritmo *Xgboost* [Chen et al. 2015], devido à sua simplificação ao ser paralelizável. Onde se avalia a paralelização em cada estrutura condicional da árvore usando *openMP*, para criar ramificações de forma independente.

Ao final deste trabalho, espera-se identificar, o algoritmo de aprendizado de máquina com melhor acurácia, menor custo de processamento e melhor precisão, ou seja, capaz de superar o desempenho do método tradicional de QPE e gerando um campo meteorológico consistente.

Agradecimentos

Agradeço ao Sistema Meteorológico do Paraná (SIMEPAR) pelo suporte financeiro e ao grupo High Performance and Efficient Systems (HiPES) da UFPR, pelo apoio.

Referências

- Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.
- Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., et al. (2015). Xgboost: extreme gradient boosting. *R package version 0.4-2*, 1(4):1–4.
- Dulong, C. (1998). The ia-64 architecture at work. *Computer*, 31(7):24–32.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29:1189–1232.
- McGovern, A., Lagerquist, R., Gagne, D. J., Jergensen, G. E., Elmore, K. L., Homeyer, C. R., and Smith, T. (2019). Making the black box more transparent: Understanding the physical implications of machine learning. *Bulletin of the American Meteorological Society*.
- Shin, J.-Y., Ro, Y., Cha, J.-W., Kim, K.-R., and Ha, J.-C. (2019). Assessing the applicability of rf, sgb model, and elm methods to the quantitative precipitation estimation of the radar data: A case study to gwangdeoksan radar, south korea, in 2018. *Advances in Meteorology*, 2019:6542410.