

AVALIAÇÃO DE CLASSIFICADORES PARA O MONITORAMENTO DE PRECIPITAÇÃO EM ÁREAS AGRÍCOLAS

GREICE MARTINS DE FREITAS¹, ANA MARIA HEUMINSKI DE ÁVILA ¹,
HILTON SILVEIRA PINTO ¹, JOÃO PAULO PAPA²

¹ CEPAGRI/UNICAMP – Centro de Pesquisas Meteorológicas e Climáticas Aplicadas à Agricultura

CEP 13083-886 - Campinas - SP, Brasil

{ greice, avila, hilton }@cpa.unicamp.br

² IC/UNICAMP - Instituto de Computação

CEP 13083-971 - Campinas - SP, Brasil

jpaulo@ic.unicamp.br

Apresentado no XV Congresso Brasileiro de Agrometeorologia – 02 a 05 de julho de 2007 – Aracaju – SE

RESUMO: Dentre os fatores que mais influenciam a produção agrícola no país, a precipitação pode ser citada como um dos mais influentes, por ter relação direta com o balanço hídrico, o qual afeta praticamente todos os aspectos relacionados ao desenvolvimento e produtividade das culturas agrícolas. Faz-se então necessário um monitoramento contínuo das variáveis meteorológicas nestes locais de produção. Métodos de estimativa de precipitação utilizando canais espectrais dos satélites meteorológicos têm sido cada vez mais utilizados para monitoramento de precipitação nestes locais. O presente trabalho tem por objetivo correlacionar informações disponíveis no canal espectral infravermelho do satélite meteorológico GOES-12, com a refletividade obtida pelo radar do IPMet/UNESP localizado no município de Bauru, visando o desenvolvimento de um modelo para a detecção de nuvens de precipitação baseado em classificadores de padrões. Nos experimentos foram comparados três algoritmos de classificação: redes neurais (ANN), máquinas de vetores de suporte (SVM) e floresta de caminhos ótimos (OPF), o qual obteve melhor resultado que os demais, tanto em eficácia quanto em efetividade.

PALAVRAS-CHAVE: Floresta de Caminhos Ótimos, GOES, Estimativa de Precipitação.

CLASSIFIERS EVALUATION FOR THE MONITORING OF THE AGRICULTURAL AREAS PRECIPITATION

ABSTRACT: Inside the factors that have most influence in the agricultural production of the country, the precipitation can be addressed as one of the most influential, due to its direct relationship with the hydric balance, which affects all the aspects related to the development and productivity of the agricultural cultures. In that way, it is necessary a continuous monitoring of the meteorological variables in these places. Precipitation estimation methods using spectral channels of the meteorological satellites have been using in the last years. The present work has as objective to correlate the informations available from GOES-12 satellite infrared spectral channel with the reflectivity obtained by IPMet/UNESP radar located at Bauru city, aiming the development a model for precipitation clouds detection based on patterns classification. In the experiments we compared tree classification algorithms: Artificial Neural Networks (ANN), Support Vector Machines (SVM) and Optimum Path Forests (OPF), which have the best results in terms of accuracy rate and running time.

KEYWORDS: Optimum Path Forest, GOES, Precipitation Estimation

INTRODUÇÃO: A produção agrícola está diretamente ligada às condições climáticas, podendo, desta forma, ser substancialmente afetada desde a época de plantio até a colheita. A falta de informações sobre o balanço hídrico do solo não permite uma orientação técnica adequada sobre a

aptidão climática das culturas às regiões, determinando espécies e variedades a serem cultivadas e as melhores épocas de plantio.

A estimativa de precipitação a partir de dados transmitidos por satélites, em função de sua boa cobertura espacial e temporal, apresenta-se como uma opção aos métodos convencionais, devido ao alto custo de instalação e manutenção de estações de superfícies e radares meteorológicos. Porém, as medidas obtidas por tais sensores remotos são indiretas, o que motiva o desenvolvimento de modelos que relacionem características dos canais espectrais dos satélites a parâmetros associados a estas informações.

O objetivo deste trabalho é desenvolver um modelo que permita correlacionar informações disponíveis nos canais espectrais do satélite meteorológico *Geostationary Operational Environmental Satellite* (GOES) com a refletividade obtida por radar visando o desenvolvimento de modelos de identificação de nuvens de precipitação através de classificadores de padrão. Dentro deste objetivo, está compreendida a comparação entre três algoritmos de classificação: Redes Neurais Artificiais (ANN) (HAYKIN, 1994), Máquinas de Vetores de Suporte (SVM) (VAPNIK, 1992) e Florestas de Caminhos Ótimos (OPF) (PAPA, 2007).

MATERIAL E MÉTODOS: Informações sobre a variabilidade espacial e temporal de índices de precipitação têm fundamental importância em inúmeros campos de aplicação, especialmente aqueles relacionados à agricultura. Devido a grande quantidade de dados produzidos por radares e satélites meteorológicos, tornou-se necessário o desenvolvimento de vários sistemas automáticos para predição de índices de precipitação. O principal objetivo de tais metodologias é construir modelos de precipitação baseados nas diferentes informações oriundas das imagens de diversas bandas espectrais. Tais modelos são interpretados por algoritmos de classificação de padrões, devido principalmente a sua complexidade e não linearidade. Grande parte dos trabalhos utiliza técnicas baseadas em Redes Neurais Artificiais (*Artificial Neural Networks* - ANN) para a predição de precipitação (MCCULLAGH *et al.*, 1995; LEE *et al.*, 1998; PALMEIRA *et al.*, 2004).

MCCULLAGH *et al.* (1995) foi um dos pioneiros a desenvolver um modelo de redes neurais para estimação de chuva, usando dados dos canais visível e infravermelho para estimativa de precipitação acumulada em seis horas. LEE *et al.* (1998) utilizou outras configurações de redes neurais com o objetivo de estimar quantidade de precipitação diária. PALMEIRA *et al.* (2004) utilizou uma abordagem probabilística de precipitação, o qual correlaciona a presença de raios e partículas de gelo com a presença de chuva. Para a realização de seu trabalho, usou dados do canal infravermelho do satélite GOES. Recentemente, alguns trabalhos têm utilizado outros algoritmos para interpretação de modelos climáticos, devido à instabilidade e demora na convergência das redes neurais. Devemos atentar ao fato que sistemas de predição de precipitação podem ser incluídos na categoria de sistemas críticos, ou seja, sistemas os quais possuem fundamental importância para economia e sobrevivência. Possíveis pontos de precipitação podem ser identificados, porém um algoritmo eficiente e rápido torna-se desejável. UMEHARA *et al.* (2005) propôs um sistema para estimação de chuvas baseado em Máquinas de Vetores de Suporte (*Support Vector Machines* - SVM). Tal abordagem tem demonstrado certa superioridade em relação às redes neurais em várias áreas de aplicação. Entretanto, a teoria na qual a SVM está embasada pode torná-la impraticável ou com resultados não satisfatórios. Uma SVM propõe resolver algum problema de classificação qualquer (classificar um dado como sendo precipitação ou não, por exemplo) em um espaço amostral de alta dimensão, ou seja, sua teoria é comprovada em espaços de dimensão infinita, o que nem sempre é válido na prática.

Dentro de tais limitações, o presente trabalho aborda também a utilização de um novo método de classificação de padrões chamado Florestas de Caminhos Ótimos (*Optimum Path Forest* - OPF) (PAPA *et al.*, 2007), o qual demonstrou superioridade em relação às Redes Neurais e Máquinas de Vetores de Suporte. O algoritmo OPF é uma técnica de classificação baseada na teoria de grafos, tendo como principais vantagens a rapidez e o fato de não tratar o problema de classificação de

dados como uma abordagem geométrica do espaço amostral. Esta é a primeira vez que um estudo relacionado à estimativa de precipitação utilizando OPF é apresentado na comunidade científica.

Para a realização do experimento foram utilizadas imagens da área localizada dentro do raio de cobertura quantitativa, de 240 km, do Radar Meteorológico do IPMet/UNESP (Instituto de Pesquisas Meteorológicas da Universidade Estadual Paulista Julio de Mesquita Filho), instalado no município de Bauru, região central do Estado de São Paulo, coordenadas 22°21'30'' de Latitude Sul e 49°01'38'' de Longitude Oeste. Os dados utilizados foram do tipo CAPPI (*Constant Altitude Plan Position Indicator*). Estes dados de radar fornecem ao estudo a verdade terrestre com a resolução espacial de 1 km.

Dados do canal infravermelho termal (10,2 – 11,2 μm) do satélite meteorológico GOES – 12 no período de dezembro/2003 a fevereiro/2004 foram utilizados em horários tão próximos quanto possíveis dos horários dos dados obtidos de radar, com no máximo 5 minutos de defasagem e com resolução espacial de 4 km.

Foram utilizadas 135 imagens do canal termal do satélite GOES as quais foram calibradas com suas 135 correspondentes no radar. Como as imagens foram obtidas com resoluções espaciais diferentes, os dados do satélite e do radar foram redimensionados, utilizando a técnica de interpolação vizinho mais próximo, para uma grade com tamanho $0.1^\circ \times 0.1^\circ$. A grade analisada neste estudo tem dimensão 20×20 *pixels*, ou seja, 54.000 *pixels* do canal infravermelho foram analisados dentro da área de cobertura de 240 km do radar de Bauru, com a região central da grade coincidindo com as coordenadas do radar.

Um dos fatores determinantes para a busca de *pixels* que representam nuvens de precipitação na imagem do canal infravermelho é a temperatura dos mesmos. Precipitações intensas estão sempre associadas a topos frios de nuvens, mas a recíproca nem sempre é verdadeira (ALDER & NEGRI, 1988).

SCOFIELD (2003) fez uma adaptação da técnica NAW (*Negri Alder Wetzel Technique*) desenvolvida por NEGRI *et. al.* (1984) para a região de cobertura do radar meteorológico do IPMet/UNESP, encontrando um limiar de temperatura de brilho de 235K para condição de precipitação. Esta foi a temperatura utilizada nos experimentos para delimitar se um *pixel* pode representar precipitação ou não. *Pixels* com temperatura de brilho maior que 235K foram classificados como *pixels* de não precipitação e excluídos de análise.

Uma outra dificuldade inerente à estimativa de precipitação aparece quando são utilizadas informações apenas do canal infravermelho. Nuvens cirrus podem levar à superestimação da área de precipitação, sendo identificadas como precipitáveis quando na verdade não são por apresentar temperaturas mais frias, aparecendo facilmente nas bandas do infravermelho (RAO *et al.*, 1990).

Um método desenvolvido por ADLER & NEGRI (1988) foi utilizado para identificar este tipo de nuvem. Neste método um parâmetro *Slope* é calculado para cada *pixel*. O parâmetro é definido como

$$S = T_m - T,$$

onde T_m é a média das temperaturas dos oito *pixels* ao redor do *pixel* analisado T . O parâmetro calculado foi comparado à equação linear desenvolvida por PANOFSKY & BRIER (1968)

$$\text{Slope} = 0.568(T - 217).$$

Esta equação linear delimita se a temperatura calculada é de uma nuvem cirrus ou uma nuvem de precipitação. Valores de S menores que o *Slope* calculado foram classificados como *pixels* de nuvens cirrus, já valores de S maiores ou iguais ao *Slope* foram classificados como *pixels* de nuvens de precipitação.

Os modelos descritos acima foram utilizados como entrada para os algoritmos de classificação avaliados: ANN, SVM e OPF. Os testes foram realizados de acordo com o seguinte procedimento: os dados obtidos das imagens foram divididos em dois subconjuntos: treinamento e teste, com a proporção de 50% cada. Os conjuntos de dados gerados foram os mesmos para as três metodologias em questão. Com o intuito de minimizar o efeito da instabilidade ocasionada pelas redes neurais, os

experimentos foram repetidos 10 vezes, com diferentes conjuntos de treinamento e avaliação, sendo calculados a média e o desvio padrão da acurácia destes classificadores no conjunto de teste. O tempo de execução médio também foi computado.

RESULTADOS E DISCUSSÃO: A Tabela 1 mostra os valores médios de acurácia e desvio padrão nos conjuntos de teste após 10 execuções dos algoritmos de classificação. Também são apresentados os tempos médios de execução de cada abordagem em segundos.

ALGORITMO	ANN	SVM	OPF
Acurácia média	0.663	0.806	0.865
Desvio padrão	0.163	0.019	0.128
Tempo de execução (s)	1756.07	176.587	108.170

Tabela 1: Valores médios obtidos da aplicação dos algoritmos

Nota-se a superioridade de desempenho do OPF em relação aos demais algoritmos. A abordagem OPF foi cerca de 16 vezes mais rápida que uma ANN, com uma taxa de acerto na classificação de uma região como sendo chuvosa ou não de aproximadamente 86%. O algoritmo SVM obteve um desempenho mais próximo, porém foi mais lento que o OPF.

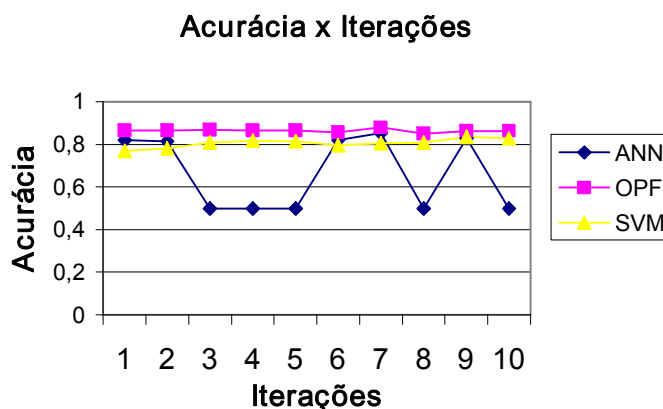


Figura 1: Desempenho dos classificadores ao longo das 10 iterações

Uma acurácia de 86% para a estimativa de precipitação pode ser considerada como satisfatória, tendo em vista a quantidade de erros oriundos do tratamento das imagens, como a perda de informação no processo de interpolação dos dados. MCCULLAGH *et al.* (1995) foi um dos trabalhos pioneiros a utilizar algoritmos de classificação para estimação de chuvas, conseguindo uma taxa de acerto de até 77.4% utilizando redes neurais artificiais para estimar precipitação acumulada em 6 horas, porém utilizando duas bandas espectrais do satélite GOES, sendo estas os canais infravermelho e visível. No seu trabalho, utilizou apenas 1000 amostras, enquanto no presente estudo foram utilizadas 54.000 amostras. A Figura 1 mostra os valores das acurácias nos diferentes conjuntos de teste ao longo das 10 iterações para os três algoritmos. Podemos perceber o comportamento instável das redes neurais artificiais.

CONCLUSÕES: Devido a grande importância da predição de precipitação em ambientes de culturas agrícolas, o presente trabalho propôs a comparação de três algoritmos para classificação e estimação de áreas de precipitação. Foram utilizadas nos experimentos 135 imagens do satélite GOES-12, calibradas com suas respectivas imagens obtidas do radar do IPMet/UNESP-Bauru. Os modelos estudados foram utilizados como entrada para os algoritmos de classificação ANN, SVM e OPF. Os experimentos foram repetidos 10 vezes, sendo calculadas as taxas médias de acerto e

tempo de execução em segundos. Também foi apresentado o desvio padrão das taxas de acerto no conjunto de teste. O presente trabalho detectou a superioridade do algoritmo OPF nesta área de aplicação, tanto com relação ao tempo de execução bem como à taxa média de acerto. Trata-se de um estudo preliminar para selecionar um classificador para ser utilizado em futuros estudos para a estimativa de precipitação diária em áreas rurais. Como próximos trabalhos, pretendemos ampliar o estudo para a identificação de quantidade de precipitação, bem como utilizar mais bandas espectrais deste e do satélite meteorológico NOAA.

REFERÊNCIAS BIBLIOGRÁFICAS

ADLER, R. F.; NEGRI A. J. *A satellite infrared technique to estimate tropical convective stratiform rainfall*. J. of App. Meteor., v. 27, p. 30-51, 1988.

BOSER, B.E.; GUYON, I.M.; VAPNIK, V.N., *A training algorithm for optimal margin classifiers*. Proc. 5th Workshop on Computational Learning Theory, ACM Press, p. 144-152, Pittsburgh, Pennsylvania, United States, 1992.

HAYKIN, S. *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1st, 1994

LEE, S.; CHO, S.; WONG, P.M. *Rainfall prediction using artificial neural networks*. J. Geogr. Inf. Dec. Anal. v.2, p. 253 – 264, 1998.

MCCULLAGH, J.; BLUFF, K.; EBERT, E. *A Neural Network Model for Rainfall Estimation*. II New Zealand Two-Stream International Conference on Artificial Neural Networks and Expert Systems, New Zealand, 1995.

NEGRI, A.J.; ADLER, R.F.; WETZEL, P.J. *Rainfall estimation from satellites: an examination of the Griffith-Woodley technique*. J. Clim. Appl. Meteor., v.23, p.102 -116, 1984.

PALMEIRA, F. L. B.; MORALES, C. A.; FRANÇA, G. B.; LANDAU, L. *Rainfall Estimation Using Satellite Data For Paraíba do Sul Basin – Brazil*. XXth International Society for Photogrammetry and Remote Sensing, p. 1-8, Istambul, 2004.

PANOFSKY, H.A.; BRIER, G.W. *Some Applications of Statistics to Meteorology*, The University of Pennsylvania, University of Park, PA, 224 p., 1968.

PAPA, J. P.; FALCÃO, ALEXANDRE X.; MIRANDA, P. A. V. ; SUZUKI, C. T. N.; MASCARENHAS, N. D. A. *A New Pattern Classifier based on Optimum Path Forest*. Institute of Computing, State University of Campinas, IC-07-13, jun., 2007.

RAO, P. K.; HOLMES S. J.; ANDERSON R. K.; WINSTON J. S.; LEHR P. E. *Weather Satellites: Systems, data, and environmental applications*. Am. Meteor. Soc. Boston, 1990.

SCOFIELD, G. B.; VENDRAME, I.; *Adaptação da técnica de estimativa de precipitação NAW para a área de abrangência do radar de Bauru, SP*. Anais XI SBSR, Belo Horizonte, INPE, p. 1203-1210, 2003.

UMEHARA, S.; YAMAZAKI, T.; SUGAI, Y. *A Precipitation Estimation System Based on Support Vector Machine and Neural Network*. Electronics and Communications in Japan, part 3: Fundamental Electronic Science, v. 89; n. 3, p. 38-47, 2006.