Fakulteta za elektrotehniko,
računalništvo in informatiko

Visualization of Numerical Association Rules by Hill Slopes

Iztok Fister Jr., Dušan Fister, Andres Iglesias, Akemi Galvez,
Eneko Osaba, Javier Del Ser, Iztok Fister

Contributions of this study

- A new method for visualization of numerical association rules is proposed
- Proposed method takes an inspiration in Tour de France cycling race
- The proposed visualization method is practically evaluated

Association rule mining

An association rule can be defined as implication:

$$X \Rightarrow Y, \tag{1}$$

where $X \subset O$, $Y \subset O$, in $X \cap Y = \emptyset$. The following two measures are defined for evaluating the quality of association rule:

$$conf(X \Rightarrow Y) = \frac{n(X \cup Y)}{n(X)}, \tag{2}$$

$$supp(X \Rightarrow Y) = \frac{n(X \cup Y)}{N}, \tag{3}$$

where $conf(X \Rightarrow Y) \geq C_{min}$ denotes confidence and $supp(X \Rightarrow Y) \geq S_{min}$ support of association rule $X \Rightarrow Y$.

Inspiration

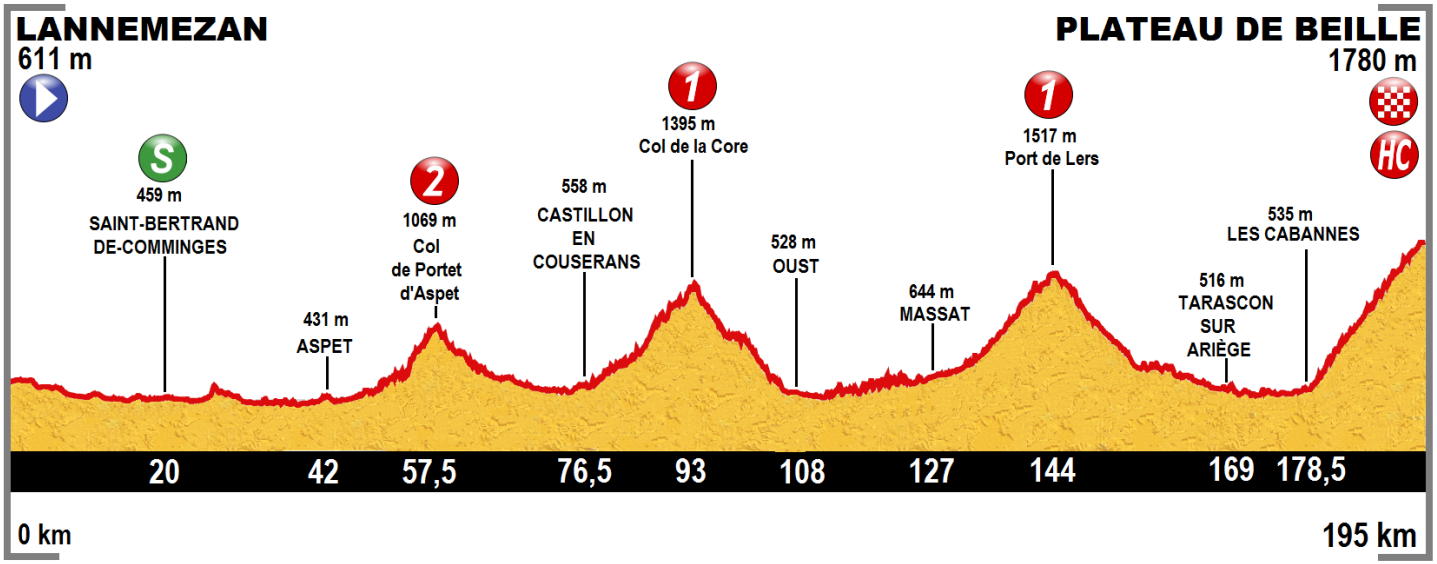

Figure: Example of TDF Stage 12, TDF 2015.

Image source: Wikimedia

Mathematical model 1/2

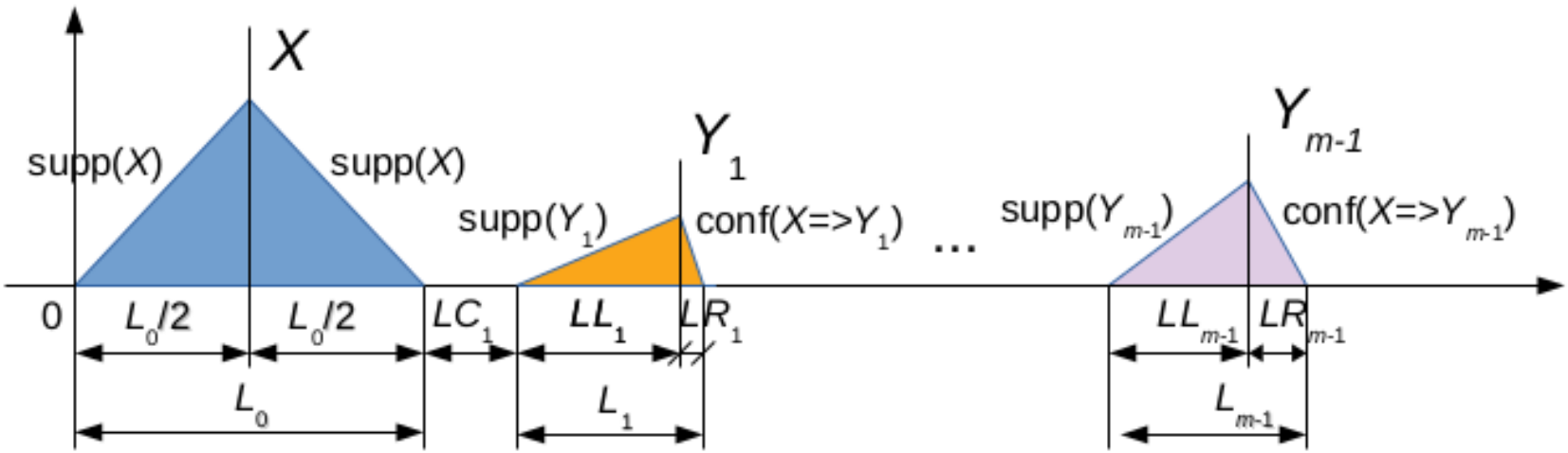

Figure: Mathematical model of virtual hills, on which the new visualization method is founded.

$$supp(X) \equiv conf(X \Rightarrow Y)$$

Mathematical model 2/2

The position of the triangle on the line is determined as follows:

$$pos_i = L_0 + \sum_{j=1}^{i-1} (LC_j + LL_j + LR_j) + (LC_i + LL_i)\,, \tag{4}$$

where L_0 denotes the diagonal length of the model triangle, $LC_j \propto conf(X \Rightarrow Y_j)$ is the distance between two subsequent triangles, LL_j, expressed as follows:

$$\begin{aligned} \cos\alpha &= \frac{supp(Y_j)}{L_j}, \quad \text{for } j = 1, \ldots, m-1, \\ LL_j &= supp(Y_j) \cdot \cos\alpha = \frac{supp^2(Y_j)}{L_j}, \end{aligned} \tag{5}$$

while LR_j as:

$$\begin{aligned} \cos\beta &= \frac{conf(X \Rightarrow Y_j)}{L_j}, \quad \text{for } j = 1, \ldots, m-1, \\ LR_j &= conf(X \Rightarrow Y_j) \cdot \cos\beta = \frac{conf^2(X \Rightarrow Y_j)}{L_j}. \end{aligned} \tag{6}$$

Experiments & results

- The goal of our experimental work was to show that the mined association rules can be visualized using the proposed visualization method.
- Two selected association rules are visualized.
- Association rules were mined using ARM-DE
- Visualization was done in Matlab software framework, using the colored 3-D ribbon plot

Transaction database

Nr.	Feature	Domain
F-1	Duration	[43.15, 80.683]
F-2	Distance	[0.00, 56.857]
F-3	Average HR	[72, 151]
F-4	Average altitude	[0.2278, 1857.256]
F-5	Maximum altitude	[0.0, 0.0]
F-6	Calories	[20.0, 1209]
F-7	Ascent	[0.00, 1541
F-8	Descent	[0.00, 1597]

Table: The transaction database consists of seven numerical features, whose domain of feasible values are illustrated in Table.

Association rules

Feature	Scenario 1	Scenario 2
Duration	[76.67,78.07]	[46.95,65.87]
Distance	[14.28,26.32]	[26.24,53.30]
Average HR	[78.79,114.92]	[104.12,141.40]
Average altitude	[631.70,1809.21]	[17.59,547.05]
Calories	[774.92,1161.43]	[1096.82,1209.00]
Ascent	[0.00,10.00]	[0.00,74.19]
Descent	[0.00,54.19]	[0.00,623.88]

Table: Mined numerical features in association rules.

Scenario 1

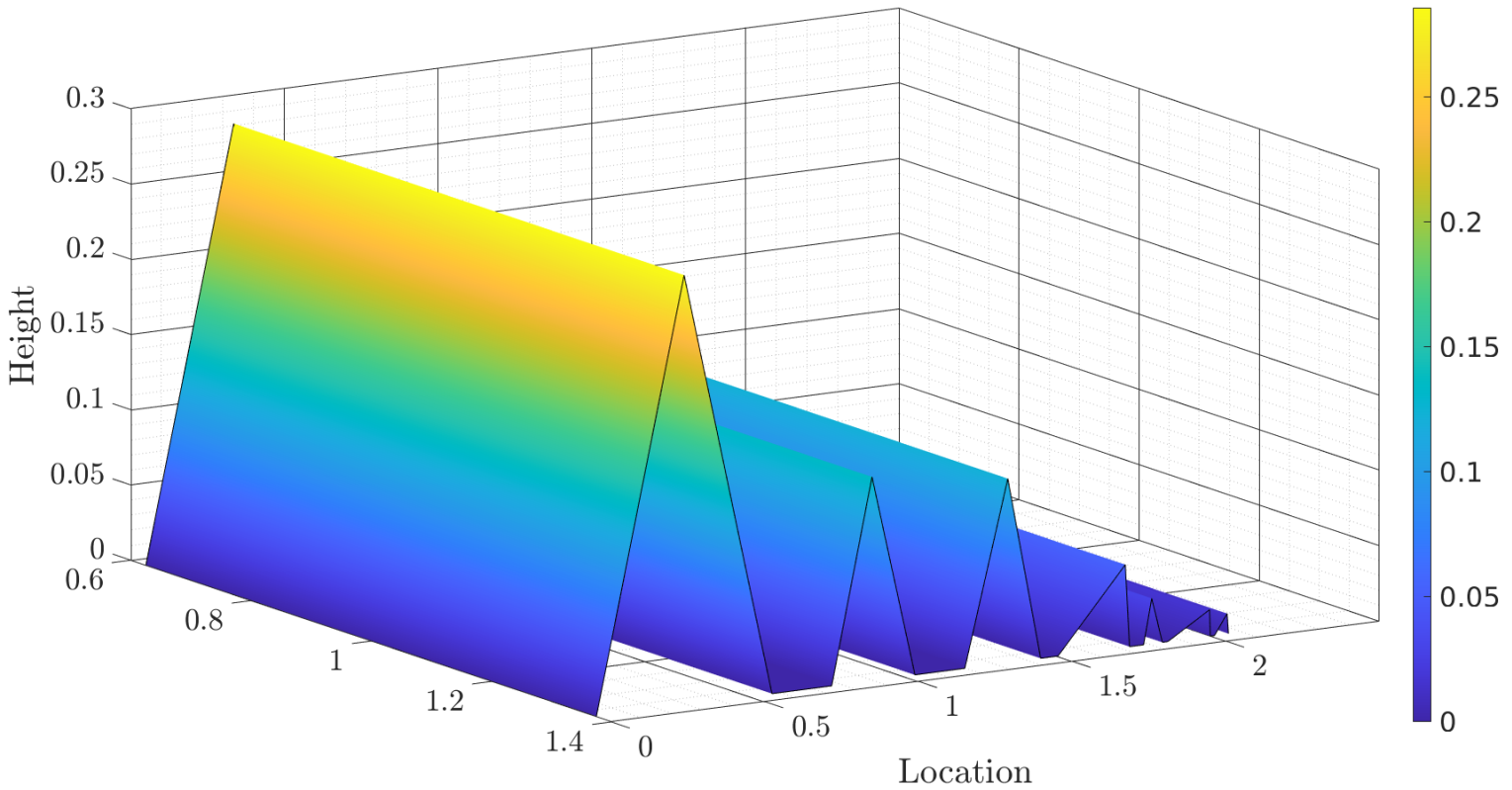

Figure: Visualization of scenario 1

Scenario 2

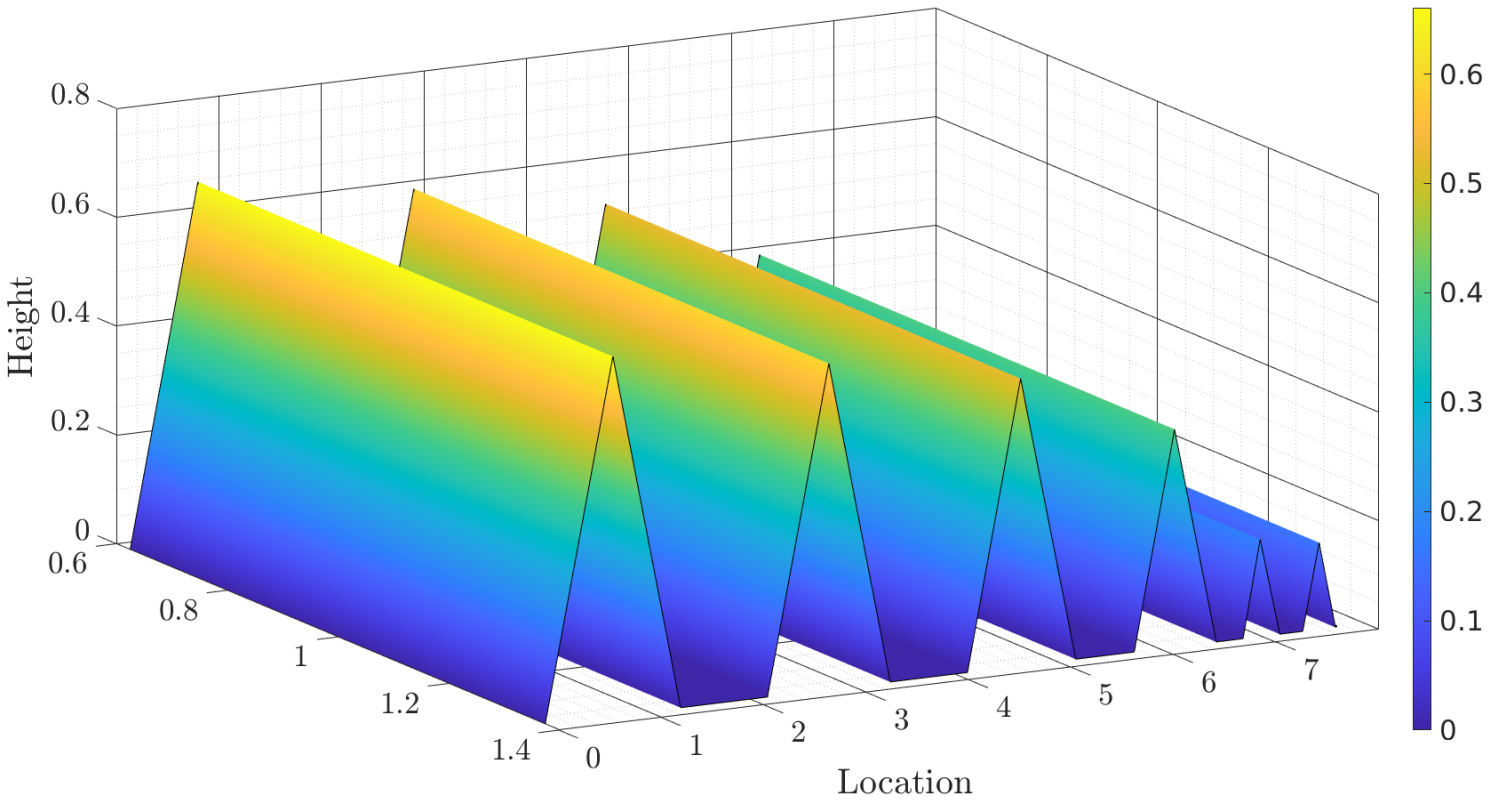

Figure: Visualization of scenario 2

Conclusion

- Paper presented a new visualization method inspired by the TDF.
- The virtual hill slopes, reflecting a probability of one attribute to be more interesting than the other, help a user to understand the relationships among attributes in a selected association rule.
- The visualization method was employed on a transaction database consisting of features characterizing the realized sports training sessions.
- The results of visualization showed the potential of the method, that is able to illustrate the hidden relationships in a transaction database in an easy and understandable way to the user.
- In the future, the method should be applied to another transaction databases.