SKA machine learning perspectives for imaging, processing and analysis

Slava Voloshynovskiy
Stochastic Information Processing Group
University of Geneva
Switzerland

with contribution of:
D. Kostadinov, S. Ferdowsi, M. Diephuis, O. Taran and T. Holotyak
Outline

- Machine learning challenges in SKA
- Challenge 1 – imaging-reconstruction
- Challenge 2 – data compression for transfer and storage
- Challenge 3 – analytics and processing of big data
- Conclusions
Remark

- Compilation of expertise from
  
  Computer Science department
  Section of mathematics
  Observatory
Machine learning realities and SKA

New perspectives of machine learning based image processing due to:

- large amount of collected observations (training data)
- new powerful computational facilities
- modern phased antenna arrays
- optimisation algorithms
Main SKA challenges

- **Challenge 1: Imaging-reconstruction**
  - Huge amount of computation for pair-wise correlations, calibration, reconstruction

- **Challenge 2: Data transfer and storage**
  - Data transfer from correlators to reconstruction servers, data centers, SDP and end users

- **Challenge 3: Analytics**
  - Automatic processing of produced data (recognition, mining, search, tracking,...)
Challenge 1: Imaging – generic approach

- Input image
- Blurred image
- Restored image
- Map

\[ y = Hx + z \]

\[ \hat{x} = \arg \max_x p(y|x)p(x) \]

Main issue:
How to model \( p(x) \) to obtain accurate, tractable and low-complexity solution?
Challenge 1: Imaging – “hand-crafted” approach

Traditional approaches to definition of $p(x)$

- **Statistical/deterministic approaches**
  - **Direct domain**
    - Smoothness of solution, local correlations.....
    
    $$\hat{x} = \arg \min_{a} \|y - Hx\|_2^2 + \lambda \Omega(x)$$
    $$\Omega(x) = -\ln p(x)$$
  - **Transform domain**
    - (decorrelation, energy compaction, directivity, ...)
    
    $$\hat{a} = \arg \min_{a} \|y - HDa\|_2^2 + \lambda \Omega(a)$$
    $$\hat{x} = Da$$
    fixed, signal independent (DCT, DWT....)
    $$\Omega(a) = -\ln p(a)$$
    i.i.d. GGD
    i.i.d. Student
    i.i.d. Mixture of Gaussians etc

- **Sparsity-based approach**
  - Overcomplete and can be learned
    
    $$\hat{a} = \arg \min_{a} \|y - H\Psi a\|_2^2 + \lambda \Omega(a) \Rightarrow \hat{x} = \Psi\hat{a}$$
    $$\Omega(a) = \|a\|_0$$

Nonconvex and NP-hard problem: relaxation/greedy approaches

Complex optimization tools:
- proximal algorithms
- prime-dual methods
- augmented Langrangian ......leading to parallel and distributed solutions
Challenge 1: Imaging – “machine learning” approach

Given: a lot of training data
Learn: statistical model $p(x)$

Various imaging configurations

Radiowaves, Microwaves, Infrared, Visible, Ultraviolet, X-Ray

Training data

+ Simulation tools
  Faraday, ASKAP, CASA...

ALMA, EVLA, LOFAR, VLBI, ..., SKA
Challenge 1: Imaging – as learning problem

**Given:** training data 
\[(y_j, x_j), 1 \leq j \leq K\]

Reconstruction algorithm = mapper 
\[\varphi(\cdot)\]

Learn a mapper \[\varphi(\cdot)\]
Challenge 1: Imaging – learning as encoding-decoding

**Given:** training data \((y_j, x_j), 1 \leq j \leq K\)
Challenge 1: Imaging – learning as encoding-decoding

**Given:** training data $(y_j, x_j), 1 \leq j \leq K$

**Encoder-decoder training**

$$(\hat{D}_e, \hat{D}_d) = \arg\min_{D_e, D_d} \sum_{j=1}^{K} \left| \left| a_j - D_e y_j \right| \left| \left| + \lambda_1 \Omega_a(a_j) + \lambda_2 \Omega_{ab}(a_j, b_j) + \lambda_3 \left| \left| x_j - D_d b_j \right| \right| \right|_2^2 \right)$$

**Encoder**

**Decoder**

**Projection problem**

$$\hat{a} = \arg\min_{a} \left| \left| a - D_e y \right| \left| \left|_2 + \gamma_1 \Omega_a(a) + \gamma_2 \Omega_{ab}(a, \hat{b}) \right)$$

$$\hat{x} = D_d \hat{b}$$

**Reconstruction problem**

$$\hat{b} = \arg\min_{b} \left| \left| y - D_d b \right| \left| \left|_2 + \lambda \Omega_{ab}(\hat{a}, b) \right)$$

$$\hat{x} = D_d \hat{b}$$
Challenge 1: Imaging – learning optimal imaging configurations

Objective: joint optimization of reconstruction and imaging (CS – random sampling)

\[
\hat{H}, \hat{D}_e, \hat{D}_d = \arg \min_{H, D_e, D_d} \sum_{j=1}^{K} \| a_j - D_e H x_j \|_2^2 + \lambda_1 \Omega_a(a_j) + \lambda_2 \Omega_{ab}(a_j, b_j) + \lambda_3 \| x_j - D_d b_j \|_2^2
\]

Constraints on number and possible geometry of synthesized arrays

Intermediate representation

Encoder

Decoder

Imaging system-encoder-decoder training

Training
Challenge 1: Imaging – learning for “adaptive” imaging

Objective: minimize the load on correlators \( \Rightarrow \) adaptive “light-weight” imaging

\[ \tilde{H}_i \] - estimation configuration (trained)

\[ y_i \]

\[ \tilde{H}_j \] - adaptive configuration

\[ \tilde{x} \]

Given input image

Image spatial spectrum

Estimation of dominant components

Reconstruction

Reconstructed image
Challenge 2: Compression for transfer, storage and distribution

- Traditional approach to compression

\[ y = Hx + z \]

- Alternative - compressive sensing \( \Rightarrow \) quantized compressive sensing

\[ y = Q(Hx + z) \]

observations are quantized (even to several bits)

\[ \hat{x} = \arg \min_{a} \|y - Q(Hx)\|_2^2 + \lambda \Omega(x) \] - inverse problem

Generic and “image independent”
Challenge 2: Compression – machine learning approach

- **Option 1:** Replace generic JPEG/JPEG2K codecs by special algorithms trained on RI images
  - We suppose that the image is already reconstructed and the problem is how to deliver it to the end users

- **Option 2:** Sparse code representation between the Encoder-Decoder
  - Encoder-decoder pairs in reconstruction are trained with the entropy constrained sparse code
  - Efficient coding based on **structured** codebooks vs **random** ones

---

![Diagram](image)
Challenge 2: Compression – information-theoretic approach

**Codebook training (like Shannon R(D) theory) from** $p(x)$

- Information-theoretic approach (limits)
- $p(x)$ is not known
- Codebook size is exponential in $n$
- Codebook is unstructured
- High compression complexity and memory
Structured codebooks with successive refinement

- $L$ trained codebooks (structured)
- Polynomial complexity
- Approaching Shannon lower bound
- Outperforming JPEG/JPEG2K for very low bit rates
Main issue: dimensionality and amount of images for automatic processing

Intermediate representation

Feature vector = compact and discriminative image representation

Supervised “hashing”
010011110010101

Automatic tools for Big Data analysis
Conclusions

- **SKA is a great tool for research**
- .... and not only in astronomy and astro-physics

- It is a “test” platform for many ideas in **machine learning based image processing** thanks to **big data** and **flexible imaging**

- In turns, it will lead to new alternative approaches to three SKA challenges covering:
  - imaging and reconstruction algorithms
  - compression algorithms
  - automatic processing of big data