Medical Mining Topic – Summer Term 2018

Multifactorial Diseases & Disorders

Which risk factors and characteristics are associated with a disease or disorder?

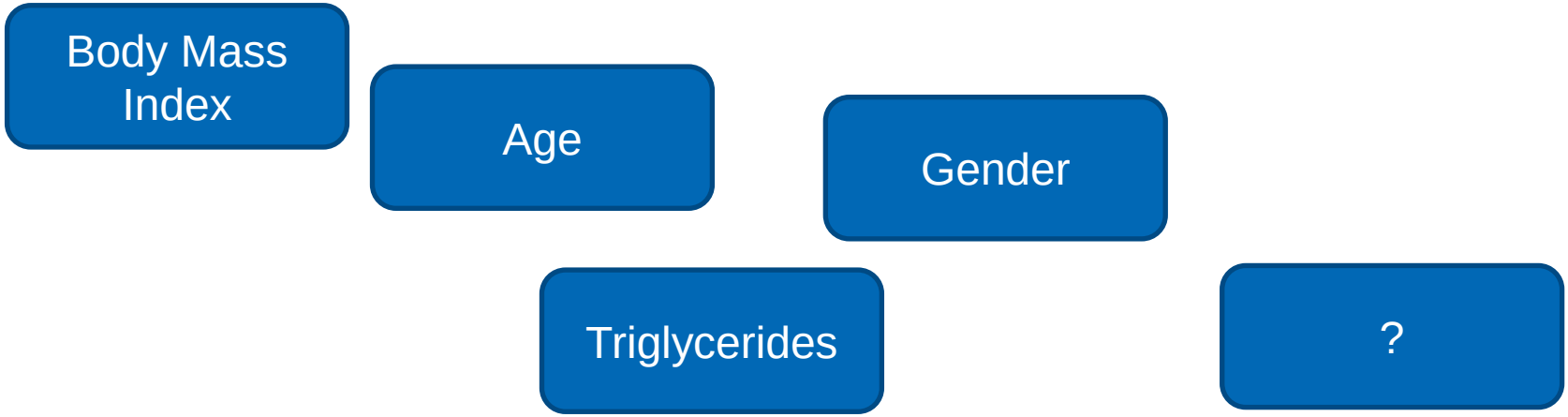
- Lifestyle factors like daily alcohol intake
- Genetic factors about individual predispositions
- Environmental factors like water pollution

Longitudinal epidemiological studies

- Multiple study waves
- Same–participant followups
- Repeated medical examination programmes

Hepatic Steatosis

- "Fatty liver"
- Example of a multifactorial disorder
- Multiple independent risk factors and important features
- Potential predictors according to [BBM+06]:



How do we find relevant factors? – One example:

- Given a study participant dataset with feature set *F* and target disease *C*:

  (1) Enumerate a number of subspaces of *F*

  (2) Use each subspace to predict the *C*

  (3) Compare the performance of each space and analyze the best ones

- (1), (2) and (3) is easy when a lot of ground truth is available
- However, how do we proceed when only **limited** ground truth w.r.t. the disease under study is available?
- One way is to enumerate and evaluate subspaces according to their best match between internal structure of the data and the limited knowledge w.r.t. the medical outcome under study

TP : Subspace Clustering on (Evolving) Objects with Constraints

How to utilize limited knowledge about the target disease/disorder?

- Limited knowledge stating that two participants have or have not the disease / disorder under study (Must–Link / Not–Link constraints)
- Assessing relevancy of subspaces

TODO

- Literature has to be reviewed and compared: Find and review different subspace methods.
- Extend a method for constraints exploitation.
- Develop a quality function which scores subspaces according to their relevancy w.r.t. the target concept, which incorporates the agreement between internal structure and constraints satisfaction.
- Implement and evaluate your method on real–world data.

The End

Thank you very much!