Proteins, Particles, and Pseudo-Max-Marginals: A Submodular Approach

Jason Pacheco
Erik Sudderth

Department of Computer Science
Brown University, Providence RI
Protein Side Chain Prediction

Estimate side chains from backbone.

\[ p(x) \propto \prod_{s \in V} \psi_s(x_s) \prod_{(s,t) \in E} \psi_{st}(x_s, x_t) \]

1D – 4D Continuous state.

[ Image: Harder et al., BMC Informatics 2010 ]
Reweighted Max-Product (RMP)

Message passing on discrete side chains

Max-marginal: \( \mu_s(x_s) \propto \max_{\{x': x'_s = x_s\}} p(x') \)

Pseudo-max-marginal: \( \nu_s(x_s) \propto \psi_s(x_s) \prod_{u \in \Gamma(s)} m_{us}(x_s) \rho_{us} \)

Edge Appearance Probability
Rotamer discretization

Fit to side chain marginal statistics

Fails to capture side chain placement…

Penicillin Acylase Complex, Trp154 [Shapovalov & Dunbrack 2007]
Particle Max-Product (PMP)

Latent space is continuous...

...particle approximation of continuous RMP messages.
Sample new particles from proposals:
( Random Walk, Likelihood, Neighbor, … )
Particle Max-Product (PMP)

1. **Augment Particles**

2. **RMP Update**

Update RMP messages on augmented particles.

$$m_{ts}(x_s) = \max_{x_t} \psi_t(x_t) \psi_{st}(x_s, x_t) \frac{1}{\rho_{st}} \prod_{u \in \Gamma(t) \backslash s} m_{ut}(x_t)^{\rho_{ut}} m_{st}(x_t)^{1-\rho_{st}}$$

**Edge Appearance Probability**
Particle Max-Product (PMP)

Select subset of good particles...

...Need particle selection method.
Greedy PMP (G-PMP)

Select **best** particle, sample from random walk Gaussian. [Trinh '09, Peng '11]

Naïve proposals do not exploit model.
Top-N PMP (T-PMP)

Select **N-best** particles ranked by pseudo-max-marginal values.  
[Besse ’12, Pacheco ‘14]

Particles collapse to single solution.
Diverse PMP (D-PMP)

Select particles to preserve messages.

- Encourages particle diversity
- Robust to initialization
Diverse Particle Selection

At node \( t \in V \) select particles to minimize maximum outgoing message error:

\[
\min_z \max_{s \in \Gamma(t), a} m_{ts}(a) - \hat{m}_{ts}(a; z)
\]

Binary Selection Vector

RMP message over subset:

\[
\hat{m}_{ts}(a; z) = \max_b z(b) \psi_t(b) \psi_{st}(a, b) \frac{1}{\rho_{st}} \prod_{u \in \Gamma(t) \setminus s} \frac{m_{ut}(b)^{\rho_{ut}}}{m_{st}(b)^{1-\rho_{st}}}
\]

Approximate IP with greedy algorithm.
Diverse Particle Selection

Pacheco et al. ICML 2014

- Good empirical results
- Difficult to analyze
- Limited to tree-structured MRFs
Diverse Particle Selection

Equivalent to minimizing $L_\infty$ norm.

Consider other norms, e.g. $L_1$:

$$\text{minimize} \sum_{z \in \Gamma(t)} \| m_{ts} - \hat{m}_{ts}(z) \|_1$$

subject to $\| z \|_1 \leq N, \ z \in \{0, 1\}^{\alpha N}$

Easier to analyze...

**Property 1:** Message error upper bounds pseudo-max-marginal error:

$$\| \nu_s - \hat{\nu}_s \|_1 \leq \sum_{t \in \Gamma(s)} \| m_{ts} - \hat{m}_{ts} \|_{1}^{\rho_{ts}}$$
Property 2: Selection IP equivalent to submodular maximization.

Set function \( f : 2^Z \rightarrow \mathbb{R} \) is submodular iff diminishing marginal gains.

\[
\frac{f(Y \cup \{e\}) - f(Y)}{f(X \cup \{e\}) - f(X)} \geq 0
\]

P. 3: Efficient \textsc{LazyGreedy} selection within 
\((1 - \frac{1}{e})\) factor of optimal value.
Selection Objective:

\[
\text{minimize} \sum_{s \in \Gamma(t)} \| m_{ts} - \hat{m}_{ts}(z) \|_1
\]
LAZYGREEDY Selection

Selection Objective:

\[
\text{minimize } \sum_{z} \left\| m_{ts} - \hat{m}_{ts}(z) \right\|_1
\]

Joint Probability

Message

Margin

Source Node (degrees)

Destination Node (degrees)
LAZYGREEDY Selection

Selection Objective:

\[
\text{minimize} \sum_{s \in \Gamma(t)} \left\| m_{ts} - \hat{m}_{ts}(z) \right\|_1
\]
Selection Objective:

\[
\minimize_{z} \sum_{s \in \Gamma(t)} \| m_{ts} - \hat{m}_{ts}(z) \|_1
\]
LAZYGREEDY Selection

Selection Objective:

\[
\text{minimize} \quad \sum_{s \in \Gamma(t)} \| m_{ts} - \hat{m}_{ts}(z) \|_1
\]
LAZYGREEDY Selection

Selection Objective:

\[
\minimize \sum_{s \in \Gamma(t)} \left\| m_{ts} - \hat{m}_{ts}(z) \right\|_1
\]
Protein Side Chain Prediction

Pairwise Markov random field (MRF):

\[ p(x) \propto \prod_{s \in \mathcal{N}} \psi_s(x_s) \prod_{(s,t) \in \mathcal{E}} \psi_{st}(x_s, x_t) \]

- Estimate side chain for fixed backbone
- 1D to 4D continuous states.

[ Image: Harder et al., BMC Informatics 2010 ]
Protein Side Chain Prediction

T-PMP

Ground Truth
Protein Side Chain Prediction
Protein Side Chain Prediction

Log-probability of MAP estimate for...

20 Proteins (11 Runs)

370 Proteins

G-PMP, T-PMP, D-PMP $L_1$, D-PMP $L_\infty$

Rosetta simulated annealing [Rohl et al., 2004]
Protein Side Chain Prediction

Root mean square deviation (RMSD) from x-ray structure.

Oracle selects best configuration in current particle set.
Optical Flow

Estimate 2D motion for every superpixel.

Middlebury optical flow benchmark
[Baker et al. 2011]
Optical Flow

Estimate 2D motion for every superpixel.

Middlebury optical flow benchmark
[Baker et al. 2011]
Optical Flow

Estimate 2D motion for every superpixel.

Middlebury optical flow benchmark
[Baker et al. 2011]
Optical Flow

Flow ambiguity near object boundaries…

D-PMP Particles

D-PMP Estimate

D-PMP particles reflect this.
Optical Flow

D-PMP accuracy equivalent to Classic-C

[Sun et al. 2014]
Summary

General purpose particle-based max-product for continuous graphical models with cycles.

Code Available: cs.brown.edu/~pachecoj
Minimize the sum of errors ($L_1$):

$$\sum_{s \in \Gamma(t)} \| m_{ts} - \hat{m}_{ts}(z) \|_1$$

subject to $\| z \|_1 \leq N, \quad z \in \{0, 1\}^{\alpha N}$

Easier to analyze than $L_\infty$ selection...

**P1:** Message error upper bounds pseudo-max-marginal error:

$$\| \nu_s - \hat{\nu}_s \|_1 \leq \sum_{t \in \Gamma(s)} \| m_{ts} - \hat{m}_{ts} \|_1^{\rho_{ts}}$$
A function \( f : 2^Z \rightarrow \mathbb{R} \) is submodular iff diminishing marginal gains:

\[
f(Y \cup \{e\}) - f(Y) \geq f(X \cup \{e\}) - f(X)
\]

- Diverse particle selection is submodular maximization with cardinality constraint
- Efficient greedy approximation algorithm
Resolving Ties

Particle diversity leads to more conflicts:

**Side Chain Particles**

- T-PMP
- D-PMP
Submodular Particle Selection

**Property 1:** Message reconstruction error bounds pseudo-max-marginal error:

\[
\|\nu_s - \hat{\nu}_s\|_1 \leq \sum_{t \in \Gamma(s)} \|m_{ts} - \hat{m}_{ts}\|_1^{\rho_{ts}}
\]

**Property 2:** IP is equivalent to submodular maximization subject to cardinality constraints.
Submodular Particle Selection

Select particles to minimize sum of errors:

\[
\text{minimize} \sum_{z \in \Gamma(t)} \| m_{ts} - \hat{m}_{ts}(z) \|_1 \\
\text{subject to} \| z \|_1 \leq N, \quad z \in \{0, 1\}^{\alpha N}
\]

Good empirical results and we can analyze!

**Property 1**: Message error bounds pseudo-max-marginal:

\[
\| v_s - \hat{v}_s \|_1 \leq \sum_{t \in \Gamma(s)} \| m_{ts} - \hat{m}_{ts} \|_1^{\rho_{ts}}
\]

**Property 2**: Equivalent to submodular maximization subject to cardinality constraints
Reweighted Max-Product (RMP)

Message passing on discrete side chain states.

But latent space is continuous...

RMP Messages:

\[ m_{ts}(x_s) = \max_{x_t} \psi_t(x_x) \psi_{st}(x_s, x_t) \frac{1}{\rho_{st}} \prod_{u \in \Gamma(t) \backslash s} m_{ut}(x_t)^{\rho_{st}} m_{st}(x_t)^{1-\rho_{st}} \]

Edge Appearance Probability

Pseudo-max-marginal:

\[ \nu_s(x_s) \propto \psi_s(x_s) \prod_{u \in \Gamma(s)} m_{us}(x_s)^{\rho_{us}} \]
Optical Flow

Estimate motion vector for every pixel.

Diverse Particle Selection (D-PMP)
Diverse Particle Selection

Minimize maximum message error ($L_\infty$):

$$\min z \sum_{s \in \Gamma(t)} \left\| m_{ts} - \hat{m}_{ts}(z) \right\|_\infty$$

subject to $\|z\|_1 \leq N$, $z \in \{0, 1\}^{\alpha N}$

Augmented Messages $\rightarrow$ Subset Messages $\leftarrow$ Selection Vector

Pose Estimation

[ Pacheco et al., ICML 2014 ]

- Good empirical results
- No analysis/guarantees
- Limited to tree MRFs