1. 実験概要

本実験では，2目的ナップサック問題の解が与えられたときに，その解が制約を違反するかを推論する識別器の獲得を目標とする．具体的には，与えられた解について，

* 両方の目的の制約を違反: class1
* 目的1の制約を違反: class2
* 目的2の制約を違反: class3
* 両方の制約を違反しない: class4

として，各クラスを正確に識別する識別器の獲得を試みる．

本実験の実験設定は，

* ナップサック問題
  + アイテム数: 100
  + ナップサック容量: (全アイテムの大きさの総和)×0.5
* 学習データ
  + 決定変数: 各変数について0.5の確率で0か1に設定
  + データ数: 10,000
* 識別器
  + *k*-Nearest Neighbor (*k*-NN)
  + Random Forest (RF)

とする．なお，識別器のハイパーパラメータは，scikit-learnのデフォルトの設定を用いた．なお，*k*-NNのデフォルトの設定は[1]，RFのデフォルトの設定は[2]に記載されている．

学習データの教師ラベルの分布を図1に示す．ここで，*x*軸のラベルは，制約を違反している目的の番号を表しており，制約を満たしている場合は ”feasible” と表記している．図1より，このデータは，クラス分布に偏りがあるデータということがわかる．

また，より簡単な問題設定である， ”目的1を制約違反しているか” および ”目的2を制約違反しているか” を識別する問題についても実験を行った．両問題における学習データの教師ラベルの分布を図2・3にそれぞれ示す．ここでは，元の問題を”Problem 1”，目的1に関する問題を”Problem 2”，目的2に関する問題を”Problem 3”と表記する．

C:\Users\t.urita\AppData\Local\Microsoft\Windows\INetCache\Content.Word\fig1.emf

図1: Problem 1の教師ラベルの分布



図2: Problem 2の教師ラベルの分布



図3: Problem 3の教師ラベルの分布

1. 実験結果

実験結果を表1に示す．表1より，いずれの問題においても，RFは*k*-NNよりも優れた識別性能を示したことがわかる．また，Problem 1ではRFであっても，識別性能が70%に満たずあまりよい結果とは言えなかった．しかし，簡単化した問題では，80%を超える性能を示し，最低限実用に耐えうるレベルといえるのではないか．

表1: 各問題で得られた識別器の識別精度(Accuracy)[%]

|  |  |  |  |
| --- | --- | --- | --- |
| Classifier | Problem 1 | Problem 2 | Problems3 |
| *k*-NN | 56.34 | 69.98 | 69.34 |
| RF | 68.13 | 81.55 | 80.94 |

参考文献

[1]https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighbors  
Classifier.html

[2] https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForest  
Classifier.html