Automatically assigning UCDs using machine learning

Norman Gray
University of Glasgow, UK

IVOA Interop, Madrid, 2014 May 21

University of Glasgow

...funded this work

```
CREATE TABLE vvvDetection(
multiframeID bigint not null,  --/D the UID of the relevant multiframe
extNum       tinyint not null, --/D the extension number of frame
cuEventID    int not null,     --/D UID of curation event
seqNum       int not null,     --/D the running detection number
[...]
```

what are the UCD1+ for these columns?

schemas: what we have

- *UCD1* on many columns
- *HIERARCH tags* on some columns
- *comments* on almost all columns

anything else?

- There is some units information
- ...which *eventually* turns out not to be a lot of help

want...

‘Intrinsic rms in H-band’

⇒ `stat.error;em.IR.H`

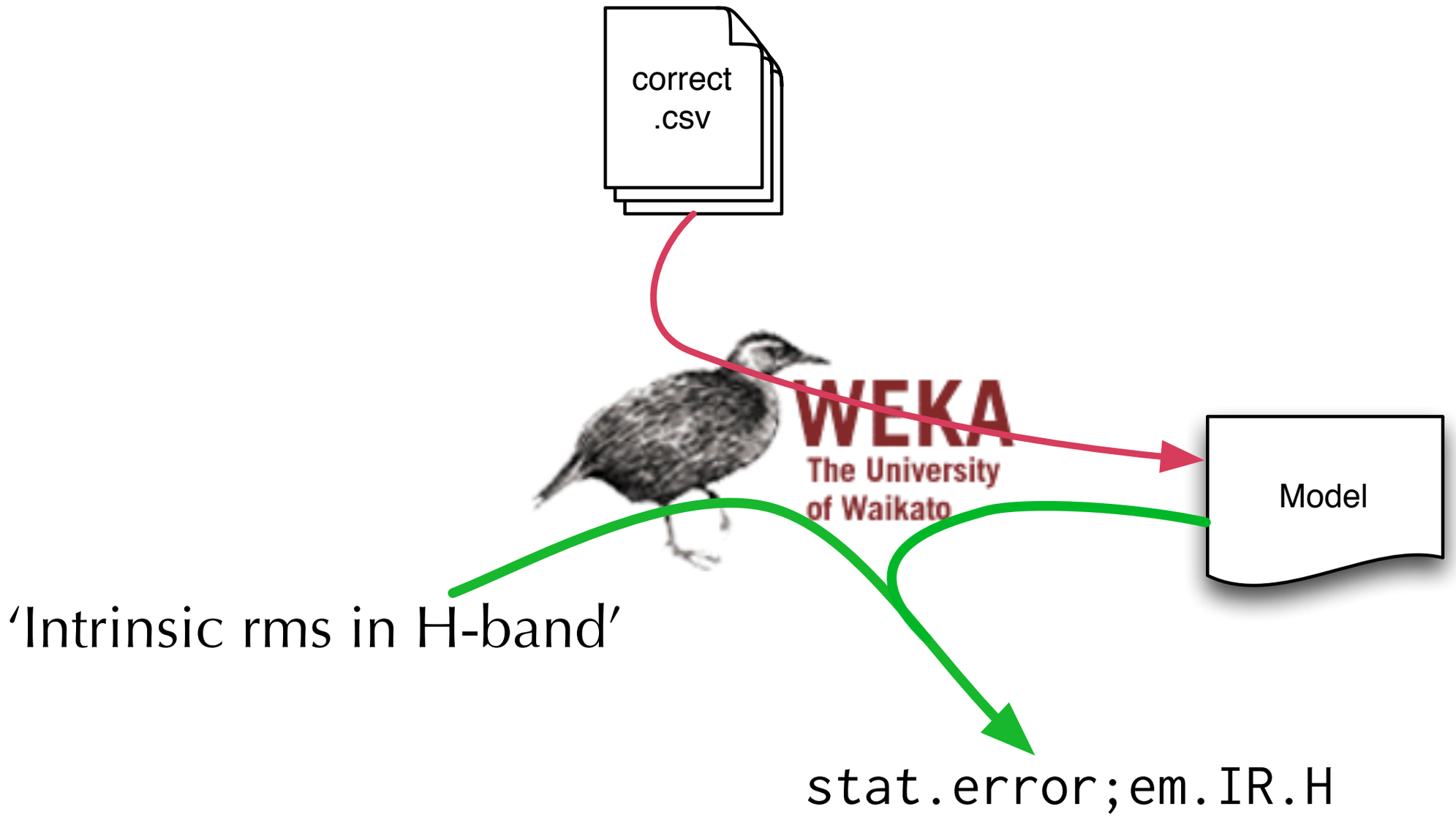
‘Classification of variability in this band’

⇒ `meta.code.class;src.var`

‘Angular separation between neighbours’

⇒ `pos.angDistance`

WEKA
The University
of Waikato


correct
.csv
WEKA
The University
of Waikato
Model
'Intrinsic rms in H-band'
stat.error;em.IR.H

training data

pos.angDistance
  ⇒ Angular separation between neighbours

stat.likelihood;em.IR.H
  ⇒ average confidence in 2 arcsec diameter default aperture (aper3) H

meta.code.class;em.IR.H
  ⇒ Best aperture (1-6) for photometric statistics in the H band

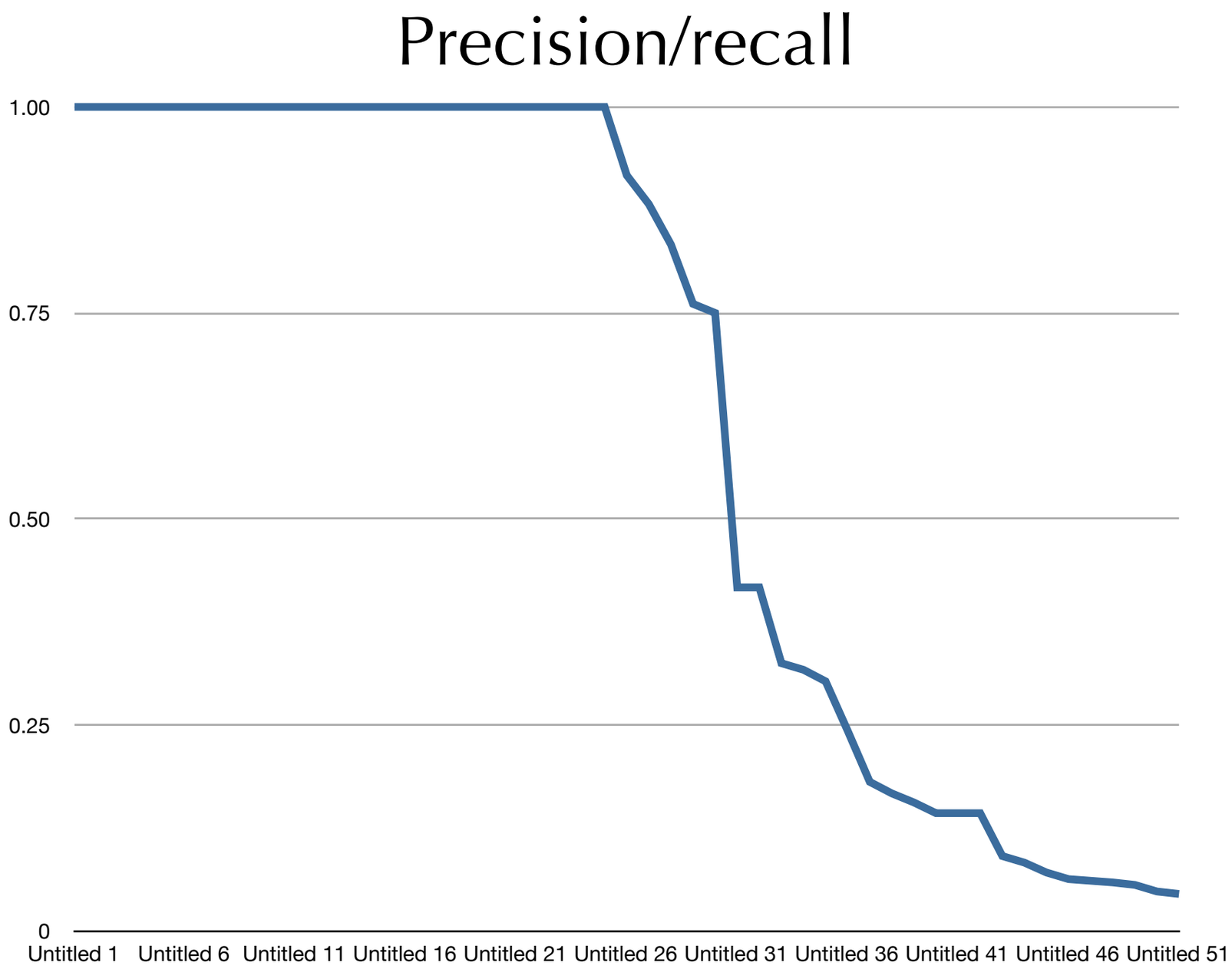
stat.fit.chi2
  ⇒ Chi square (per degree of freedom) fit to data (mean and expected rms)

phot.mag;stat.max;em.IR.H
  ⇒ Expected magnitude limit of frameSet in this in H band.

stat.fit.goodness;em.IR.H
  ⇒ Goodness of fit of Strateva function to astrometric data in H band

stat.error;em.IR.H
  ⇒ Intrinsic rms in H-band

time.interval;obs;stat.median
  ⇒ median gap between observations


Precision/recall
1.00
0.75
0.50
0.25
0
Untitled 1
Untitled 6
Untitled 11
Untitled 16
Untitled 21
Untitled 26
Untitled 31
Untitled 36
Untitled 41
Untitled 46
Untitled 51

- use other features in input
- use other features (units/dimensions) to veto assignments
- enlarge training set (might be quite biased right now)
- package and release