Fourier-Information Duality in the Identity Management Problem

Xiaoye Jiang[1], Jonathan Huang[2], Leonidas Guibas[1]

[1]Stanford University, California, USA
[2]Carnegie Mellon University, Pennsylvania, USA

2011 European Conference on Machine Learning

Sep 7, 2011

Problem in Identity Management [Shin et. al. 2003]

Problem in Identity Management

Reasoning over Permutation Group

Existing approaches model uncertainty in identity management with distributions over all permutations.

- Permutation Group: All bijective mappings from $\{1, 2, \ldots, n\}$ to itself.
- There are $n!$ permutations.

$n = 5$	120
$n = 10$	$3,628,800$
$n = 20$	2.432×10^{18}

- Two dueling representations: Fourier approach and Information approach.

Contributions

- Identity a duality relationship between Fourier approach and Information approach.
- Explore the problem of converting between two representations.
- Propose a hybrid approach.

Fourier Approach

Fourier approach works by collapsing a distribution over permutations to low order marginals [Kondor et. al. 2007, Huang et. al. 2007].

Example

- We can summarize a distribution using a matrix of first order marginals.
- Requires storing only $\mathcal{O}(n^2)$ numbers.

[A,B,C]	$P(\sigma)$
$[1, 2, 3]$	$1/6$
$[1, 3, 2]$	$1/12$
$[2, 1, 3]$	$1/12$
$[2, 3, 1]$	$1/4$
$[3, 1, 2]$	$1/6$
$[3, 2, 1]$	$1/4$

	A	B	C
1	$1/4$	$1/4$	$1/2$
2	$1/3$	$5/12$	$1/4$
3	$5/12$	$1/3$	$1/4$

Fourier Approach — High Order Generalizations

- First order marginals can not capture high order dependencies.
- We can summarize a distribution using a matrix of second order marginals.
- Requires storing only $\mathcal{O}(n^4)$ numbers.

Example

[A,B,C,D]	$P(\sigma)$
$[1, 2, 3, 4]$	.01
$[1, 2, 4, 3]$	.02
$[1, 3, 2, 4]$	.01
$[1, 3, 4, 2]$	.015
$[1, 4, 2, 3]$	.005
$[1, 4, 3, 2]$	.005
$\vdots$	$\vdots$

	(A,B)	(A,C)	(A,D)	(B,C)	$\cdots$
(1,2)	.03	.025	.01	.03	$\cdots$
(1,3)	.02	.015	.03	.07	$\cdots$
(1,4)	.045	.01	.035	.02	$\cdots$
(2,3)	.015	.03	.02	.04	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Information Approach

- Parametrize a distribution over permutations using an exponential family [Schumitsch et. al. 2005].
- First order information coefficients requires storing $\mathcal{O}(n^2)$ numbers.

Example

	A	B	C
1	-.1	.3	-.2
2	.3	.2	.5
3	.1	.1	.4

$$P([A, B, C] \to [1, 3, 2]) \propto \exp(-.1 + .1 + .5)$$

Information Approach — High Order Generalizations

- We can parametrize a distribution using a second order information matrix.
- Requires storing only $\mathcal{O}(n^4)$ numbers.

Example

	(A,B)	(A,C)	(A,D)	(B,C)	$\cdots$
(1,2)	-.03	.025	-.01	.03	$\cdots$
(1,3)	.02	.015	.03	.07	$\cdots$
(1,4)	.045	-.01	.035	-.02	$\cdots$
(2,3)	-.015	-.03	.02	.04	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

$$P([A,B,C,D] \to [1,2,3,4]) \propto \exp(-.03 + .015 + .035 + .04 + \cdots)$$

Two Forms of Representation

We have two representation forms.

- Fourier: linear parameterization.
- Information: exponential family.

How do the two representation forms fit into the operations required to update the distribution over permutations?

- Accuracy
- Complexity

Markov Model for Identity Management

$$
\begin{array}{ccccccccc}
\sigma_0 & \xrightarrow{\mathrm{M}_0} & \sigma_1 & \xrightarrow{\mathrm{M}_1} & \sigma_2 & \xrightarrow{\mathrm{M}_2} & \sigma_3 & \xrightarrow{\mathrm{M}_3} & \cdots \\
\Big\downarrow{\scriptstyle L(z_0|\sigma_0)} & & \Big\downarrow{\scriptstyle L(z_1|\sigma_1)} & & \Big\downarrow{\scriptstyle L(z_2|\sigma_2)} & & \Big\downarrow{\scriptstyle L(z_3|\sigma_3)} & & \\
z_0 & & z_1 & & z_2 & & z_3 & &
\end{array}
$$

σ: true state; z: observations; M: Markov matrix; $L(z|\sigma)$: likelihood function.

- Mixing Model: tracks swapped identities with some probability.
- Observation Model: identity on a particular track is observed.
- Our Problem: Find posterior over associations between identities with tracks conditioned on all past observations.

Mixing Model

- A probability distribution m characterizing mixing of tracks induces a Markov process on associations between identities and tracks.

$$h(\sigma) \leftarrow \sum_{\tau} m(\tau) h(\tau^{-1}\sigma)$$

- Suppose m is a distribution on permutation group S_n, then the simplest mixing model is

$$m(\tau) = \begin{cases} p & \tau = \text{id} \\ 1-p & \tau = (i,j) \\ 0 & \text{otherwise} \end{cases}$$

Mixing Model

Example

Suppose A, B, C are located at tracks 1, 2, 3, when a mixing event happen between tracks 1 and 2, then the prior distribution h and mixing distribution m are

$$h(\sigma) = \begin{cases} 1 & \sigma = \text{id} \\ 0 & \text{otherwise} \end{cases} \qquad m(\tau) = \begin{cases} .5 & \tau = \text{id} \\ .5 & \tau = (1,2) \\ 0 & \text{otherwise} \end{cases}$$

then after the mixing, the distribution over permutations becomes

$$h(\sigma) = \begin{cases} .5 & \sigma = \text{id} \\ .5 & \sigma = (1,2) \\ 0 & \text{otherwise} \end{cases}$$

Mixing Model

Example

Suppose A, B, C are located at tracks 1, 2, 3, when a mixing event happen between tracks 1 and 2, then the first order marginals for h and m are

$$H = \left[\begin{array}{c|ccc} & A & B & C \\ \hline 1 & 1 & 0 & 0 \\ 2 & 0 & 1 & 0 \\ 3 & 0 & 0 & 1 \end{array}\right], \qquad M = \left[\begin{array}{c|ccc} & 1 & 2 & 3 \\ \hline 1 & .5 & .5 & 0 \\ 2 & .5 & .5 & 0 \\ 3 & 0 & 0 & 1 \end{array}\right]$$

after the mixing, the first order marginal distribution over permutations are the matrix product of M and H.

- However, such a property does NOT hold for the *information form* representation. Generally, updating *information matrices* is NOT nearly as easy.

Mixing Model — Fourier and Information Form

Proposition (Convolution theorem)

Let $M^{(t)}$, $H^{(t)}$ be the first order marginal matrices for the mixing distribution $m^{(t)}$ and $h(\sigma^{(t)}|z^{(1)},\ldots,z^{(t)})$. Then the marginal matrix for $h(\sigma^{(t+1)}|z^{(1)},\ldots,z^{(t)})$ is:

$$H^{(t+1)} = M^{(t)} \cdot H^{(t)}.$$

Proposition

Let $\Omega^{(t)}$ be the first order information matrix for $h(\sigma^{(t)}|z^{(1)},\ldots,z^{(t)})$. We need to use a second order information matrix to parameterize $h(\sigma^{(t+1)}|z^{(1)},\ldots,z^{(t)})$ after a mixing event.

Observation Model

- A typical observation says that "Observing Red on Track 1".
- We look at the color histograms of each identity, suppose identities A, B, C has 70%, 40%, 60% of red color respectively, then

$$L(\sigma) = \begin{cases} .7 & \text{if } \sigma(A) = 1 \\ .4 & \text{if } \sigma(B) = 1 \\ .6 & \text{if } \sigma(C) = 1 \end{cases}$$

- $$h(\sigma) \leftarrow \frac{1}{Z} L(\sigma) \cdot h(\sigma)$$

  where the normalizing constant $Z = \sum_{\sigma} L(\sigma) h(\sigma)$.

Observation Model — Fourier and Information Form

Proposition (Kronecker conditioning)

Let $H^{(t+1)}$ be the first order marginal matrix for the distribution $h(\sigma^{(t+1)}|z^{(1)},\ldots,z^{(t)})$, We need to use a second order marginal matrix to parameterize $h(\sigma^{(t+1)}|z^{(1)},\ldots,z^{(t+1)})$ after an observation event.

Proposition (Schumitsch et al.)

If $h(\sigma^{(t+1)}|z^{(1)},\ldots,z^{(t)}) \propto \exp\left(Tr(\Omega^T M_\sigma)\right)$, then we can update

$$\Omega_{jk} \leftarrow \Omega_{jk} + \log \alpha_{jk},$$

for a particular observation on track j.

Normalization and Maximization

Proposition

Computing the normalization constant of the information form parameterization is $\#P$-complete; while it is a trivial operation in the Fourier domain.

Proposition

Computing the permutation which is assigned the maximum probability under h reduces to the same "maximal matching" problem for both the Fourier and information forms due to the fact that the exponential is a monotonic function.

Comparison of the Two Forms

Inference Operation	**Fourier (First Order)**		**Information Form (First Order)**	
	Accuracy	Complexity	Accuracy	Complexity
Prediction/Rollup	Exact	$\mathcal{O}(n)$	Approximate	$\mathcal{O}(n)$
Conditioning	Approximate	$\mathcal{O}(n^3)$	Exact	$\mathcal{O}(n)$
Normalization	Exact	$\mathcal{O}(n^2)$	Approximate	$\mathcal{O}(n^4 \log n)$
Maximization	Exact	$\mathcal{O}(n^3)$	Exact	$\mathcal{O}(n^3)$

Both Forms are Low-Dimensional Projections

- The Fourier transform is linear and orthogonal. The Fourier approximation is an ℓ_2 projection onto a low-frequency Fourier subspace.
- The information form representation is an information projection onto the same low-frequency Fourier subspace using the KL-divergence metric.

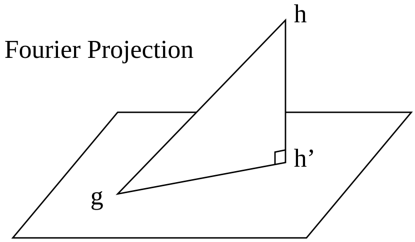
$$(IP) \quad \min_q \sum_\sigma q(\sigma) \log \frac{q(\sigma)}{h(\sigma)} \quad \text{s.t.} \quad \sum_\sigma q(\sigma) M_\sigma = Q, \quad q(\sigma) \geq 0, \forall \sigma$$

$$(ME) \quad \min_q \sum_\sigma q(\sigma) \log q(\sigma) \quad \text{s.t.} \quad \sum_\sigma q(\sigma) M_\sigma = Q, \quad q(\sigma) \geq 0, \forall \sigma$$

Pythagorean Theorem

Proposition

In both the Fourier and information domains, the Pythagorean theorem holds. If g is any function that satisfies the marginal constraints, then $D(g||h) = D(g||h') + D(h'||h)$, where h' is the projection of h in the sense of ℓ_2 or KL-divergence.



Hybrid Approach

- Hybrid approach: switch between two domains.
- Given the information coefficients Ω, we can compute the first order marginals H_{jk}, by conditioning on $\sigma(k) = j$, then normalizing.

$$H_{jk} = \sum_{\sigma:\sigma(k)=j} h(\sigma) = \frac{\exp(\Omega_{jk})\text{perm}(\exp(\hat{\Omega}_{jk}))}{\text{perm}(\exp(\Omega))}.$$

- Given the first-order marginal probabilities Q, we can compute the maximum entropy distribution consistent with the given marginals.

$$\begin{aligned} \min_q \quad & \sum_\sigma q(\sigma) \log q(\sigma) \\ \text{s.t.} \quad & \sum_\sigma q(\sigma) M_\sigma = Q \\ & q(\sigma) \geq 0 \end{aligned}$$

Hybrid Approach

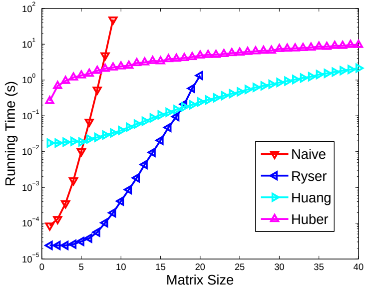
- Hybrid approach involve estimation of the matrix permanent.
  - Naive Algorithm: super-exponential.
  - Ryser Algorithm: exponential.
  - Huber et. al.: FPRAS.
  - Huang et. al.: belief propagation algorithm based on graphical models.
- Different rules for switching:
  - Myopic Switching: accuracy takes top priority.
  - Smoothness Based Switching: using Fourier (information) form to represent smooth (peaky) distributions.
  - Lagged Block Switching: look ahead k timesteps.
- Adaptively factorize the problem into independent components.

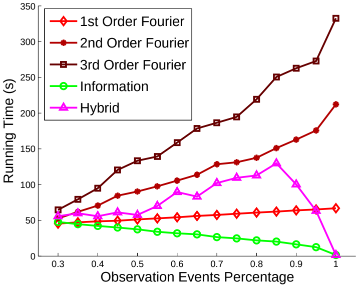
Real Camera Data from Simulation

- Simulated Data.
- Up to 100 moving targets.
- Complex movement patterns.

- Mixing event: whenever two persons get close to each other.
- Observation event: whenever a person is separated from all other persons.

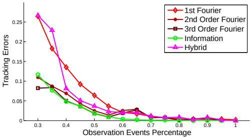
Running Time



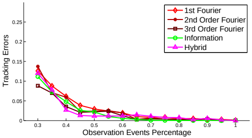


- Running time comparison of different approaches in computing matrix permanent; and the running time comparison of the three approaches.

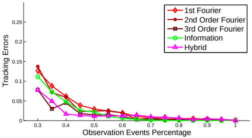
Accuracy



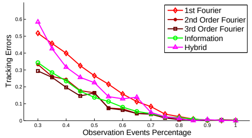
$p = .3, \ell = .55$



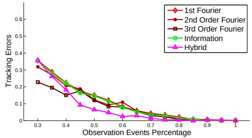
$p = .3, \ell = .75$



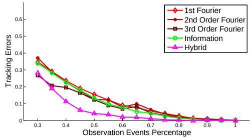
$p = .3, \ell = .95$



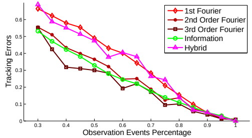
$p = .4, \ell = .55$



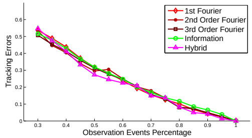
$p = .4, \ell = .75$



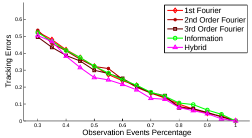
$p = .4, \ell = .95$



$p = .5, \ell = .55$

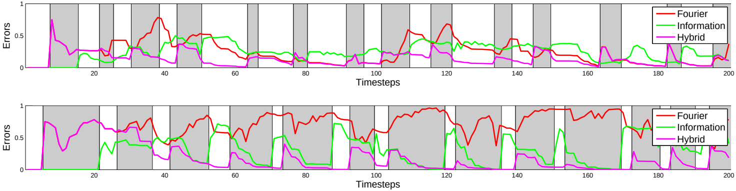


$p = .5, \ell = .75$



$p = .5, \ell = .95$

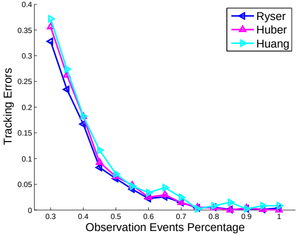
Errors in Distribution

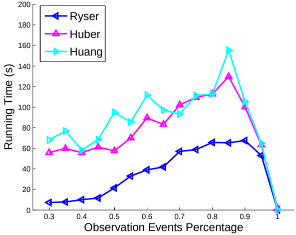


- Compare the errors in distribution of the three approaches. The white intervals denote the rollup steps and the grey intervals denote the conditioning steps.

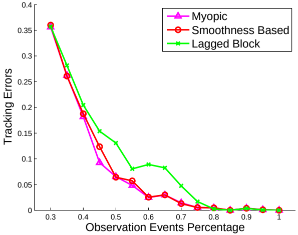
More Experiments

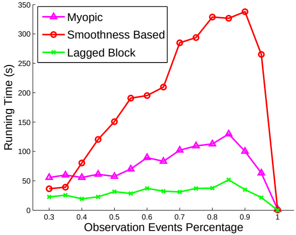
- Different matrix permanent approximation algorithms.



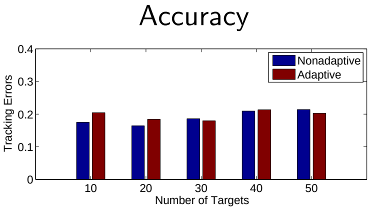


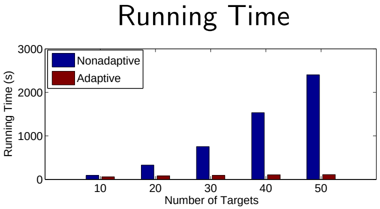
- Different switching rules.





Adaptive. vs Nonadaptive





- The tracking accuracy for the adaptive approach is comparable to the nonadaptive approach, while the running time can always be controlled using the adaptive approach.

Conclusions

- Established connections between the Fourier approach and the information approach.
- Proposed a novel hybrid approach.
- Generalized the hybrid approach to high orders.