Information Planning and Active Data Collection

John Fisher

Sensing, Learning, & Inference Group
Computer Science & Artificial Intelligence Laboratory
Massachusetts Institute of Technology
http://groups.csail.mit.edu/vision/sli/

November 20, 2014
Efficient Information Planning

Tractable greedy selection achieves near-optimal performance.

Williams et al. [2007b] reduces complexity of information gathering formulated as a Markov Decision Process.

\[ O(\left[N_s 2^{N_s}\right]^N M^N) \rightarrow O(NN_s^3) \]

Williams et al. [2007a] the optimal information gathering rate is no greater than twice the greedy information gathering rate.

\[ \frac{I(X;Z^G_N)}{I(X;Z^{*}_N)} \geq \frac{1}{2} \quad \forall N \]

\( N_s \): number of sensing actions, \( N \): planning horizon, \( M \): measurement simulation cost.
Sequential Information Planning

- $N_t$ measurements for each $X_t$, $\mathcal{V}_t = \{1, \ldots, N_t\}$, $t \in \{1, \ldots, T\}$.
- Goal

$$\mathcal{O} \in \arg \max_{|\mathcal{I}_1| \leq k_1, \ldots, |\mathcal{I}_T| \leq k_T} f(\mathcal{S})$$

$$\mathcal{O} \left( \binom{N}{k}^T \right) \leq \mathcal{O}(N^{kT}) \text{ when } N_t = N, k_t = k \forall t$$
On the Oracle Assumption

Most of the guarantees on greedy selection assume an oracle model, i.e., the complexity of reward evaluation has constant time [Nemhauser et al., 1978], [Guestrin et al., 2005], [Krause et al., 2005], [Kempe et al., 2003], [Calinescu et al., 2007], [Streeter et al., 2009].

This generally does not hold, particularly for sequential problems (aka almost all problems of interest).

Here we show

- One can exploit sparsity in the latent variable structure and selection order reduce complexity.
- The same reasoning leads to an efficient incremental approach inference in trees and poly-trees (with extensions to loopy graphs utilizing feedback vertex set graph decompositions).
Gaussian HMMs

Consider the Gaussian HMM governed by the dynamics:

\[ X_t = A_{t-1} X_{t-1} + V_{t-1} \]
\[ Y_t = C_t X_t + W_t, \]

where \( A_t, C_t \) highly sparse.

---

<table>
<thead>
<tr>
<th>Operation</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Propagation</td>
<td>( O(d^3) )</td>
</tr>
<tr>
<td>Update</td>
<td>( O(md^2) )</td>
</tr>
</tbody>
</table>

- \( d \): dimension of \( X_t \)
- \( T \): number of hidden variables
- \( N_t \): number of observations per set
- \( m \): dimension of each observation \( Y_{t,u} \)
- \( k_t \): number of constraints per set (\( d, N_t \gg m, k_t \))
Exploiting Sparsity

- **Update** - sparsity combined with the information form yields efficient updates (with some bookkeeping).

\[
J_{w_j|u \cup \mathcal{G}_{j-1}} = J_{w_j|\mathcal{G}_{j-1}} + C_{w_j,u}^T R_{w_j,u}^{-1} C_{w_j,u}
\]

\[
= \begin{bmatrix}
J_{w_j|\mathcal{G}_{j-1}}(I_u, I_u) & J_{w_j|\mathcal{G}_{j-1}}(I_u, -I_u) \\
J_{w_j|\mathcal{G}_{j-1}}(-I_u, I_u) & J_{w_j|\mathcal{G}_{j-1}}(-I_u, -I_u)
\end{bmatrix} + 
\begin{bmatrix}
C_{w_j,u}(I_u)^T R_{w_j,u}^{-1} C_{w_j,u}(I_u) & 0_{q \times (d-q)} \\
0_{(d-q) \times q} & 0_{(d-q) \times (d-q)}
\end{bmatrix},
\]

where \( I_u, |I_u| = q \ll d \).

- **Exploration.** Choose \( g_j \) as

\[
J_{w_j|u \cup \mathcal{G}_{j-1}} = J_{w_j|\mathcal{G}_{j-1}} + C_{w_j,u}^T R_{w_j,u}^{-1} C_{w_j,u} = J_{w_j|\mathcal{G}_{j-1}} + 
\begin{bmatrix}
\hat{C}_{w_j,u}^T \\
0_{(d-q) \times m}
\end{bmatrix} 
\begin{bmatrix}
\hat{C}_{w_j,u} & 0_{m \times (d-q)}
\end{bmatrix},
\]

where \( \hat{C}_{w_j,u} = R_{w_j,u}^{-1/2} C_{w_j,u}(I_u) \).

---

<table>
<thead>
<tr>
<th>Operation</th>
<th>Speedup/measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Update</strong></td>
<td>( \Theta \left( \frac{d^2}{\max{m,q}^2} \right) )</td>
</tr>
<tr>
<td><strong>Exploration</strong></td>
<td>( \Theta \left( \frac{d^3}{m \max{m,q}^2} \right) )</td>
</tr>
</tbody>
</table>
Experiment

Two hundred moving objects with different degrees of correlation.

\[ X_t = A_{t-1} X_{t-1} + V_{t-1}, \forall t \in \{1, \ldots, 20\} \]
\[ Y_t = C_t X_t + W_t, \]

where \( X_t = \begin{bmatrix} p_{t,(x,y,z)}^{1:200} & v_{t,(x,y,z)}^{1:200} \end{bmatrix}^T \),
\( V_{t-1} \sim \mathcal{N}(0, Q_{t-1}) \) driving and
\( W_t \sim \mathcal{N}(0, R_t) \) measurement noise. We of the hidden dimension and different degrees of sparsity in the measurement model.

Speedup grows with observation size and sparsity.

A measurement is available for each latent variable (position, velocity). We have \( N^{\text{max}} = 1200 \) measurements in total, while we consider different observation sizes, \{10\%, 25\%, 50\%, 75\%, 100\%\} of \( N^{\text{max}} \). We select \( k_t = 6 \) measurements from each set.
Multi-Modal Fusion

3D reconstruction using multiple data sources

Multiple Data Sources
- Full motion video (FMV)
- FMV Platform GPS/INS
- LIDAR
- Open Street Map GPS
- Open Street Map Waypoints

All of these provide complementary information about the scene.

Q. How do we combine them in a coherent way?
Multi-Modal Fusion

3D reconstruction using multiple data sources

- **Challenge:** We need a model for integration.
- **Approach:** Formulate as inference in a graphical model.
- **Reality:** Handles some aspects really well, others require new algorithmic and theoretical developments.

- Data Integration versus Queries are separated in such models.
- Uncertainty is explicitly represented.
Multi-Modal Fusion

*3D reconstruction using multiple data sources*
Multi-Modal Fusion

*3D reconstruction using multiple data sources*

- **However**, construction of such a model is often merely an intermediate step to more complex reasoning.
Multi-Modal Fusion

*3D reconstruction using multiple data sources*
Multi-Modal Fusion

3D reconstruction using multiple data sources

- We can reason over additional content of the scene.
Multi-Modal Fusion

3D reconstruction using multiple data sources

Mensuration: we can measure physical dimensions.
Multi-Modal Fusion

*3D reconstruction using multiple data sources*

- **Mensuration**: we can measure physical dimensions.
- In short, this is an intermediate step to higher-level reasoning, *i.e.*, asking questions about the scene.
Adding Contextual Variables

- Context is a loaded term.
- In this setting, it represents learned local and global priors on appearance and geometry.
- Can be shown to reduce the Vol of measurements (this is expected).

Questions:
- Is there context to exploit?
- Can we learn it from data?
- Can integrate Vol analysis that trades off measurement Vol versus contextual Vol?
Categories of Surface Normals
Categories of Appearance

Original

Cluster Mean

Labels (K=5)


**Outgrowth of Supervised Student Research**