Análise de metacaracterísticas para classificação de uso e cobertura do solo utilizando Random Forest

Roberto U. Paiva[1], **Savio S. Oliveira**[2], **Wellington S. Martins** [2], **Leandro L. Parente**[1]

[1]Laboratório de Processamento de Imagens e Geoprocessamento (Lapig)
Universidade Federal de Goiás (UFG)
Goiânia, GO

[2]Instituto de Informática - Universidade Federal de Goiás (UFG)
Goiânia, GO.

urzedabr@ufg.br, wellington@inf.ufg.br, leal.parente@gmail.com, savioteles@gmail.com

Abstract. *This work analyzes the impact of the use of meta-features generated by the TWDTW algorithm for mapping land use and cover using Random Forest. The tests were carried out by classifying nine classes, for a set of samples from Mato Grosso region, in Brazil. Our results show that the meta-features are promising for the improvement of accuracy, increasing the overall accuracy of the tested models. The most significant improvements occur in the producer accuracy of the classes more difficult to classify. The importance of meta-features in the classification was significantly greater than features extracted from the EVI and NDVI indices and NIR and MIR bands.*

Resumo. *Este trabalho analisa o impacto do uso de metacaracterísticas geradas pelo algoritmo TWDTW para mapeamento do uso e cobertura do solo utilizando Random Forest. Os testes foram realizados classificando nove classes, para um conjunto de amostras da região do Mato Grosso, no Brasil. Nossos resultados mostram que as metacaracterísticas são promissoras para a melhora de acurácia, aumentando a acurácia global dos modelos testados. As melhoras mais significativas ocorrem na acurácia do produtor das classes de maior dificuldade de classificação. A importância das metacaracterísticas na classificação foi significativamente maior do que as características extraídas dos Índices EVI e NDVI e Bandas NIR e MIR.*

1. Introdução

O uso e cobertura do solo estão em constante mudança, tornando o mapeamento desse fenômeno uma das aplicações mais comuns e importantes no sensoriamento remoto [Foody 2002]. Compreender a dinâmica de alterações no espaço e tempo dessas mudanças auxilia na criação de políticas públicas, produtividade e redução na emissão de gases de efeito estufa [Nepstad et al. 2014, Bonan 2008, Bala et al. 2007].

O uso de algoritmos de aprendizado de máquina para a classificação do uso e cobertura do solo é essencial para a criação de modelos que permitam um mapeamento automatizado e recorrente. Nos últimos anos, os algoritmos Random Forest [Breiman 2001] e Support Vector Machines (SVM) [Vapnik 1995] se tornaram referência de bons classificadores na área de sensoriamento remoto [Rodriguez-Galiano et al. 2012, Belgiu and Drăguţ 2016, Mountrakis et al. 2011].

Outros tipos de algoritmos também são utilizados para classificação de uso e cobertura do solo, como o Time-Weighted Dynamic Time Warping (TWDTW) [Maus et al. 2016], que é um método sensível a mudanças sazonais dos diversos tipos de vegetações naturais e cultivadas. O TWDTW trabalha com a geração de medidas de similaridade entre séries temporais e a utiliza para a classificação [Belgiu and Csillik 2018]. Porém, alguns estudos mostram que o algoritmo apresenta baixa acurácia quando utilizado em regiões com alta variabilidade dos dados dentro de cada classe [Dadi 2019].

[Oliveira et al. 2018] explora paralelismo para acelerar o algoritmo TWDTW e viabilizar uma nova abordagem para o uso das medidas de similaridade. Nesse cenário as medidas de similaridade são tratadas como metacaracterísticas e utilizadas como parâmetros de entrada para os algoritmos KNN e SVM. A abordagem estende a metodologia de classificação que utiliza como características as observações anuais do produto MODISQ13 das bandas NIR e MIR e dos índices NDVI e EVI, adicionando as metacaracterísticas geradas pelo TWDTW [Oliveira et al. 2019]. Os resultados mostraram que as metacaracterísticas foram capazes de aumentar a acurácia global em ambos os algoritmos.

Neste trabalho analisamos o impacto das metacaracterísticas baseadas em similaridade geradas pela versão paralela do TWDTW na classificação de uso e cobertura do solo com nove classes distintas na região do Mato Grosso - Brasil, com o classificador Random Forest. O Random Forest é um dos algoritmos mais utilizados para classificação do uso e cobertura do solo, pois possui uma configuração simples e capacidade de obter boa acurácia [Pal 2005].

Os resultados mostraram que, uma vez que se tenha acesso a essas metacaracterísticas para o conjunto de dados, elas podem melhorar a acurácia de classificação.O conjunto de meta-características geradas pelo TWDTW foi capaz de melhorar a acurácia no modelo de classificação apresentado por [Picoli et al. 2018], assim como também em modelos com espaços de características menores. Essas metacaracterísticas podem ser utilizadas em diversos modelos de classificação de uso e cobertura do solo que trabalham com algoritmos de aprendizado de máquina.

O artigo está organizado da seguinte forma. A seção 2 apresenta um resumo de classificação de uso e cobertura do solo em sensoriamento utilizando o algoritmo Random Forest. Na seção 3 iremos tratar dos dados e métodos utilizados. A seção 4 apresenta experimentos e resultados que foram obtidos durante essa a pesquisa. Por fim, na seção 5 apresentamos as conclusões e trabalhos futuros.

2. Classificação do uso e cobertura do solo utilizando Random Forest

O Random Forest é um dos algoritmos de aprendizado de máquina mais utilizados para classificação de uso e cobertura do solo [Pal 2005, Belgiu and Drăguţ 2016, Gislason et al. 2006]. Alguns trabalhos atuais utilizam o Random Forest para o mapeamento de grandes áreas [Parente and Ferreira 2018, Tsai et al. 2018, Parente et al. 2019, Ayala-Izurieta et al. 2017].

De acordo com [Pal 2005] essa popularidade do Random Forest ocorre devido a sua capacidade de atingir acurácia semelhante ao SVM, em conjunto com a facilidade de utilização, com poucos parâmetros a serem configurados pelo usuário e adaptação aos dados de sensoriamento remoto.

A composição do espaço de características para o algoritmo Random Forest pode ser crucial para o desempenho da classificação [Canovas-Garcia and Alonso-Sarria 2015, Laliberte et al. 2012]. Diversos trabalhos buscam otimizar o número de características usadas para classificação do uso e cobertura do solo com o Random Forest utilizando métodos de seleção de características [Duro et al. 2012, Ma et al. 2017]. Nenhum trabalho anterior avaliou as metacaracterísticas geradas pelo TWDTW apresentadas por [Oliveira et al. 2019] utilizando o algortimo Random Forest.

Um espaço de características otimizado pode reduzir significativamente o tempo de processamento e o espaço de armazenamento e, ao mesmo tempo, produzir maior acurácia de classificação do que o conjunto de dados inicial [Georganos et al. 2018]. Nosso trabalho apresenta experimentos com redução no espaço de características utilizadas como entrada para o algortimo Random Forest, mantendo uma acurácia global acima de 90% em todos os casos.

No geral, os modelos de classificação com Random Forest buscam utilizar um espaço de características que sejam representativas para a classificação de determinado tipo de classe de uso e cobertura do solo. As características são geradas a partir das bandas espectrais e temporais do produto e em seguida utilizadas como dados de entrada para o algoritmo. O resultado é um classificador capaz de identificar uma ou várias classes com um certo nível de acurácia, de acordo com o modelo. Essa classificação permite gerar mapas de uso e cobertura do solo que podem ser utilizados como novas referências para classificação.

3. Dados e métodos

Os experimentos foram realizados utilizando o modelo e dados apresentados por [Picoli et al. 2018] para a classificação com SVM. São dados MODISQ13, cada amostra possui 23 observações anuais das bandas *Near-infrared* (nir), *Mid-infrared* (mir) e dos índices *Normalized Difference Vegetation Index* (NDVI), *Enhanced Vegetation Index* (EVI). Os dados foram extraídos da região do Mato Grosso - Brasil, com o objetivo de classificar nove classes diferentes de uso e cobertura do solo. As amostras são distribuiodas da seguinte maneira: Cerrado (400), Algodão-pousio (34), Floresta (138), Pastagem (370), Soja-milho (398), Soja-algodão (399), Soja-pousio (88), Soja-milheto (235) e Soja-girasol (53), totalizando 2115 amostras.

Foram geradas medidas de similaridade com o TWDTW para cada ponto dentro das amostras utilizando uma versão paralela do algoritmo proposta em [Oliveira et al. 2018]. As medidas de similaridade apresentam o valor de distâncias para cada classe, sendo no geral nove valores para cada ponto.

A nossa metodologia avaliou três abordagens de espaço de características como entrada para o Random Forest:

1. Bandas/Índices - Espaço de características com as 23 observações anuais das bandas NIR e MIR e dos índices NDVI e EVI [Picoli et al. 2018]. Essa abordagem possui um espaço de características de tamanho 92.
2. Bandas/Índices + metacaracterísticas TWDTW - Nessa abordagem a ideia é testar as metacaracterísticas no modelo de classificação anterior.
3. Bandas/Índices Individuais + metacaracterísticas TWDTW - Esse experimento visa analisar o impacto das metacaracterísticas em um espaço de características

menor, testando cada banda e índice individualmente e em conjunto com as metacaracterísticas TWDTW.

A Tabela 1 apresenta um exemplo resumido para três pixels e duas observações para entrada do modelo Bandas/Índices + metacaracterísticas TWDTW.

Pixel	ndvi1	ndvi2	evi1	evi2	nir1	nir2	mir1	mir2	twdtw Cerrado	twdtw Algodão-pousio	twdtw Floresta
1	0,39	0,49	0,25	0,28	0,32	0,27	0,31	0,17	3,53	4,03	4,00
2	0,50	0,49	0,26	0,33	0,23	0,36	0,14	0,16	2,62	4,27	2,91
3	0,35	0,34	0,19	0,16	0,23	0,20	0,20	0,15	1,70	4,08	3,34

Tabela 1. Exemplo de entrada para o Random Forest no modelo Metacaracterísticas TWDTW + Bandas/Índices.

Todos os experimentos foram realizados em um computador do tipo estação de trabalho, com as seguintes configurações: Processador Intel Core i7-9700, 16GB de memória RAM e 240GB de armazenamento em SSD M2 Pcie Nvme. O Random Forest foi treinado com 500 árvores, pois não observamos melhora dos resultados com o aumento do número de árvores. Com a finalidade de estimar a acurácia, todos os experimentos foram realizados utilizando a técnica de validação cruzada K-Fold, com K = 5 [Wiens et al. 2008].

Para facilitar a análise geramos uma matriz de confusão para todos os 3 testes, com as acurácias do produtor, do usuário e global. Foi utilizada a função de importância de característica do Random Forest para analisar o impacto destas durante o treinamento do classificador com todas as 101 caracterśiticas. A Figura 1 apresenta um fluxograma da metodologia.

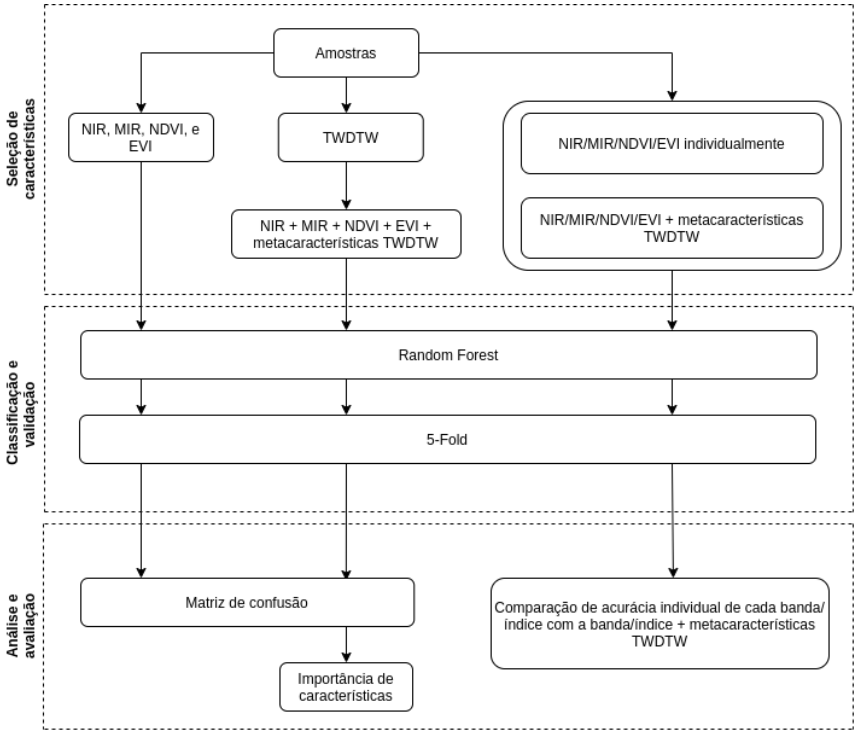

Figura 1. Fluxograma da metodologia utilizada no trabalho

4. Resultados

As tabelas de confusão para cada teste apresentaram bons índices de acurácia global. A Tabela 2 mostra os resultados do experimento utilizando a estratégia de Bandas e Índices. O classificador Random Forest treinado com a estratégia Bandas (NIR, MIR) e Índices (EVI, NDVI), com um total de 92 características por pixel de entrada, obteve uma acurácia global de 92,48%. Ao adicionar as 9 metacaracterísticas por pixel geradas pelo TWDW, a acurácia Global (AG) subiu para 93,05% (Tabela 3). Também foram calculadas a acurácia do produto (AP) e acurácia do usuário (AU).

	1	2	3	4	5	6	7	8	9	**AU**
1 - Cerrado	398	0	1	8	1	0	0	0	0	97,50%
2 - Algodão-pousio	0	25	0	0	0	2	0	0	0	94,12%
3 - Floresta	2	0	136	1	0	0	0	0	0	97,83%
4 - Pastagem	0	3	1	359	5	1	0	9	0	94,86%
5 - Soja-milho	0	0	0	0	360	26	0	27	16	82,66%
6 - Soja-algodão	0	6	0	1	8	366	0	3	0	95,49%
7 - Soja-pousio	0	0	0	0	0	0	87	1	0	98,86%
8 - Soja-milheto	0	0	0	1	23	4	1	195	7	84,68%
9 - Soja-girassol	0	0	0	0	1	0	0	0	30	98,11%
AP	99,50%	73,53%	98,55%	97,03%	90,45%	91,73%	98,86%	82,98%	56,60%	**AG = 92,48%**

Tabela 2. Matriz de confusão dos testes com a metodologia Bandas/Índices.

Podemos observar na Tabela 3 que o classificador conseguiu boa acurácia geral (maior que 90% de acurácia do produtor) para as classes Cerrado, Floresta, Pastagem, Soja-milho, Soja-algodão e Soja-pousio. Houve confusão nas classes Soja-Girassol com Soja-Milho, devido ao comportamento similar das séries temporais dessas classes. O classificador não foi capaz diferenciá-las com uma boa acurácia, fazendo com que muitas amostras de Soja-girassol fossem classificadas como Soja-Milho. O mesmo acontece entre as classes Soja-Milho e Soja Milheto. A semelhança do comportamento das séries temporais pode ser observada na Figura 2.

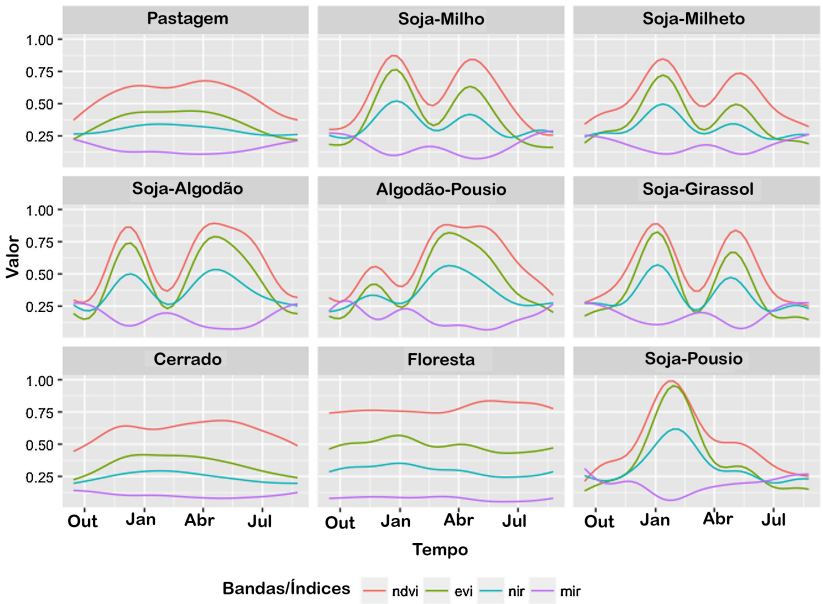

Figura 2. Comportamento das séries temporais MODIS das classes analisadas [Picoli et al. 2018]

A melhor acurácia global foi atingida utilizando a combinação de Bandas e Índices com as metacaracterísticas de distância geradas pelo TWDTW, como pode ser observado na Tabela 3. Nossos experimentos conseguiram uma acurácia global de 93,05%, utilizando todas as 101 características por amostra.

	1	2	3	4	5	6	7	8	9	**AU**
1 - Cerrado	396	0	1	6	0	0	0	0	0	98,25%
2 - Algodão-pousio	0	28	0	0	0	1	0	0	0	97,06%
3 - Floresta	2	0	136	1	0	0	0	0	0	97,83%
4 - Pastagem	2	1	1	361	5	1	0	6	0	95,68%
5 - Soja-milho	0	0	0	0	360	23	0	26	14	84,17%
6 - Soja-algodão	0	4	0	1	8	369	0	2	0	96,24%
7 - Soja-pousio	0	0	0	1	0	1	86	1	0	96,59%
8 - Soja-milheto	0	1	0	0	24	4	2	200	7	83,83%
9 - Soja-girassol	0	0	0	0	1	0	0	0	32	98,11%
AP	99,00%	82,35%	98,55%	97,57%	90,45%	92,48%	97,73%	85,11%	60,38%	**AG = 93,05%**

Tabela 3. Matriz de confusão dos testes com a metodologia Bandas/Índices + metacaracterísticas TWDTW.

A contribuição das metacaracterísticas geradas pelo TWDTW foi mais significativa nas classes difíceis (i.e., classes que, por possuírem comportamento semelhantes nas séries temporais, acabam confundindo o classificador). Traçando um comparativo com o experimento anterior, tivemos um aumento da acurácia do produtor de 8,82% na classe Algodão em pousio, 2,13% em Soja-Milheto e 3,78% em Soja-Girassol.

Realizando a análise de importância de características durante a classificação utilizando as Bandas e Índices acrescidos das metacaracterísticas geradas pelo TWDTW, identificamos que o importância dessas 9 metacaracterísticas por pixel é significativamente maior do que as características das Bandas e Índices. A Figura 3 apresenta o gráfico com a relação de importância de características.

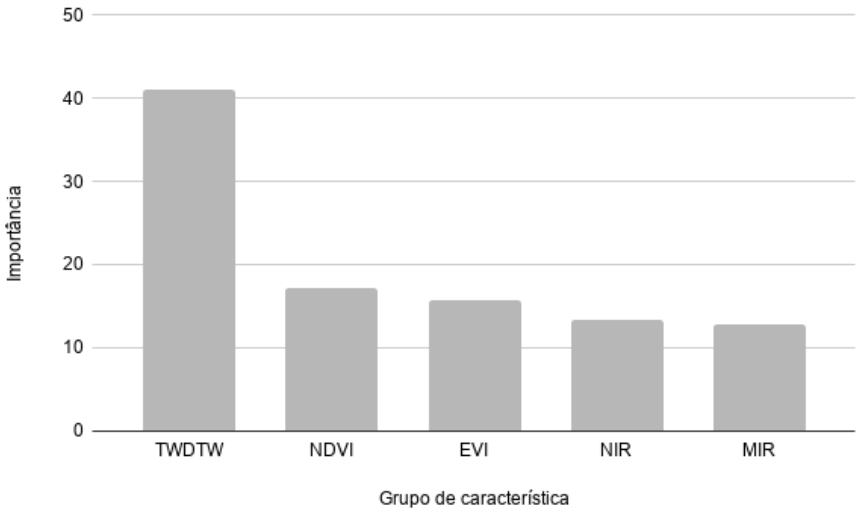

Figura 3. Importância de cada grupo de características durante a classificação.

Apesar de serem o menor grupo de características (i.e., possuir 9 metacaracterísticas enquanto as bandas e índices possuem 23 características cada grupo), o gráfico mostra que as metacaracterísticas estão em uma escala bem maior de importância.

Devido a esta alta importância das metacaracterísticas, foram realizados experimentos com espaços de características menores, utilizando individualmente cada Banda (NIR e MIR) e Índice (NDVI e EVI) e em seguida repetindo o processo acrescentando a cada modelo individual metacaracterísticas TWDTW.

Essa estratégia utiliza espaços de características compostos por 23 observações anuais de cada Banda ou Índices, e permite a avaliação das metacaracterísticas em modelos onde a acurácia global não está muito alta. Esses experimentos permitem visualizar melhor o impacto das metacaracterísticas TWDTW na acurácia global da classificação.

Na Tabela 4 observamos que o impacto positivo das metacaracterísticas na acurácia global é maior em um espaço de característica menor.

	Acurácia Global Individual	**Acurácia Global Banda/Índice + TWDTW**
NIR	86,80%	90,35%
MIR	83,26%	89,17%
EVI	87,47%	90,30%
NDVI	86,43%	90,40%

Tabela 4. Tabela de comparação de Acurácia Global da metodologia Bandas/Índices + metacaracterísticas TWDTW.

Tendo em vista que o espaço de características foi significativamente reduzido (i.e., usando apenas as 23 observações de uma Banda ou Índice, somados a 9 metacaracterísticas geradas pelo TWDTW), foi possivel obervar melhora de até 3,97% na acurácia global (modelo NDVI + metacaracterísticas TWDTW). Esse resultado mostra que as metacaracterísticas geradas pelo TWDTW podem ser aplicadas em qualquer modelo de classificação que utiliza o Random Forest, podendo obter bons ganhos de acurácia. Mesmo com a redução do espaço de características, 3 dos 4 modelos avaliados conseguiram manter a acurácia global acima de 90% em uma base de dados com 9 classes.

A redução do espaço de características, mantendo bons padrões de acurácia, tem potencial para contribuir com a otimização do tempo de execução de modelos de classificação do uso e cobertura do solo. A Tabela 5 apresenta o tempo de execução de cada experimento de acordo com a quantidade de características. Nesta tabela não foi considerado o pré-processamento para geração das metacaracterísticas com o TWDTW, levando em consideração que estas metacaracterísticas só precisam ser geradas uma vez para o grupo de amostras, podendo em seguida serem utilizadas em diversas análises.

	Acurácia Global	**Qte de características**	**Tempo de execução**
1 Bandas/Índices	92,48%	92	22,0s
2 Bandas/Índices+TWDTW	93,05%	101	23,7s
3 NDVI+TWDTW	90,40%	32	14,2s

Tabela 5. Tempo de execução de acordo com o espaço de características

A acurácia global atingida nos testes com o Random Forest nesse conjunto de dados foi ligeiramente menor do que a acurácia atingida nos testes com o algoritmo SVM (93,05% com o Random Forest e 93,80% com o SVM) em [Oliveira et al. 2018]. Entretanto, este já é um resultado esperado visto que o SVM tem uma configuração complexa

que precisa ser ajustada para toda pequena área a ser analisada. Para grandes áreas, estes ajustes de parâmetros do SVM são tão trabalhosos que faz com que o Random Forest seja o método de escolha devido à sua simplicidade na parametrização e boa acurácia. Assim, esta pesquisa demonstra que as metacaracterísticas podem também melhorar a acurácia do Random Forest, contribuindo principalmente na classificação de uso de solo relacionados a agricultura, que possuem maiores semelhanças entre as séries temporais, devido ao padrão sazonal de cultivo.

O TWDTW possui uma complexidade algorítmica quadrática e seu código foi projeto para trabalhar de forma sequencial. Devido a esses fatores, a demanda computacional para se processar as séries temporais com ele é muito elevada. Diante disto, o pré-processamento para geração das metacaracterísticas pode impactar de forma negativa no tempo de execução das análises completas. Isto pode inviabilizar análises em grandes áreas com um volume maior de dados a serem processados devido a alta demanda de recursos computacionais necessários para o processamento. Porém, uma vez que as metacaracterísticas sejam geradas para uma imagem, elas não precisam ser processadas novamente, sendo possível utilizá-las em diversos modelos e pesquisas.

5. Conclusões e trabalhos futuros

Um conjunto de características que representem bem os elementos a serem classificados é crucial para uma boa acurácia do algoritmo Random Forest [Belgiu and Drăguţ 2016]. As metacaracterísticas são geradas a partir das séries temporais das bandas e índices, se tornando um compilado de todas as informações. Por esse motivo, essas poucas características de similaridade possuem grande representação das classes, causando um impacto positivo na acurácia da classificação.

As metacaracterísticas geradas pelo TWDTW apresentaram melhora na acurácia global em todos experimentos nos modelos de classificação com o Random Forest em que foram acrescentadas. Uma vez que se tenha acesso a essas características, elas podem ser utilizadas em qualquer modelo de classificação. Por outro lado, o TWDTW possui alto custo computacional, o que pode dificultar a obtenção das características para grandes áreas.

Em trabalhos futuros pretendemos otimizar o algortimo paralelo do TWDTW apresentado por [Oliveira et al. 2018], possibilitando a geração de metacaracterísticas para grandes áreas. Essas metacaracterísticas serão testadas em outros modelos de classificação, com o objetivo de melhorar a acurácia na geração de mapas de uso e cobertura do solo.

Referências

Ayala-Izurieta, J. E., Márquez, C. O., García, V. J., Recalde-Moreno, C. G., Rodríguez-Llerena, M. V., and Damián-Carrión, D. A. (2017). Land cover classification in an ecuadorian mountain geosystem using a random forest classifier, spectral vegetation indices, and ancillary geographic data. *Geosciences*, 7(2):34.

Bala, G., Caldeira, K., Wickett, M., Phillips, T., Lobell, D., Delire, C., and Mirin, A. (2007). Combined climate and carbon-cycle effects of large-scale deforestation. *Proceedings of the National Academy of Sciences*, 104(16):6550–6555.

Belgiu, M. and Csillik, O. (2018). Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote sensing of environment*, 204:509–523.

Belgiu, M. and Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114:24–31.

Bonan, G. B. (2008). Forests and climate change: forcings, feedbacks, and the climate benefits of forests. *science*, 320(5882):1444–1449.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Canovas-Garcia, F. and Alonso-Sarria, F. (2015). Optimal combination of classification algorithms and feature ranking methods for object-based classification of submeter resolution z/i-imaging dmc imagery. *Remote Sensing*, 7(4):4651–4677.

Dadi, M. M. (2019). Assessing the transferability of random forest and time-weighted dynamic time warping for agriculture mapping. Master's thesis, University of Twente, Enschede.

Duro, D. C., Franklin, S. E., and Dubé, M. G. (2012). Multi-scale object-based image analysis and feature selection of multi-sensor earth observation imagery using random forests. *International Journal of Remote Sensing*, 33(14):4502–4526.

Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80(1):185–201.

Georganos, S., Grippa, T., Vanhuysse, S., Lennert, M., Shimoni, M., Kalogirou, S., and Wolff, E. (2018). Less is more: Optimizing classification performance through feature selection in a very-high-resolution remote sensing object-based urban application. *GIScience & remote sensing*, 55(2):221–242.

Gislason, P. O., Benediktsson, J. A., and Sveinsson, J. R. (2006). Random forests for land cover classification. *Pattern Recognition Letters*, 27(4):294–300.

Laliberte, A. S., Browning, D., and Rango, A. (2012). A comparison of three feature selection methods for object-based classification of sub-decimeter resolution ultracam-l imagery. *International Journal of Applied Earth Observation and Geoinformation*, 15:70–78.

Ma, L., Fu, T., Blaschke, T., Li, M., Tiede, D., Zhou, Z., Ma, X., and Chen, D. (2017). Evaluation of feature selection methods for object-based land cover mapping of unmanned aerial vehicle imagery using random forest and support vector machine classifiers. *ISPRS International Journal of Geo-Information*, 6(2):51.

Maus, V., Câmara, G., Cartaxo, R., Sanchez, A., Ramos, F. M., and De Queiroz, G. R. (2016). A time-weighted dynamic time warping method for land-use and land-cover mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(8):3729–3739.

Mountrakis, G., Im, J., and Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3):247–259.

Nepstad, D., McGrath, D., Stickler, C., Alencar, A., Azevedo, A., Swette, B., Bezerra, T., DiGiano, M., Shimada, J., da Motta, R. S., et al. (2014). Slowing amazon deforesta-

tion through public policy and interventions in beef and soy supply chains. *science*, 344(6188):1118–1123.

Oliveira, S. S., Cardoso, M. d. C., Bueno, E., Rodrigues, V. J., and Martins, W. S. (2019). Exploiting parallelism to generate meta-features for land use and land cover classification with remote sensing time series. *Brazilian Symposium on Geoinformatics (GeoInfo)*, pages 135–146.

Oliveira, S. S., Pascoal, L. M., Ferreira, L., Cardoso, M. d. C., Bueno, E., Rodrigues, V. J., and Martins, W. S. (2018). Sp-twdtw: A new parallel algorithm for spatio-temporal analysis of remote sensing images. *Brazilian Symposium on Geoinformatics (GeoInfo)*, pages 46–57.

Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26(1):217–222.

Parente, L. and Ferreira, L. (2018). Assessing the spatial and occupation dynamics of the brazilian pasturelands based on the automated classification of modis images from 2000 to 2016. *Remote Sensing*, 10(4):606.

Parente, L., Mesquita, V., Miziara, F., Baumann, L., and Ferreira, L. (2019). Assessing the pasturelands and livestock dynamics in brazil, from 1985 to 2017: A novel approach based on high spatial resolution imagery and google earth engine cloud computing. *Remote Sensing of Environment*, 232:111301.

Picoli, M. C. A., Camara, G., Sanches, I., Simões, R., Carvalho, A., Maciel, A., Coutinho, A., Esquerdo, J., Antunes, J., Begotti, R. A., et al. (2018). Big earth observation time series analysis for monitoring brazilian agriculture. *ISPRS journal of photogrammetry and remote sensing*, 145:328–339.

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., and Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67:93–104.

Tsai, Y. H., Stow, D., Chen, H. L., Lewison, R., An, L., and Shi, L. (2018). Mapping vegetation and land use types in fanjingshan national nature reserve using google earth engine. *Remote Sensing*, 10(6):927.

Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. Springer New York.

Wiens, T. S., Dale, B. C., Boyce, M. S., and Kershaw, G. P. (2008). Three way k-fold cross-validation of resource selection functions. *Ecological Modelling*, 212(3-4):244–255.