Association rule mining using fishbone diagrams

Students: Daria Likholetova, Nina Lukashina
Supervisors: Petr Tsurinov, Oleg Shpynov
Project concept

Bioinformatics data \(\xrightarrow{ARM}\) Ishikawa diagram

Conditions \(\rightarrow\) Target
Why do we need new ARM method?

_Fishbone ARM method:_ ARM + information theory

<table>
<thead>
<tr>
<th></th>
<th>FP-growth</th>
<th>Decision Tree</th>
<th>Fishbone ARM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hierarchical rules</strong></td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td><strong>Visualization</strong></td>
<td>Table</td>
<td>Table, tree</td>
<td>Table, graph, fishbone</td>
</tr>
<tr>
<td><strong>Interpretability</strong></td>
<td>Ok</td>
<td>Good</td>
<td>Good</td>
</tr>
</tbody>
</table>
Project goals and objectives

The goal: validate ARM approach with existing data and apply it to non-published data

- Improve method usability
- Explore the bioinformatics article and extract scientific statements
- Check obtained statements with ARM approach
- Find and investigate new statements from non-published data
Improving service usability

Old service: command line tool to run fishbone algorithm + web page for visualization

New service: client-server application with UI to run analysis using different algorithms and visualize results
Validating method with public data

A Validated Regulatory Network for Th17 Cell Specification (Ciofani et al., 2012)

TF feedback loops in Th17 (from the article)

Fishbone diagram based on article’s data

<< BATF is associated with IRF4 >>

Complex formation on DNA
Exploring InCHIANTI data

- Aging study
- Data: biochemical blood analysis, nutrition profiling, physical examinations, self-reported health, etc.
- Number of study participants: 1308
Applying ARM approach to the data: Old

Target: Age over 75 years
Applying ARM approach to the data: Old

Target: Age over 75 years

- TNF-alpha receptor 1
- Creatinine clearance
- Homocystein
- IGFbp-1
- PIIINP
- Urine nitrites
- Homocystein
- X_AGEL_is_higher_75

Levels:
- High
- Low
- Not normal
Applying ARM approach to the data: Young

Target: Age less than 75 years

Creatinine clearance

High
Not normal
Not low
Low
Applying ARM approach to the data: Young

- Creatinine clearance
- Cortisol:Dehydroepiandrosterone sulfate ratio
- Urinary hepcidin
- Dehydroepiandrosterone sulfate
- TNF-alpha receptor I
- Urinary hepcidin
- Dehydroepiandrosterone sulfate

Levels:
- High
- Not normal
- Not low
- Low
Results

- Provided client-server application with extended UI for Fishbone ARM method
- Method was validated using known rules from published data
- InCHIANTI data analysis using Fishbone algorithm showed key factors of old age
- Literature support the Fishbone-retrieved factors association with aging
- More biological rules to be investigated
References

1. Creatinine: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4354997/
3. PIIINP (Intact N-terminal propeptide of type-3 procollagen)
   https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4587692/
4. Homocysteine
   https://academic.oup.com/biomedgerontology/article/60/9/1190/560525
5. IGF binding protein-1 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3286655/
6. Urinary hepcidin https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3752551/
7. Dehydroepiandrosterone sulfate
   https://www.tandfonline.com/doi/abs/10.1080/13685538.2019.1592151
8. Cortisol:DHEAS ratio
   https://www.sciencedirect.com/science/article/pii/S0889159111000766