OPTIMUM FEATURE ORDERING FOR DYNAMIC INSTANCE–WISE
JOINT FEATURE SELECTION AND CLASSIFICATION

Yasitha Warahena Liyanage, Daphney–Stavroula Zois
Electrical and Computer Engineering Department
University at Albany, SUNY, Albany
{yliyanage, dzois}@albany.edu

Introduction

- In many real-world applications (e.g., medical diagnosis)
  - time-sensitive and interpretable decisions are needed
  - features are not freely available to acquire
- Example: doctor wants to diagnose patient
  - must diagnose [classification decision] quickly by conducting minimum number of tests [features]
  - different set of tests may be appropriate for each individual patient [data instance]
  - order by which tests are conducted [feature ordering] is important

Related Work

- Feature selection methods
  - features used are same for all instances
- Instance–wise feature selection methods
  - reveal all feature assignments and do not scale for large feature spaces
- Our prior work
  - order by which features are reviewed is fixed
- In contrast, proposed method
  - optimizes both order by which feature is reviewed and number of features per data instance
  - dynamically selects features and scales for large feature spaces

Solution

Optimization Problem

\[ J(\sigma, R), D(\sigma, R) = E \left( \sum_{t=1}^{L} \ell(F_t | \mathcal{C}) \right) = \sum_{t=1}^{L} \sum_{j=1}^{K} q_{C_j} \ell(F_t | j, C = c_j) \]

Cost of evaluating features
Misclassification cost

Optimum Classification

\[ D^*_\sigma(R) = \min_{1 \leq j \leq L} \left[ Q_j^T \pi_{\sigma(R)} \right] \]

Optimum Stopping

\[ J_{\pi_j}(\pi_j) = \min \left[ g(\pi_j), A_k(\pi_j) \right] \]

Cost of stopping
Cost of continuing

Theoretical Results

- Function \( g(\pi)\) is continuous, concave, and piecewise linear and represented by set \( \{ Q_j^T \}_{j=1}^L \) at \( L \) vectors.

Lemma

\[ \hat{A}_k(\pi_j) = \min_{F_{k+1} \in \mathcal{E}_k} \left[ F_{k+1} \right] \]

Optimal feature ordering

\[ F^*_\sigma = \left\{ F_1, F_2, \ldots, F_K \right\} \]

Problem Description

- \( F \triangleq \{ F_1, F_2, \ldots, F_K \} \) set of features
- \( C \in \{ C_1, \ldots, C_L \} \) class variable
- \( \ell(F_t) \) cost of evaluating features
- \( Q_{ij} \) misclassification cost of selecting class when \( C_j \) is true \( C_i \)

- \( \sigma(R) \) feature at which the sequential process stops [stopping feature]
- \( \sigma(R) \) feature at which the sequential process stops after reviewing the second feature \( F_2 \)

Conclusions

- Contributions
  - framework to select both order and number of features for each data instance individually
  - properties of optimum solution
- IFCO algorithm and validation of its performance on real-world datasets

- Future directions
  - extend framework to regression settings

This research is based upon work supported by the National Science Foundation (NSF) under Grants ECCS–1737443 & CNS–1942330.