The nutcracker framework for ensemble interpretability

Zeev Ben Mordehay, O.; Duivesteijn, W.; Pechenizkiy, M.

Published: 01/10/2017

Document Version
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
• A submitted manuscript is the author's version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 07. Dec. 2018
The Nutcracker Framework for Ensemble Interpretability

Oren Zeev-Ben-Mordehai
o.zeev.ben.mordehay@tue.nl

Introduction

The basic principles behind ensembles (e.g., Random Forest [1], AdaBoost [2]) are simple. But we’re still in trouble when attempting to explain the logic taken. Where does the problem lie? The reason that ensembles are effective is that the base estimators “work together” and compensate each for the others’ shortcomings.

The Nutcracker Framework

Given a trained ensemble and the relevant training / test dataset, construct prediction matrix, \( M \), cases (rows) against predictions (columns). Bi-cluster \( M \) to a given number of \( R \times C \) bi-clusters [3]. Now, investigate performance per bi-cluster (\( R \times C \)). Identify feature importance per base estimators group (C). Describe each of the \( R \) cases subgroups in terms of features and values. We use Exceptional Model Mining [4] for that task.

Performance of the ensemble against the dataset compared to performance of base estimator groups against subgroups of cases, adds transparency.

Table 1 Descriptions for the subgroups

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SG1</td>
<td>alcohol &gt; 9.97 and residual sugar &lt;= 19.23 and total sulfur dioxide &lt;= 192.0</td>
</tr>
<tr>
<td>SG2</td>
<td>alcohol &lt;= 9.97 and chlorides &lt;= 0.18 and quality != 4</td>
</tr>
<tr>
<td>SG3</td>
<td>alcohol &gt; 9.97 and alcohol &lt;= 12.93 and sulphates &lt;= 0.73</td>
</tr>
<tr>
<td>SG4</td>
<td>alcohol &gt; 9.97 and total sulfur dioxide &lt;= 192.0 and fixed acidity &lt;= 10.71</td>
</tr>
<tr>
<td>SG5</td>
<td>quality != 6 and density &lt;= 0.99 and color != 1</td>
</tr>
<tr>
<td>SG6</td>
<td>alcohol &gt; 9.97 and volatile acidity &lt;= 0.72 and pH &gt; 3.09</td>
</tr>
</tbody>
</table>

References