Feature Set Embedding for Incomplete Data – W 13

D. Grangier & I. Melvin NEC

Missing Features: examples as sets of (FeatureId, FeatureValue) pairs

- Missing Features

$$x = \begin{array}{c|ccccccc} \text{Feature} & \text{A} & \text{B} & \text{C} & \text{D} & \text{E} & \text{F} & \text{G} \\ \hline \text{Value} & 0.15 & ? & ? & ? & 0.28 & 0.77 & ? \end{array}$$

- Prior Art: Examples are vectors

impute | integrate out | example-specific subspace

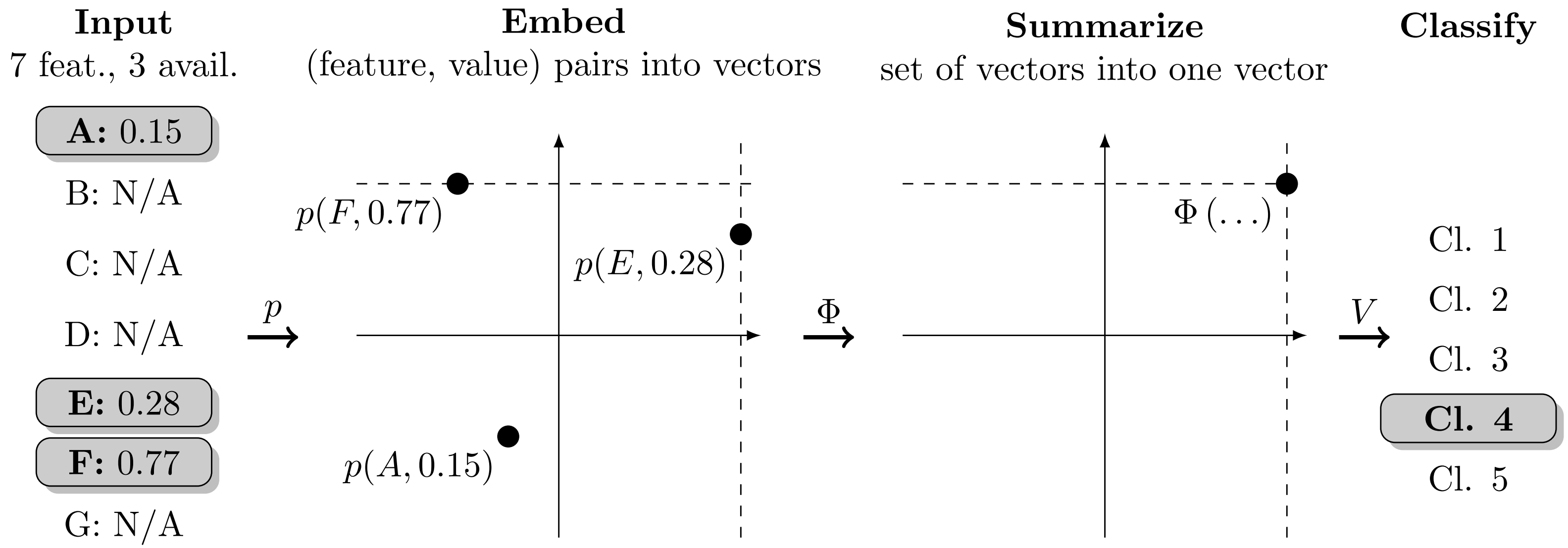
- This Work: Examples are sets of (FeatureId, FeatureValue) pairs

$$x = \{(A, 0.15),\ (E, 0.28),\ (F, 0.77)\}$$

Feature Set Embedding for Incomplete Data – W 13

D. Grangier & I. Melvin NEC

FSE classifies sets of (FeatureId, FeatureValue) pairs



1. p embeds each pair (FeatureId, FeatureValue) into a latent space

2. Φ summarizes the latent vectors into one vector (e.g. average or max)

3. V classifies this vector

Feature Set Embedding for Incomplete Data – W 13

D. Grangier & I. Melvin NEC

Flexibility, Empirical Advantage, Beyond Missing Feat.

- flexibility: p allows
  - mixing continuous/discrete features
  - encoding prior knowledge about the features...
- outperform recent alternatives for feat. missing at
  - train & test time
  - train time only
  - test time only
- FSE allows
  - prediction before all features are computed
  - multi-instance learning
  - active feature selection

See you at Poster # 13