

# Association discovery in two-view data

Matthijs van Leeuwen  
m.van.leeuwen@liacs.leidenuniv.nl  
LIACS, Leiden University, NL

Esther Galbrun  
esther.galbrun@inria.fr  
Inria Nancy – Grand Est, FR

Finding **small** and **non-redundant** sets of associations

that describe **how the two sides of a two-view dataset are related**.

## PROBLEM

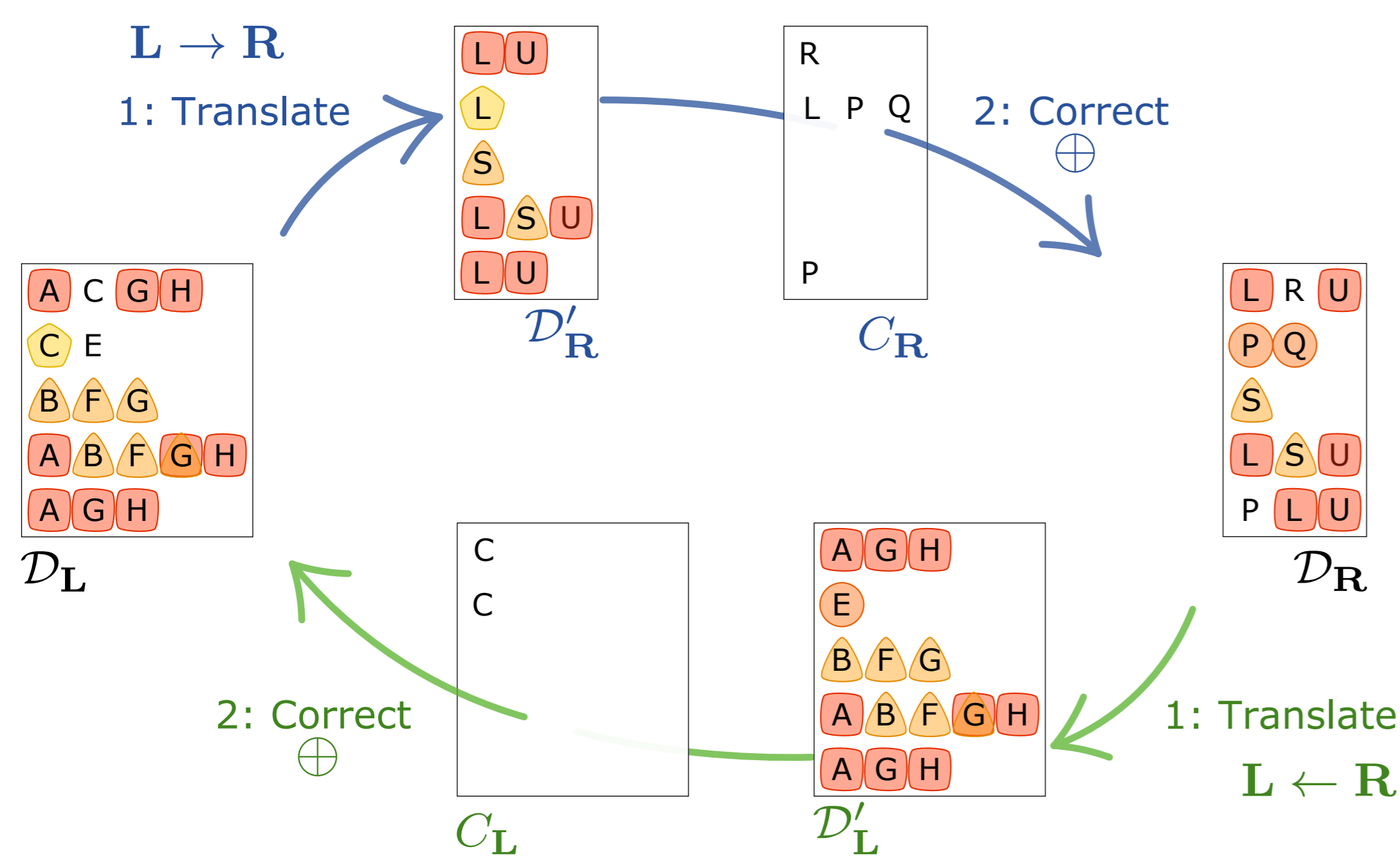
Two-view datasets are datasets whose attributes are naturally split into two sets, each providing a different view on the same set of objects.

What associations are present in these views?

A typical **exploratory data mining** question:

discovering patterns that together describe the structure of the data.

We are interested in associations that span both views. Associations might be unidirectional or hold in both directions, implying that the two observations occur mostly together.



Translation of a toy dataset consisting of the two views  $\mathcal{D}_L$  and  $\mathcal{D}_R$ , in both directions with translation table  $T$ .

## RELATED WORK

- Association rule mining
  - closed itemsets
  - statistical testing
- Pattern set mining
  - MDL-based modelling
- Exceptional Model Mining
- Redescription Mining

## EXPERIMENTS

Average compression ratios (14 datasets):

<i>T-SELECT(1)</i>	————	79.73%
<i>REREMI</i>	————	97.74%
<i>MAGNUM OPUS</i>	————	123.81%
<i>KRIMP</i>	————	300.73%

## CONTRIBUTIONS

- Introduction of pattern-based models for Boolean two-view data, i.e., *translation tables*;
- Model selection based on the Minimum Description Length (MDL) principle;
- Three *TRANSLATOR* algorithms.

<i>TRANSLATOR-SELECT(1)</i>		$c^+$
Red Fox $\leftrightarrow$ European Hedgehog $\wedge$ Least Weasel		0.98
Bank Vole $\leftrightarrow$ European Water Vole $\wedge$ Common Shrew $\wedge$ Eurasian Pygmy Shrew $\wedge$ Red Squirrel $\wedge$ Brown rat $\wedge$ Least Weasel		0.97
Brown long-eared bat $\wedge$ Field Vole $\wedge$ European Badger $\rightarrow$ Eurasian Pygmy Shrew		0.97
<i>MAGNUM OPUS</i>		
European Polecat $\wedge$ European Mole $\rightarrow$ European Hare		0.98
European Badger $\wedge$ European Mole $\rightarrow$ European Hare		0.97
Eurasian Water Shrew $\wedge$ European Mole $\rightarrow$ European Hare		0.97
<i>REREMI</i>		
European Mole $\wedge$ Red Fox $\leftrightarrow$ Harvest Mouse $\wedge$ European Hare		0.92
Brown long-eared bat $\wedge$ European Mole $\wedge$ Red Fox $\leftrightarrow$ Eurasian Pygmy Shrew $\wedge$ European Hare $\wedge$ Least Weasel		0.91
Bank Vole $\wedge$ Red Fox $\leftrightarrow$ European Pine Marten $\wedge$ Red Squirrel		0.91

## REFERENCES

M. van Leeuwen and E. Galbrun (2015) *Association discovery in two-view data*. TKDE **27**(12).  
<http://patternsthatmatter.org/software.php#translator>

[MAGNUM OPUS] G. I. Webb (2007) *Discovering significant patterns*. Machine Learning.

[KRIMP] J. Vreeken, M. van Leeuwen, and A. Siebes (2011) *Krimp: mining itemsets*

that compress. Data Mining and Knowledge Discovery.

[REREMI] E. Galbrun and P. Miettinen (2012) *From black and white to full color:*

extending redescription mining outside the boolean world. Statistical Analysis and Data Mining.



Universiteit  
Leiden  
The Netherlands

