Fast greedy algorithms for dictionary selection with generalized sparsity constraints

Kaito Fujii & Tasuku Soma (UTokyo)

Neural Information Processing Systems 2018, spotlight presentation
Dec. 7, 2018
Dictionary

If real-world signals consist of a few patterns, a "good" dictionary gives sparse representations of each signal.
If real-world signals consist of a few patterns, a “good” dictionary gives sparse representations of each signal patch.
If real-world signals consist of a few patterns, a "good" dictionary gives sparse representations of each signal.
If real-world signals consist of a few patterns, a “good” dictionary gives sparse representations of each signal.
Dictionary selection [Krause–Cevher’10]

Union of existing dictionaries
- DCT basis
- Haar basis
- Db4 basis
- Coiflet basis

Selected atoms as a dictionary

Atoms for each patch $y_t$ ($\forall t \in [T]$)
### Dictionary selection [Krause–Cevher’10]

**Union of existing dictionaries**

- **DCT basis**
- **Haar basis**
- **Db4 basis**
- **Coiflet basis**

**Selected atoms as a dictionary**

Atoms for each patch $\mathbf{y}_t \ (\forall t \in [T])$
Dictionary selection [Krause–Cevher’10]

Union of existing dictionaries

- DCT basis
- Haar basis
- Db4 basis
- Coiflet basis

Selected atoms as a dictionary

Atoms for each patch $y_t$ ($\forall t \in [T]$)

$\approx w_1 + w_2 + w_3$
Dictionary selection [Krause-Cevher’10]

Union of existing dictionaries

- DCT basis
- Haar basis
- Db4 basis
- Coiflet basis

Selected atoms as a dictionary

Atoms for each patch $y_t (\forall t \in [T])$

$\approx w_1 + w_2 Z_1 + w_3$
Dictionary selection with sparsity constraints

Maximize
\[
\max_{(Z_1, \ldots, Z_T) \in \mathcal{I} : Z_t \subseteq X} \sum_{t=1}^{T} f_t(Z_t) \quad \text{subject to } |X| \leq k
\]
Dictionary selection with sparsity constraints

Maximize \( \max_{X \subseteq V} \sum_{t=1}^{T} f_t(Z_t) \) subject to \( |X| \leq k \)

2nd maximization:
selecting a set \( Z_t \subseteq X \) of atoms
for a sparse representation of each patch
under sparsity constraint \( \mathcal{I} \)
Dictionary selection with sparsity constraints

Maximize \(X \subseteq V\) \[\max_{(Z_1, \ldots, Z_T) \in \mathcal{I}: Z_t \subseteq X} \sum_{t=1}^{T} f_t(Z_t)\]
subject to \(|X| \leq k\)

set function representing the quality of \(Z_t\) for patch \(y_t\)
sparsity constraint
Dictionary selection with sparsity constraints

Maximize \( X \subseteq V \)

\[
\begin{align*}
\max_{(Z_1, \ldots, Z_T) \in \mathcal{I}} \quad & \sum_{t=1}^{T} f_t(Z_t) \\
\text{s.t.} \quad & |X| \leq k
\end{align*}
\]

Our contributions

1. Replacement OMP:
   A fast greedy algorithm with approximation ratio guarantees
Dictionary selection with sparsity constraints

Maximize \[ X \subseteq V \]
subject to \[ |X| \leq k \]

\[ \max_{(Z_1, \ldots, Z_T) \in \mathcal{I}} \sum_{t=1}^{T} f_t(Z_t) \]

Our contributions

1. Replacement OMP:
   A fast greedy algorithm with approximation ratio guarantees

2. \( p \)-Replacement sparsity families:
   A novel class of sparsity constraints generalizing existing ones
Replacement Greedy for two-stage submodular maximization [Stan+’17]
Replacement Greedy for two-stage submodular maximization [Stan+’17]

1st result: application to dictionary selection

Replacement Greedy: $O(s^2dknT)$ running time
Replacement Greedy for two-stage submodular maximization [Stan+’17]

1st result
Replacement Greedy \( O(s^2dknT) \) running time

2nd result
Replacement OMP \( O((n + ds)kT) \) running time

application to dictionary selection

O(\(s^2d\)) acceleration with the concept of OMP
# Replacement OMP

| Algorithm          | Approximation Ratio | Running Time | Empirical Performance |
|--------------------|---------------------|--------------|-----------------------|
| SDS$_{MA}$ [Krause–Cevher’10] | ✓                    | ✓            | ✓                     |
| SDS$_{OMP}$ [Krause–Cevher’10] |                      | ✓            | ✓                     |
| Replacement Greedy | ✓                    | ✓            | ✓                     |
| Replacement OMP    | ✓                    | ✓            | ✓                     |
2. $p$-Replacement sparsity families

- Average sparsity [Cevher–Krause’11]
- Average sparsity w/o individual sparsity
- Block sparsity [Krause–Cevher’10]
- Individual sparsity [Krause–Cevher’10]
- Individual matroids [Stan+’17]
\section{p-Replacement sparsity families}

- (3k − 1)-replacement sparse
  - Average sparsity
    - [Cevher–Krause’11]
  - Average sparsity w/o individual sparsity
- (2k − 1)-replacement sparse
- k-replacement sparse
  - Block sparsity
    - [Krause–Cevher’10]
  - Individual sparsity
    - [Krause–Cevher’10]
  - Individual matroids
    - [Stan+’17]
We extend Replacement OMP to \( p \)-replacement sparsity families

**Theorem**

Replacement OMP achieves
\[
\frac{m_{2s}^2}{M_{s,2}^2} \left( 1 - \exp \left( - \frac{k M_{s,2}}{p m_{2s}} \right) \right) \text{-approximation}
\]
if \( \mathcal{I} \) is \( p \)-replacement sparse

**Assumption**

\[
f_t(Z_t) \triangleq \max_{w_t: \text{supp}(w_t) \subseteq Z_t} u_t(w_t)
\]

where \( u_t \) is \( m_{2s} \)-strongly concave on \( \Omega_{2s} = \{ (\mathbf{x}, \mathbf{y}) : \| \mathbf{x} - \mathbf{y} \|_0 \leq 2s \} \)

and \( M_{s,2} \)-smooth on \( \Omega_{s,2} = \{ (\mathbf{x}, \mathbf{y}) : \| \mathbf{x} \|_0 \leq s, \| \mathbf{y} \|_0 \leq s, \| \mathbf{x} - \mathbf{y} \|_0 \leq 2 \} \)
Overview

1 Replacement OMP: A fast algorithm for dictionary selection
2 $p$-Replacement sparsity families: A class of sparsity constraints

Other contributions

- Empirical comparison with dictionary learning methods
- Extensions to online dictionary selection

Poster #78 at Room 210 & 230 AB, Thu 10:45–12:45