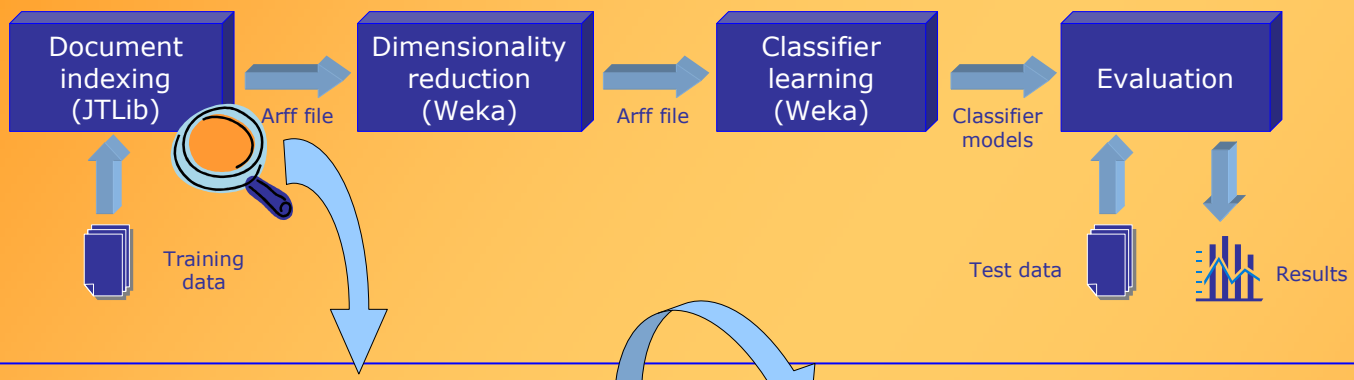
Feature Engineering and Quick Prototyping of PPI Classifiers*

Manuel J. Maña López
Jacinto Mata Vázquez
Universidad de Huelva
{manuel.mana, mata}@diesia.uhu.es

Francisco Carrero García
José Mª Gómez Hidalgo
Universidad Europea de Madrid
{francisco.carrero, jmgomez}@uem.es

Classification Process



Taxonomy of Attribute Types

Intrinsic
Information used to compute the attribute comes only from the same example

Contextual extrinsic
The information is obtained from the processed example, but also from other examples that have a strong relation with it

Global extrinsic
The information comes from all the examples in the set

JTLib

attribs
- Common pre-defined attributes

jtFormatter
- Pre-processing of attributes
- Processing of intrinsic attributes
- Processing of contextual and global extrinsic attributes
- Generation of data set

tasks
- Sub-pakages that implement several classifiers

Selected Attributes

Attribute type	Attribute Name
Global extrinsic	Unigrams Bigrams Trigrams

Ranking of Attributes

N-tuples with higher correlation coefficient		
interaction	two hybrid	yeast two hybrid
domain	domain of	two hybrid system
binding	yeast two	we show that
hybrid	with the	the yeast two
yeast	interacts with	a yeast two
with	in vitro	the interaction of
interacts	the interaction	to interact with
terminal	interact with	in vitro and
complex	interaction between	is required for
interact	interaction with	two hybrid screen

N-tuples with lower correlation coefficient		
ms	<number> de	dimensional gel electrophoresis
de	two dimensional	two dimensional gel
spots	protein spots	by <number> de
proteome	<number> d	maldi tof ms
proteomic	gel electrophoresis	<number> protein spots

Best Run Results

	Adaboost with Naïve Bayes Nº of unigrams/bigrams/trigrams 5140 / 18510 / 10224
Precision	0.583
Recall	0.952
F-measure	0.723

* Funding: Spanish Ministry of Education and Science and the European Union from the European Regional Development Fund (ERDF) - (TIN2005-08988-C02-01 and TIN2005-08988-C02-02); Madrid Regional Government under the IV PRICIT programme through the Joint Research Programme MAVIR (S-0505/TIC-0267).