Vold-driven learning over varying tasks and data-types

Al Hero

1. Bounds and tradeoffs: high dimensional scaling laws
2. Data integration: multimodal information fusion
• Value of information per sample is quantified by learning rates

• Co-PI’s on the MURI have studied high dimensional scaling laws for different types of tasks, constraints, side information, e.g.

• Insights into task-dependence of scaling laws for correlation mining
Sample complexity scaling laws

- \( n = g(p) \) = required number of samples for desired level of performance

- VoI and sample complexity
  - Question 1: what is intrinsic value of contextual info for a task?
    - How would knowledge about model reduce sample complexity?
  - Question 2: how does the VoI intrinsically depend on the task?
    - How many samples required to attain performance benchmarks?
• $d$ dimensional sample $X$
• Sparse covariance matrix: $\Sigma = \text{cov}(X), \ (p \times p)$
• $p = \binom{d}{2}$ number of unknown correlations in $\text{cov}(X)$
• $n= \text{number of available i.i.d. samples} \ \{X_1, ..., X_n\}$
The sample complexity scaling laws depend on the task.

**Regimes:** The sample complexity decreases as task becomes easier.

<table>
<thead>
<tr>
<th>Task</th>
<th>Screening</th>
<th>Detection</th>
<th>Support recovery</th>
<th>Param. estimation</th>
<th>Perform. estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk</td>
<td>$P(N_e &gt; 0)$</td>
<td>$P(N_e &gt; 0)$</td>
<td>$P{S \Delta \hat{S}} = \phi$</td>
<td>$E[|\Omega - \hat{\Omega}|^2_F]$</td>
<td>$\int E[(f_\Omega(x) - \hat{f}(x))^2]dx$</td>
</tr>
<tr>
<td>Bound</td>
<td>$1 - e^{-\kappa n}$</td>
<td>$pe^{-n\beta}$</td>
<td>$2^{p'\alpha} e^{-n\beta}$</td>
<td>$\frac{p \log p}{n} \beta$</td>
<td>$n^{-2/(1+p)} \beta$</td>
</tr>
<tr>
<td>Regimes</td>
<td>$\frac{\log p}{n} \to \infty$</td>
<td>$\frac{\log p}{n} \to \alpha$</td>
<td>$\frac{p'}{n} \to \alpha$</td>
<td>$\frac{p \log p}{n} \to \alpha$</td>
<td>$\frac{p}{\log n} \to \alpha$</td>
</tr>
<tr>
<td>Threshold</td>
<td>$\rho_c \to 1$</td>
<td>$\rho_c \to \rho^*$</td>
<td>$\rho_c \to 0$</td>
<td>$\rho_c \to 0$</td>
<td>$\rho_c \to 0$</td>
</tr>
</tbody>
</table>

Unachievable region for support recovery.
Sequence tasks in order of increasing sample complexity requirements (Firouzi, Hero, Rajaratnam, 2014)

Detection Task

Estimation Task

Uncertainty Quantification Task

Implication: value of graduated matching of tasks to sample size

Number of samples

A few samples

More samples

Lots of samples

Detect

Estimate

Quantify

Time
Sampling, prediction and adaptive regression by correlation screening (SPARCS)

Experiment: Stage 1
- p probes
- q responses
- n replicates

Experiment: Stage 2
- k probes
- q responses
- t-n replicates

Predictive Correlation Screening ($\delta=1$)

Pooled OLS predictor:
$$\arg\min_A \sum_{exp1\cup exp2} |Y^q - AX^k|^2$$

Firouzi, Rajaratnam, H 2014
SPARCS: two stage predictor design algorithm
1. Select variables from small amount of full dimensional data
2. Collect larger amount of low dimensional data with selected variables

Extension of distilled sensing and sure independence screening

<table>
<thead>
<tr>
<th>Symptom</th>
<th>RMSE: LASSO</th>
<th>RMSE: SIS</th>
<th>RMSE: SPARC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Runny Nose</td>
<td>0.7182</td>
<td>0.6896</td>
<td>0.6559</td>
</tr>
<tr>
<td>Stuffy Nose</td>
<td>0.9242</td>
<td>0.7787</td>
<td>0.8383</td>
</tr>
<tr>
<td>Sneezing</td>
<td>0.7453</td>
<td>0.6201</td>
<td>0.6037</td>
</tr>
<tr>
<td>Sore Throat</td>
<td>0.8235</td>
<td>0.7202</td>
<td>0.5965</td>
</tr>
<tr>
<td>Earache</td>
<td>0.2896</td>
<td>0.3226</td>
<td>0.3226</td>
</tr>
<tr>
<td>Malaise</td>
<td>1.0000</td>
<td>0.7563</td>
<td>0.9125</td>
</tr>
<tr>
<td>Cough</td>
<td>0.5879</td>
<td>0.7505</td>
<td>0.5564</td>
</tr>
<tr>
<td>Shortness of Breath</td>
<td>0.4361</td>
<td>0.5206</td>
<td>0.4022</td>
</tr>
<tr>
<td>Headache</td>
<td>0.7896</td>
<td>0.7500</td>
<td>0.6671</td>
</tr>
<tr>
<td>Myalgia</td>
<td>0.6372</td>
<td>0.5539</td>
<td>0.4610</td>
</tr>
<tr>
<td>Average for all symptoms</td>
<td>0.6953</td>
<td>0.6463</td>
<td>0.6016</td>
</tr>
</tbody>
</table>

Firouzi, Rajaratnam, H 2014
Data integration: Multimodal information fusion

• Multimodality image fusion aims to extract high value information from different data types

• Several MURI co-Pis have studied multimodality fusion

• Here: a multimodal factor analysis method for dichotomous data types: continuous, discrete, event-based
• Factor analysis is a matrix factorization $X = MA$
• Each column of the matrix $X = [X(1), \ldots, X(N)]$ obeys model

$$X(t) = Ma(t) + W(t) = \sum_{k=1}^{p} m_k a_k(t) + W(t)$$

$m_k$ are factor loadings and $a_k(t)$ are factor scores
• Factor analysis was introduced by Thurstone (1931) to
  – Reduce dimension of a high dimensional data matrix
  – Cluster the rows and columns of the data matrix
• Standard factor analysis is inapplicable when rows of $X$ come from mixed alphabet datatypes (modalities)
  – Continuous, discrete, mixed cts-discrete, categorical, symbolic,…
• Such cases frequently arise in data-rich applications
  – Combined sensor data, contextual annotations, social media, event tags,…
Multimodal factor analysis

Previous work

- Fisher et al. 02, audio-video fusion using mutual information
- Barnard et al. 03, joint processing of images and annotation
- Wu et al. 04, modality selection and fusion for image and video
- Khan et al. 10, factor analysis for Gaussian and multinomial data
- Sui et al. 12, multimodal fusion of brain imaging data
- Otto 13, fusion of data from heterogeneous sensors for advanced driver assistance systems
- Salazar et al. 15, Bayesian multimodality analysis for mental health

In [1] we propose a multimodal factor analysis method that integrates

- Continuous data, e.g., Gaussian, VonMisesFisher or Watson
- Discrete data, e.g. multinomial
- Event-based data, e.g., Poisson

Multimodal factor analysis model

\[ y_{in} | x_n \sim \text{Pois}(e^{c_i^T x_n}), \]
\[ x_n \sim \mathcal{N}(\zeta, R), \]
\[ z_{im} | w_m \sim \mathcal{N}(c_i^T w_m, \sigma_i^2), \]
\[ w_m \sim \mathcal{N}(\alpha, S), \]
\[ h_{id} | \{v_d\} \sim \text{Mult} \left( L_i; \frac{e^{c_i^T v_1}}{\sum_{d=1}^{D} e^{c_i^T v_d}}, \ldots, \frac{e^{c_i^T v_D}}{\sum_{d=1}^{D} e^{c_i^T v_d}} \right), \]
\[ v_d \sim \mathcal{N}(\beta, Q), \]
\[ \theta = \{c_i, \zeta, R, \sigma_i^2, \alpha, S, \beta, Q\} : \text{parameters} \]
Twitter Dataset from August 2014

• ~1M Tweets with
  – ~10K Hashtags
  – Geolocation data (Latitude and Longitude)
  – Count of Hashtags over 744 hours
  – Bag of Words representation of Hashtags

• Multimodal Factor Analysis [1]
  • For each Hashtag: Factor Loading Matrix mixes von Mises-Fisher, Poisson, and Multinomial distributions in their natural parameters
  • EM Algorithm to find the Factor Loading coefficients for Hashtags, and Factor Scores for Distributions
  • Factors localize both in terms of Geolocation and Words correspond to generative events for Hashtags used for dimension reduction, clustering, ...

Some of the top factors found

Loved robinwilliams
Suicide Peace
Depression

#RIP
#hook
#RIPobinwilliams
#Robin
#RIPobin

crime
dead
uk usainbott
Glasgow

#CommonwealthGames
#Glasgow2014

Law
Divine
Bandhan
jail
justice

#WhyMediastAntHindu
#VedicaRakshaBanchanWithBapuji
#PoliceMisuSeOfPOCSOinBapujiCase
#UnfairProbeByJodhpurPolice
#365DaysOfPOCSOmisuse

challenge
donate
nominate
ice
bucket

#alsicebucketchallenge
#ASILIceBucketChallenge

Police
white black
cops
Protest

#Ferguson
#MikeBrown
#medablackout

ice
accepted
donate
waters
als

#IceBucketChallenge

Paul
George
USA
heart
Sorry

#USABasketball
#paulgeorge
#PrayForPG
#PrayForPaulGeorge
Geographic Distribution of Factors

Hindu Religious Leader in Jail

* Indian Locations
△ Indian Center
Geographic Distribution of Factors

Paul George Injury

- Indian Locations
- Indian Center
- Commonwealth Locations
- Commonwealth Center
- PaulGeorge Locations
- PaulGeorge Center
Geographic Distribution of Factors

Ferguson Unrest

- Indian Locations
- Indian Center
- Commonwealth Locations
- Commonwealth Center
- PaulGeorge Locations
- PaulGeorge Center
- Ferguson Locations
- Ferguson Center
Geographic Distribution of Factors

Global Ice Bucket Challenge

- Indian Locations
- Indian Center
- Commonwealth Locations
- Commonwealth Center
- PaulGeorge Locations
- PaulGeorge Center
- Ferguson Locations
- Ferguson Center
- RobinWilliams Locations
- RobinWilliams Center
- Local IceBucket Locations
- Local Ice Bucket Center
- Local Ice Bucket True Center
- Global IceBucket Locations
- Global Ice Bucket Center
- Global Ice Bucket True Center
Efficient Clustering
Value of Geolocation Information

![Graph showing the value of location information](image-url)

- **x-axis**: Weight of Location Relative to Words (in computing Factor Loadings)
- **y-axis**: Adjusted Rand Index

The graph illustrates the relationship between the weight of location information and its adjusted rand index. The peak in the graph suggests an optimal weight, indicating the maximum value of geolocation information in this context.
Conclusions

• Accomplishments
  – Task dependent scaling laws for correlation mining
  – Multimodal factor analysis developed and applied to social media

• Advantages and benefits
  – Task dependent scaling laws quantify VoI/sample and suggest graduated information exploitation
  – Multimodal factor analysis unifies dichotomous data types

• Work done in collaboration with
  – Bala Rajaratnam, Stanford University
  – Brian Sadler, Nasser Nasrabadi, Lance Kaplan, Ed Zelnio ARL