

TerraSAR-X データを用いたランダムフォレストによる作付作物の分類 Crop Classification by Random Forest Using TerraSAR-X Data

○山谷 祐貴*, 菌部 礼**, 谷 宏***, 王 秀峰***, 小林 伸行****, 望月 貫一郎*****

○Yuki Yamaya*, Rei Sonobe**, Hiroshi Tani***, Xiufeng Wang***,

Nobuyuki Kobayashi****, Kanichiro Mochizuki*****

1. 背景と目的

農業分野におけるリモートセンシングの利用は、従来の地上調査に代わる有望な統計情報収集手段の一つとなりうる。特にマイクロ波を用いた能動型センサである合成開口レーダ (SAR) は、光学センサと異なり天候の影響を受けにくく (Blaes *et al.*, 2005), 多時期のデータによる農業分野での利用に効果的であると考えられる。本研究では、Xバンドの SAR を搭載した TerraSAR-X が観測したデータを使用し、機械学習の分類アルゴリズムから各圃場における作付作物の分類を行い、その精度を算出した。分類を行うにあたり、SAR の主要な数値である後方散乱係数だけでなく、固有値解析法から算出される固有値を変数として使用した。また、このような衛星データの取得にはコストがかかるため、高い分類の精度を保持しつつ、いかに要素数や時期数を削減することが可能であるかについても併せて検討した。

2. 研究方法

(1) 対象地域および対象作物 研究対象地域は、北海道十勝管内の西部に位置する河西郡芽室町・上川郡清水町の畑作地帯とした。対象作物は、小麦、豆類、馬鈴薯、甜菜、牧草、トウモロコシの 6 種類とし、全 5,057 圃場を解析に使用した。

(2) 使用した衛星データ 本研究で使用したデータは、11 日間隔で取得された 2013 年 6 月 17・28 日、7 月 9・20・31 日の計 5 時期の TerraSAR-X 二重偏波データである。分類を実行するにあたり、これらのデータから各偏波の後方散乱係数 (HH, VV) に加えて、固有値解析法 (Cloude, 2007) を用い、固有値 (エントロピ H, アルファ角 α) を算出した。以上の 5 時期・4 要素からなる計 20 変数を分類に使用した。

(3) 解析方法 まず、算出した変数について、1 圃場ごとの平均値を抽出した。次に、20%の圃場を分類処理のための教師データとして無作為に抽出し、残りの 80%をテストデータとした。このテストデータについて変数から教師付き分類を行い、その精度を算出した。分類には機械学習アルゴリズムであるランダムフォレスト (Breiman, 2001) を使用した。

(4) 分類処理の流れ まず、5 時期 4 要素の全 20 変数を用いて分類を行い、その精度を算出した。次に、これらの要素数や時期数を削減した場合について、再度ランダムフォレストを適用した。それらの結果から、使用するデータ数を減少させたことによって、十分な精度を保持した分類を行うことが可能であるか検討した。

* 北海道大学大学院農学院, Graduate School of Agriculture, Hokkaido University

** 静岡大学農学部, Faculty of Agriculture, Shizuoka University

*** 北海道大学大学院農学研究院, Research Faculty of Agriculture, Hokkaido University

**** 株式会社スマートリンク北海道, Smart Link HOKKAIDO Inc.

***** 株式会社パスコ, PASCO Corporation

[キーワード] TerraSAR-X, 偏波, ランダムフォレスト, 作物分類

3. 結果と考察

(1) 20 変数すべてを使用した分類

20 変数すべてを使用した結果を表 1 に示す。すべての時期・要素を用いた場合、全体精度は 0.895、カッパ係数は 0.869 であり、ランダムフォレストを TerraSAR-X データに適用することによって、高い精度で分類を行うことが可能であった。特に小麦は高い精度で分類が可能であったが、トウモロコシは豆類などに誤分類される傾向があり、高い精度が得られなかった。この原因として、トウモロコシは豆類などと群落における散乱の形態が近かったことが考えられる。

(2) 使用する要素数の検討 本研究で用いた 4 要素のうち、いずれか 1 要素を用いた場合、後方散乱係数の 2 要素を用いた場合 (HH+VV)、固有値の 2 要素を用いた場合 (H+ α) について、それぞれ精度の保持は可能であるか検討した (表 2)。1 要素のみ用いた場合の全体精度は 0.730~0.778 と大幅に低く、2 要素を用いた場合についても全体精度は 0.830~0.841 程度であった。したがって、要素数を限定することは精度の低下につながり、高精度な分類を保持することが困難になると判断した。

(3) 使用する時期数の検討 本研究で使用した 5 時期のデータのうち、いずれか数時期のデータのみ使用した場合、精度の保持が可能であるか検討した (図 1)。5 時期中 2 時期を使用することでカッパ係数は 0.8 を超え、ある程度高い精度を保持していた。特に 5 時期中 4 時期を使用した場合の全体精度は 0.886 であり、5 時期すべてを用いた場合と同じ水準の分類精度が得られた。

4. 結論

本研究では、X バンドの二重偏波 SAR データにランダムフォレストを適用し、作付作物の分類を行った。その結果、5 時期 4 要素の計 20 変数を使用した場合、約 0.9 の高い精度で作付作物が分類された。使用する要素数を限定して分類を行ったところ、精度は大幅に下がり、高精度の保持は望めなかった。しかし、時期数を限定した場合、4 時期以上あれば 5 時期すべてを使用した際とほぼ変わらない精度で分類がなされ、より低コストでの作物分類が可能となると結論付けた。

引用文献

- Blaes, X., L. Vanhalle and P. Defourny (2005) *Efficiency of crop identification based on optical and SAR image time series. Remote Sensing of Environment*, .96(3~4), 352~65
- Breiman, L (2001), *Random forests. Machine Learning*, 45(1), pp. 5-32.
- Cloude, S. R. (2007), *The dual-polarization entropy/alpha decomposition: a pulsar case study*, Proc. POLInSAR, Frascati, Italy, pp. 1-6

表 1 20 変数すべてを利用した判別効率表
Table 1. Confusion matrix, using all 20 variables

分類結果	参照データ							合計	UA
	豆類	甜菜	牧草	トウモロコシ	馬鈴薯	小麦	PA		
豆類	753	15	2	41	46	6	863	0.873	
甜菜	11	442	1	12	4	0	470	0.940	
牧草	13	4	423	28	9	46	523	0.809	
トウモロコシ	29	26	15	375	8	13	466	0.805	
馬鈴薯	30	2	2	12	490	8	544	0.901	
小麦	1	1	25	11	7	1,147	1,192	0.962	
合計	837	490	468	479	564	1,220	4,058		
PA	0.900	0.902	0.904	0.783	0.869	0.940			
							全体精度	0.895	
							カッパ係数	0.869	

表 2 一部の要素を使用した分類精度
Table 2. Classification accuracy, using some parameters

使用した要素	分類精度	カッパ係数
HH	0.778	0.725
VV	0.770	0.715
H	0.730	0.665
α	0.766	0.710
HH+VV	0.830	0.789
H+ α	0.841	0.803
すべて	0.895	0.869

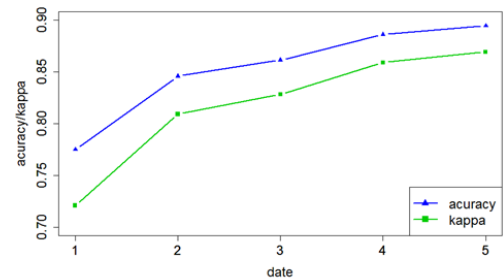


図 1 使用した時期数別の分類精度の変化
Figure 1. Change of accuracy using some scenes