Hannah Blocher and Georg Schollmeyer
Ludwig–Maximilians–Universität München

A Statistical Depth Function for Non–Standard Data based on Formal Concept Analysis

ECDA 2021, 07.07.2021

Overview

1. The closure system and family of implications given by a formal context
2. The definition of the statistical depth function based on a reduced family of implications
3. Examples for the statistical depth function

Formal Concept Analysis

Definition (Formal Context)

A formal context is given by a triple $\mathbb{K} = (G, M, I)$. G corresponds to the set of *objects*, M to the set of *attributes* and I defines a binary relation between G and M.

- A formal context defines a closure system on G with closure operator

$$\gamma(A) = \{g \in G \mid \forall m \in M \text{ with } \forall a \in A, \ (a, m) \in I \text{ is true: } (m, g) \in I\}$$

  for $A \subseteq G$.

- The closure system can be completely described by a family of implications.

Example: Spatial Data

Formal Context:

$$\left(\mathbb{R}^2,\ \{H \mid H \text{ closed halfspace}\},\ \{(p, H) \mid p \text{ lies in halfspace } H\}\right)$$

In this case:

- The closure system corresponds to the closed convex sets in $\mathbb{R}^2$ with the convex hull as closure operator *ConvHull*.
- Each implication which describes the closure system is given by $A \to B$ with $A, B \subseteq \mathbb{R}^2$ and $B \subseteq ConvHull(A)$.

Union–free Generic Family of Implications and the resulting Depth Function

Definition (Union–free Generic Family of Implications)

The *union–free generic* family of implications, $\mathcal{UFG}$, consists of implications $A \to B$ for which the following is true:

1. they are non–trivial,
2. they have a minimal premise and a maximal conclusion, and
3. they can not be divided by a further set of implications $\mathcal{I}$ with minimal premise and maximal conclusion. We say an implication $A \to B$ can be divided by a set of implications $\mathcal{I}$ if the union of the premise A equals the union of the premises of the implications in $\mathcal{I}$ and the same holds for the conclusions.

$\to$ This family of implications is sufficient to describe the corresponding closure system.

Union–free Generic Family of Implications and the resulting Depth Function

Definition (Statistical Depth Function of a Formal Context)

For a given formal context $\mathbb{K} = (G, M, I)$ we define the statistical depth function based on the corresponding union–free generic family of implications $\mathcal{UFG}$ by

$$D(\cdot, \mathbb{K})\colon G \to [0,1],\ g \mapsto \frac{\#\{A \to B \in \mathcal{UFG} \mid g \in B\}}{\#\{A \to B \in \mathcal{UFG}\}}.$$

Example: Spatial Data

Formal Context for a subset $G \subseteq \mathbb{R}^2$:

$$\mathbb{K} = \big(G,\ \{H \mid H \text{ closed halfspace}\},\ \{(p, H) \mid p \text{ lies in halfspace } H\}\big).$$

The union–free generic family of implications corresponds to

$$\mathcal{UFG} = \{A \rightarrow ConvHull(A) \mid \#A \in \{2,3\} \text{ and } A \text{ not degenerated}\}$$

and therefore the statistical depth function is given by

$$D(\cdot, \mathbb{K})\colon G \rightarrow [0,1],$$
$$p \mapsto \frac{\#\{ConvHull(A) \mid A \subseteq G, \#A = \{2,3\} \text{ and } p \in ConvHull(A)\}}{\#\{ConvHull(A) \mid A \subseteq G, \#A = \{2,3\}\}}.$$

Further Examples

For ordinal data, we obtain a formal context $\mathbb{K} = (G, M, I)$ by the use of interordinal scaling and get

$$\begin{aligned} argmax_{g \in G} D(g, \mathbb{K}) &= argmax_{g \in G} \frac{\#\{[a,b] \mid g \in [a,b], a, b \in G\}}{\#\{[a,b] \mid a, b \in G\}} \\ &= median(G). \end{aligned}$$

By the use of nominal scaling (with grouped classes), we obtain a formal context $\mathbb{K} = (G, M, I)$ for nominal data with

$$\begin{aligned} argmax_{g \in G} D(g, \mathbb{K}) &= argmax_{g \in G} \frac{\#\{a \in G \mid g \text{ in the same class as } g\}}{\#G} \\ &= modus(G). \end{aligned}$$

Summary: Statistical Depth Function for Non–Standard Data

- We defined a statistical depth function which only depends on a suitable formal context for the non-standard data.
- For spatial data it is possible to obtain a similar definition as the simplicial depth function.
- For ordinal and nominal data, we get the median and respectively the modus as the argument of the maximum of the statistical depth function.

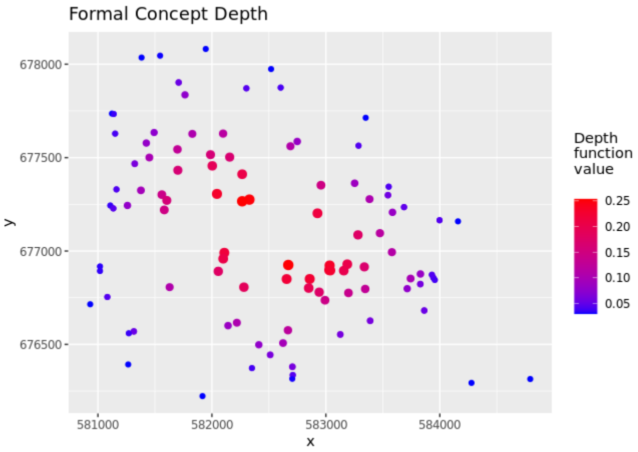
Literatur I

[1] Yves Bastide et al. *Mining Minimal Non-redundant Association Rules using Frequent Closed Itemsets*. Springer Berlin Heidelberg, 2000, pp. 972–986.

[2] Bernhard Ganter, Gerd Stumme, and Rudolf Wille. *Formal Concept Analysis: Foundations and Applications*. Springer Berlin Heidelberg, 2005.

[3] Bernhard Ganter and Rudolf Wille. *Formal Concept Analysis: Mathematical Foundations*. Springer Berlin Heidelberg, 2012.

Appendix



This is a subset of the GORILLAS data set which is stored and documented within the R-package SPATSTAT.DATA