Vol.2021-NL-248 No.35
Vol.2021-CVIM-226 No.35
2021/5/21

ブドウ果粒検出の評価用画像データセットの構築

大村 海雲[1,a)]　内海 ゆづ子[1,b)]　三輪 由佳[2,c)]　岩村 雅一[1,d)]　黄瀬 浩一[1,e)]

概要：日本のブドウ生産は年々減少している．この原因として，高齢化による栽培者の減少とブドウ栽培作業の数の多さ，困難さが挙げられる．そこで，我々は，ブドウの栽培支援を目的として，摘粒支援システムを開発している [1]．このシステムの評価のためには，画像上からの実の検出位置精度の評価が必要であるが，ブドウの実の画像とその位置を評価するためのデータセットは我々の知る限り存在していない．そのため，本研究では，摘粒支援システムの評価用データセットを作成する．このデータセットは，3210 枚のブドウ画像とブドウの実の中心位置情報から成り，ブドウの画像から実の中心位置を検出する手法を評価できる．データセットの難易度を評価するため，従来研究でブドウの実の検出に成功していた深層学習以前に提案されていた Roscher らの手法を適用し，ブドウの実の中心位置検出の精度評価を行った．実験の結果，一部のブドウの実のみ検出することができたが，多くのブドウの実の検出には失敗した．このことから，提案データベースのタスクは，深層学習以前の手法に対して難しいことが明らかとなった．

キーワード：物体検出，ブドウ栽培支援，データベース評価

1. はじめに

農林水産省の統計によると，日本のブドウ収穫量，出荷量，生産量は年々減少している [2]．ブドウ栽培の縮小の原因として，高齢化による栽培者の減少が挙げられる．また，ブドウ栽培は他の果樹と比較して作業数が多く，1 つ 1 つの作業が困難である．そのため，日本のブドウ栽培を拡大するにはブドウ生産者の負担を減らすことが重要である．

本稿ではブドウ栽培作業の中でも，摘粒と呼ばれる作業に注目する．摘粒とは，ブドウの房の整形と実の数の調整のために房から不要な実を間引く作業である．この摘粒作業はブドウ栽培の中でも特に困難な作業であるとされている．なぜなら，摘粒には図 1 のように明確な基準があるものの，ブドウの房の形状や実の付き方は様々であるため，実際に摘粒を行うには生産者の長年の経験が必要になるからである．そのため，摘粒作業にかかる手間を小さくすることができれば，ブドウ生産者の負担を大幅に減るので，日本のブドウ栽培の拡大が期待できる．

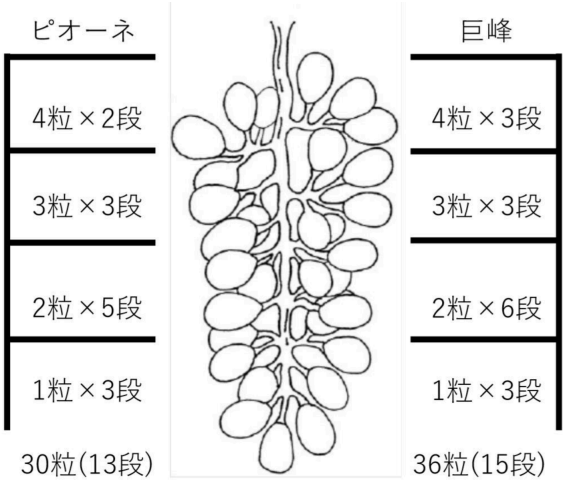

図 1: ブドウ摘粒作業の基準 [3]

我々の研究グループではブドウの栽培支援を目的としたシステムを開発している．摘粒作業を支援する方法として，カメラでブドウを撮影した後，自動で間引くべき実を選択し，生産者に提示するものが挙げられる．そこで，図 2 のようなシステムを提案した [1]．まず，ブドウ農場でカメラを用い，ブドウを撮影する．次に，インスタンスセグメンテーションにより撮影された動画像からブドウの実を検出し，ステレオマッチングによりブドウの実の 3 次元位置を推定し，ブドウの実の 3 次元的な位置関係を復元する．そして，得られた復元結果に対して摘粒基準に合わせて間

1 大阪府立大学 大学院工学研究科 知能情報工学分野
Graduate School of Engineering, Osaka Prefecture University
2 大阪府環境農林水産総合研究所
Reseach Institute of Environment, Agriculture and Fisheries, Osaka Prefecture
a) omura@m.cs.osakfu-u.ac.jp
b) yuzuko@cs.osakafu-u.ac.jp
c) MiwaY@mbox.kannousuiken-osaka.or.jp
d) masa@cs.osakafu-u.ac.jp
e) kise@cs.osakafu-u.ac.jp

© 2021 Information Processing Society of Japan

引く実を自動で選択し，ブドウ農家に提示するシステムである．

摘粒支援システムの完成には，評価用データセットを用いた検出精度の評価が必要である．我々の知る限り，ブドウの実検出の位置精度評価が可能なデータセットは存在しない．そこで，本研究では，ブドウの実の検出のためのデータセットを構築する．また，一般的に，データセットは広く知られた既存手法で評価実験が行われてから公開される．そこで本稿では，はじめの取組として，深層学習の出現よりも以前に提案されたブドウの実の検出手法である Roscher らの円構造検出器 [4] を用いてデータセットの難易度を評価した．実験の結果，Roscher らの手法では，ごく一部のブドウの実の検出には成功したが，大半のブドウの実の検出には失敗した．そのため，深層学習を用いない方法では画像からブドウの実を検出することは非常に困難であることが分かった．

2. 関連研究

本章では関連研究として画像セグメンテーションに関する研究，およびデータセットに関する研究について述べる．

2.1 画像セグメンテーション

画像セグメンテーションは，入力された画像に対してピクセルレベルで物体が占める領域を検出するタスクである．画像から自動で物体の位置を検出できるため，この技術は自動車の自動運転やシーンの理解への応用が期待されている [5–7]．近年，畳み込みニューラルネットワーク (Convolutional Neural Network) の発展とディープラーニングの導入に伴い，機械学習を用いた高レベルなセグメンテーション手法が多数提案されている．中でも，R-CNN [8] をはじめ，Fast R-CNN [9] や Faster R-CNN [10]，Mask R-CNN [11] は高い精度での領域検出を実現したことで様々な用途に利用されている．

画像セグメンテーションはモデルを学習させるために大量のデータを必要とする [12,13]．Mask R-CNN も 80,000 枚の画像データを用いて学習を行っている [11]．画像セグメンテーションが可能なのは，大規模なデータセットの存在のおかげであると言える [14]．しかし，データセットの作成には膨大な時間とコスト，人手が必要になる．そのため，新たな学習用データセットを入手することは非常に困難である．そして現在，ブドウを対象にしたデータセットは数が少なく，ブドウの実を対象にしたデータセットに至っては存在しない．そこで，我々はブドウの検出を評価するためのデータセットを提案する．

2.2 データセット

多くの研究に用いられるデータセットとして，COCO データセットが挙げられる [15]．COCO データセットは 328,000 枚の画像から成る大規模なデータセットである．その中には 91 種類のクラス，計 2,500,000 個の物体が含まれている．しかし，このデータセットにはブドウというクラスがないため，本研究では使用できない．

ブドウに関するデータセットに Santos らのデータセットがある [16]．Santos らはブドウ 4,432 房が写った 300 枚の画像を対象に房単位でのラベル付けを行ったデータセットを提案した．ブドウの種類は Chardonnay，Cabernet Franc，Cabernet Sauvignon，Sauvignon Blanc，Syrah の 5 種類ある．ブドウ画像は Canon EOS REBEL T3i DSLR と Motorola Z2 Play smartphone を用いて，ブドウから 1–2 m の距離から撮影された．Canon EOS REBEL T3i DSLR で撮影された画像のサイズは 2,048×1,365 画素で，Motorola Z2 Play smartphone の画像は 2,048×1,536 画素である．ラベル付けは房ごとにバウンディングボックスを付けることで行われた．また，4,432 房のブドウのうち 2,020 房が写った 137 枚の画像に対してはピクセルレベルでのラベル付けも行われた．Santos らのデータセットは房単位でラベル付けされたデータセットであるため，実を検出する本研究では利用することができない．

2.1 で述べたように，データセットの作成には非常に大きなコストがかかる．そのため近年では，動画像を取得すると同時にラベル付けを行う手法も提案されている．Stephan らはコンピュータゲームのプレイ画像に対して，グラフィックハードウェアから得られるオブジェクトの情報を基にピクセルレベルでラベル付けを行った [17]．コンピュータゲーム内ではオブジェクトの位置や名前を言った情報が既知であるため，自動で高速かつ高精度なラベル付けが可能である．また，モーションキャプチャや TOF カメラ，複数のカメラを使ったキャリブレーションにより自動で人間の骨格点情報を得る手法も存在する [18] [19–21]．自動でラベル付けを行うことにより，データセット作成にかかるコストが削減されるだけでなく，作成にかかる時間も減る．しかし，現在のところブドウ画像を自動でラベル付けする手法はなく，人間が手でラベル付けをする必要がある．

3. ブドウの実検出評価用データセット

ブドウに関するデータセットに注目して見ると，ブドウを対象にしたデータセットは数が少なく，ブドウの実を対象にしたデータセットに至っては存在しない．そのため，摘粒支援システムを開発しても，その評価ができない．そこで，我々は，画像中のブドウの実の中心座標を記録したブドウの実評価用データセットを作成する．データセット画像の例を図 3 に示す．図 3 中の赤色の点は記録された実の中心位置を表す．

提案データセットには 3,210 枚のブドウ画像が含まれており，これらの画像は大阪府立環境農林水産総合研究所の

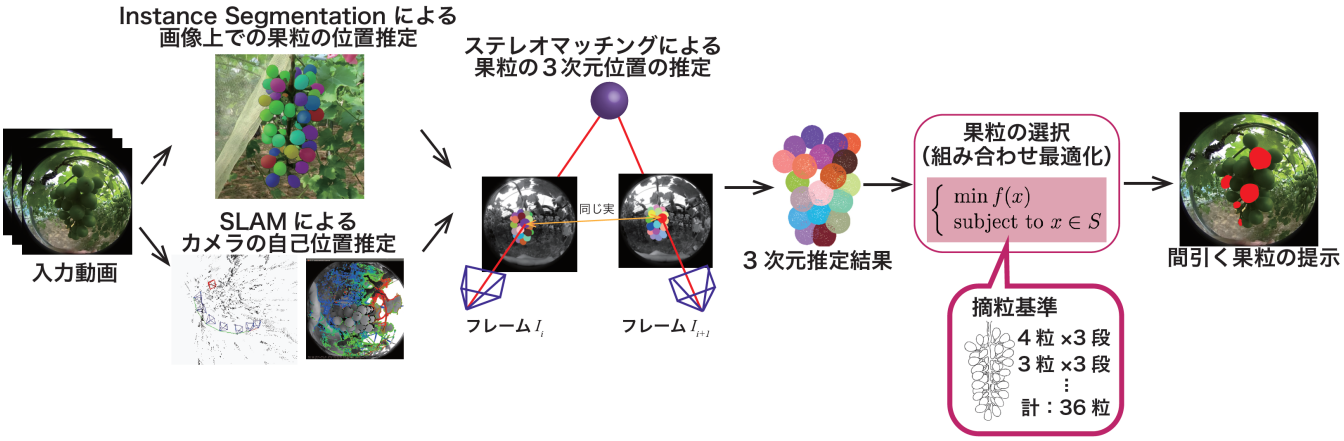

図 2: 摘粒支援システム概要 [1]

図 3: ブドウの実検出評価用データセットの画像例．赤点はブドウの実の中心位置を示す．

果樹ほ場で撮影された．ブドウの種類はシャインマスカット，サンヴェルデ，藤稔，巨峰の 4 種類である．ブドウ画像は Canon Power Shot G7 X Mark II を用いて，ブドウ 1 房につき房の周囲を角度を変えながら 3 枚撮影されている．撮影日は 2019 年 5 月 31 日と 6 月 13 日で， 5 月 31 日に撮影されたブドウは摘粒前のブドウであり， 6 月 13 日に撮影されたブドウは摘粒後のブドウである．ブドウの実の中心位置は株式会社 ABEJA に依頼して，ブドウの実の中心位置のポイントアノテーションをした．基準として，実が完全に見える場合は実の中心，一部が隠れている場合は見える範囲の実の中心の座標を記録した．また，完全に隠れている実は無視することとした．

4. Roscher らの手法 [4]

一般的に，データセットは広く知られた既存手法を用いて評価実験が行われる．実際に，新しく作成されたデータセットで学習とテストが行われ，どの程度の精度が実現できるか測定される．本研究では，データセットの評価の最初の試みとして，Roscher らの手法 [4] を用いて評価をする．

Roscher らは入力された画像内にあるブドウの実のサイズを推定する手法を提案した．Roscher らの手法の概要図を図 4 に示す．まず初めに，Roscher らの手法では画像から円構造を持つ物体が検出される．次に，検出された領域の特徴量をそれぞれ抽出することにより，その領域がブドウかブドウでないかの判定が行われる。最終的に，検出された領域が占めるピクセル数から実のサイズが計算される．4.1 では Roscher らの円検出手法について，4.2 では抽出する特徴量について，4.3 では実の判別方法について説明する．

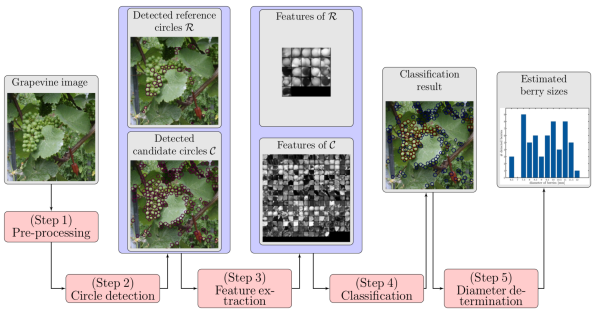

図 4: Roscher らの手法の流れ [4]

4.1 円検出

Roscher らの手法では実際のブドウの実のサイズを推定するために，画像における実のサイズを正確に求める必要がある．Roscher らの円検出手法の概要を図 5 に示す．初めに，ソーベルフィルタを用いて入力画像から勾配の向きと大きさを計算する．画像のエッジ部分に注目すると，勾配の向きはブドウの実の中心に向いているか，中心と正反対の向きを指している．この性質から，エッジ上の点から勾配方向に直線を引き，その直線が交わる点が円の中心点として推定できる．実際の画像では照明の影響等でエッジ点が正しく検出されていないことも多い．そのため実装上では，勾配方向の直線が通った回数の多い画素を中心位置として検出する．図 5 の右は，直線の通る回数が多いものが明るく，少ないものが暗くなるように作成された画像で

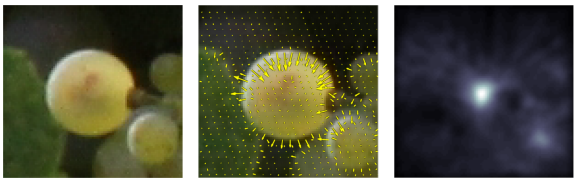

図 5: 円検出の過程 [4]．左から入力画像，勾配の向きと大きさ，投票結果を示す．

ある．この図より，実の中心位置が最も明るくなっていることが分かる．Roscher らの手法では Peng らの標準偏差リッジ検出器 [22] を用い，高速かつ高精度でエッジを抽出した後，上記の円検出を用いて円の検出を行っている．

4.2 特徴量

検出した領域がブドウの実であるか判断するために，Roscher らの手法では色特徴量，HoG 特徴量 [23]，GIST 特徴量 [24] の 3 つの特徴量が用いられる．色特徴量では画像の RGB 色空間が使用され，人工物や地面といった背景とブドウの実を分けることができる．HoG 特徴量では画像の勾配情報が特徴量として表現される．ブドウの実のように円形構造をもつ物体は特徴的な HoG 記述子を示すため，実の判定に利用できる．GIST 特徴量画像全体から求められる大域的特徴量である．

4.3 分類

検出された領域をがブドウの実かブドウの実でないかの分類は 2 つの情報を基に行われる．1 つは事後確率である．この事後確率は，4.2 で述べた特徴量を確信度に変換することで求めることができる．もう 1 つは空間の事前知識である．ブドウの実は房に成るため，ある程度集まっている．すなわち，検出された候補円が集まって分布していれば，その候補円がブドウの実である可能性は大きくなる．この性質を利用して事前知識が得られる．Roscher らはこれら 2 つの情報を条件付き確率場 [25] に与えることでブドウの実とそれ以外の分類を実現した．

5. 実験

作成したブドウの実検出評価用データセットの画像を入力とし，4.1 の円検出手法により画像からブドウの実の候補領域を検出した．今回の実験では Roscher らの手法のうち，円検出を行う部分のみを用い，特徴量によるブドウの実の判別は行っていない．また，最終的なブドウの実の中心座標の決定には，図 5 の右の図に対してピーク検出を行い，中心座標を決定した．

5.1 実験条件

画像は 5,476×3,648 画素から 684×456 画素にリサイズして評価を行った．計算には Opteron 6348 を用いた．コア数は 24 で，メモリ容量は 512 GB，クロック数は 2.8 GHz である．円検出器の実装には OpenCV を使用した．円検出の精度評価は目視による評価で行った．

図 6: Roscher らの手法をデータセットに適用した結果．赤い円は検出されたブドウの実の候補領域を示す．

5.2 実験結果

実験の結果の一例を図 6 に示す．画像中の赤色の円は検出されたブドウの候補領域を表す．Roscher らの円検出手法を用いることで，画像からブドウの実を検出することができた．

一方で Roscher らの円検出手法のみでは真円に近い形をしたブドウの実しか検出することができなかった．また，実とは関係のない領域も検出された．そのため，深層学習を用いない Roscher らの方法では実の検出が困難であり，画像からブドウの実を検出することは困難であることがわかった．

6. まとめと今後の課題

ブドウの実検出手法評価のために作成された橋本のデータセットは難しさの評価を行っていなかった．そこで本稿では，橋本のデータセットの公開を目的として，Roscher らの手法のうち円検出手法を用い，画像からブドウの実の候補領域を検出することでデータセットの評価を行った．結果，一部の実の候補領域の検出に成功した．しかし，実とは関係のない領域まで検出されたため，深層学習を用いない方法ではブドウの実の検出が困難であることが分かった．

今後の予定として，深層学習ベースの手法でも実の検出の評価を行う．用いる手法としては，Mask R-CNN を予定している．そして，最終的に摘粒データセットの評価を行う．

© 2021 Information Processing Society of Japan

謝辞 本研究は 2019 年度電気普及財団研究調査助成，大阪府信用農業協同組合連合会令和 2 年度産学連携研究支援事業による研究成果に基づく．

参考文献

[1] 内海ゆづ子，三木啓輔，尾形亮輔，大林拓実，三輪由佳，岩村雅一，黄瀬浩一：ブドウ摘粒支援のための果房 3 次元構造推定，第 21 回画像の認識・理解シンポジウム予稿集，No. DS-7 (2018).

[2] 農林水産省大臣官房統計部：「農林水産統計令和元年産日本なし、ぶどうの結果樹面積、収穫量及び出荷量」(2020).

[3] 社団法人山梨県果樹園芸会：葡萄の郷から ～ おいしいブドウのできるまで ～ (2016).

[4] Roscher, R., Herzog, K., Kunkel, A., Kicherer, A., Töpfer, R. and Förstner, W.: Automated Image Analysis Framework for the High-Throughput Determination of Grapevine Berry Sizes Using Conditional Random Fields, *Computers and Electronics in Agriculture*, pp. 148–158 (2014).

[5] Xu, S., Liu, D., Bao, L., Liu, W. and Zhou, P.: MHP-VOS: Multiple Hypotheses Propagation for Video Object Segmentation, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 314–323 (2019).

[6] Xu, K., Wen, L., Li, G., Bo, L. and Huang, Q.: Spatiotemporal CNN for Video Object Segmentation, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1379–1388 (2019).

[7] Wang, L., Li, D., Zhu, Y., Tian, L. and Shan, Y.: Dual Super-Resolution Learning for Semantic Segmentation, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3774–3783 (2020).

[8] Girshick, R., Donahue, J., Darrell, T. and Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587 (2014).

[9] Girshick, R.: Fast R-CNN, *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1440–1448 (2015).

[10] Ren, S., He, K., Girshick, R. and Sun, J.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, *Proceedings of Advances in Neural Information Processing Systems*, Vol. 28 (2015).

[11] He, K., Gkioxari, G., Dollár, P. and Girshick, R.: Mask R-CNN, *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2961–2969 (2017).

[12] Hung, W.-C., Jampani, V., Liu, S., Molchanov, P., Yang, M.-H. and Kautz, J.: SCOPS: Self-Supervised Co-Part Segmentation, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 869–878 (2019).

[13] Shen, Y., Ji, R., Wang, Y., Wu, Y. and Cao, L.: Cyclic Guidance for Weakly Supervised Joint Detection and Segmentation, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 697–707 (2019).

[14] Ahn, J., Cho, S. and Kwak, S.: Weakly Supervised Learning of Instance Segmentation With Inter-Pixel Relations, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2209–2218 (2019).

[15] Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. and Zitnick, C. L.: Microsoft COCO: Common Objects in Context, *Proceedings of European Conference on Computer Vision*, pp. 740–755 (2014).

[16] Santos, T. T., de Souza, L. L., dos Santos, A. A. and Avila, S.: Grape detection, segmentation, and tracking using deep neural networks and three-dimensional association, *Computers and Electronics in Agriculture*, Vol. 170, No. 105247 (2020).

[17] Richter, S. R., Vineet, V., Roth, S. and Koltun, V.: Playing for Data: Ground Truth from Computer Games, *Proceedings of European Conference on Computer Vision*, Vol. 9906, pp. 102–118 (2016).

[18] Ionescu, C., Papava, D., Olaru, V. and Sminchisescu, C.: Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 36, No. 7, pp. 1325–1339 (2013).

[19] Sigal, L., Balan, A. O. and Black, M. J.: HUMANEVA: Synchronized Video and Motion Capture Dataset and Baseline Algorithm for Evaluation of Articulated Human Motion, *International Journal of Computer Vision*, Vol. 87, No. 1-2, p. 4 (2010).

[20] Mehta, D., Rhodin, H., Casas, D., Fua, P., Sotnychenko, O., Xu, W. and Theobalt, C.: Monocular 3D Human Pose Estimation in the Wild Using Improved CNN Supervision, *Proceedings of 2017 International Conference on 3D Vision (3DV)*, pp. 506–516 (2017).

[21] Joo, H., Liu, H., Tan, L., Gui, L., Nabbe, B., Matthews, I., Kanade, T., Nobuhara, S. and Sheikh, Y.: Panoptic Studio: A Massively Multiview System for Social Motion Capture, *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3334–3342 (2015).

[22] Hidayat, R. and Green, R. D.: Real-time Texture Boundary Detection from Ridges in the Standard Deviation Space, *Proceedings of the British Machine Vision Conference*, pp. 1–10 (2009).

[23] Dalal, N. and Triggs, B.: Histograms of oriented gradients for human detection, *Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 886–893 (2005).

[24] Oliva, A. and Torralba, A.: Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope, *International Journal of Computer Vision volume*, Vol. 42, pp. 145–175 (2001).

[25] Lafferty, J. D., McCallum, A. and Pereira, F. C. N.: Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data, *Proceedings of the Eighteenth International Conference on Machine Learning*, pp. 282–289 (2001).

© 2021 Information Processing Society of Japan