Optimal Continuous DR-Submodular Maximization and Applications to Provable Mean Field Inference

Yatao (An) Bian, Joachim M. Buhmann, Andreas Krause

ETH Zurich

Motivation and Background

Product recommendation

Ground set $\mathcal{V}$: n products, n usually large

Which subset $S \subseteq \mathcal{V}$ to recommend?

Given a parameterized submodular utility $\mathrm{F(S)}$

→ Graphical model: $\mathrm{p(S)} \propto \mathrm{e}^{\mathrm{F(S)}}$

Mean Filed Approximation provides:

1, A differentiation technique to learn $\mathrm{F(S)}$ end-to-end

2, Approximate inference though the surrogate distribution q

Mean field inference aims to approximate $p(S)$ with a product distribution $q(S|\mathbf{x}) := \prod_{i \in S} x_i \prod_{j \notin S}(1 - x_j), \mathbf{x} \in [0,1]^n$

$$\max_{\mathbf{x}\in[\mathbf{0},\mathbf{1}]} f(\mathbf{x}) := \overbrace{\mathbb{E}_{q(S|\mathbf{x})}[F(S)]}^{\text{multilinear extension of } F(S):f_{\text{mt}}(\mathbf{x})} - \sum_{i=1}^{n}[x_i \log x_i + (1 - x_i)\log(1 - x_i)]$$

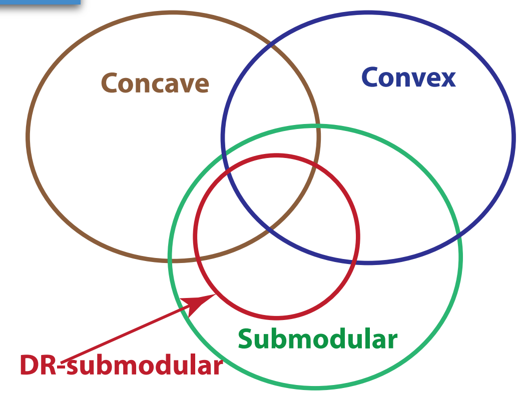
$$= f_{\text{mt}}(\mathbf{x}) + \sum_{i \in \mathcal{V}} H(x_i), \qquad \text{(ELBO)}$$

Highly non-convex →

😀 Continuous DR-Submodular wrt $\mathbf{x}$

Mean Field Inference as a Continuous DR-Submodular Maximization Problem

Guaranteed Non-Convex Optimization Problem:
Continuous DR-Submodular (Diminishing Returns) Maximization



$$\underset{\mathbf{x}\in[\mathbf{a},\ \mathbf{b}]}{\text{maximize}} \quad f(\mathbf{x})$$

$f(\mathbf{x})$ is continuous DR-submodular

DR-submodularity [BMBK17]: $\forall \mathbf{x} \leq \mathbf{y}, \forall i \in [n], \forall k \in \mathbb{R}_+$ it holds,

$$f(k\mathbf{e}_i + \mathbf{y}) - f(\mathbf{y}) \leq f(k\mathbf{e}_i + \mathbf{x}) - f(\mathbf{x})$$

Hardness: Box-constrained continuous DR-submodular maximization is *NP-hard*. There is no $(\frac{1}{2} + \epsilon)$-approximation for any $\epsilon > 0$ unless RP=NP

Provable Algorithm

Proposed **DR-DoubleGreedy**, which has a 1/2-approximation guarantee → *Optimal* Algorithm

Input: $\max_{\mathbf{x}\in[\mathbf{a},\mathbf{b}]} f(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^n$, $f(\mathbf{x})$ is DR-submodular

$\mathbf{x}^0 \leftarrow \mathbf{a}$, $\mathbf{y}^0 \leftarrow \mathbf{b}$; → Maintain two solutions

for $k = 1 \to n$ **do**

- let v_k be the coordinate being operated;
- find u_a such that $f(\mathbf{x}^{k-1}|_{v_k} u_a) \geq \max_{u'} f(\mathbf{x}^{k-1}|_{v_k} u') - \frac{\delta}{n}$,
- $\delta_a \leftarrow f(\mathbf{x}^{k-1}|_{v_k} u_a) - f(\mathbf{x}^{k-1})$;

} Solve 1-D problem on $\mathbf{x}$

- find u_b such that $f(\mathbf{y}^{k-1}|_{v_k} u_b) \geq \max_{u'} f(\mathbf{y}^{k-1}|_{v_k} u') - \frac{\delta}{n}$,
- $\delta_b \leftarrow f(\mathbf{y}^{k-1}|_{v_k} u_b) - f(\mathbf{y}^{k-1})$;

} Solve 1-D problem on $\mathbf{y}$

- $\mathbf{x}^k \leftarrow \mathbf{x}^{k-1}|_{v_k} (\frac{\delta_a}{\delta_a+\delta_b} u_a + \frac{\delta_b}{\delta_a+\delta_b} u_b)$;
- $\mathbf{y}^k \leftarrow \mathbf{y}^{k-1}|_{v_k} (\frac{\delta_a}{\delta_a+\delta_b} u_a + \frac{\delta_b}{\delta_a+\delta_b} u_b)$;

} Change coordinate to be a **convex** combination 😎

Output: $\mathbf{x}^n$ or $\mathbf{y}^n$ $(\mathbf{x}^n = \mathbf{y}^n)$

Empirical Evaluation

DR-DoubleGreedy outperforms SOTA algorithms for extensive real-world experiments

		ELBO objective			PA-ELBO objective		
Category	D	Sub-DG	BSCB	DR-DG	Sub-DG	BSCB	DR-DG
carseats	2	2.089±0.166	2.863±0.090	**3.045**±0.069	1.015±1.081	2.106±0.228	**2.348**±0.219
	3	1.890±0.146	3.003±0.110	**3.138**±0.082	1.309±1.218	2.414±0.267	**2.707**±0.208
n=34	10	1.390±0.232	**3.100**±0.140	3.003±0.157	1.599±1.317	2.684±0.271	**2.915**±0.250
safety	2	1.934±0.402	2.727±0.212	**2.896**±0.098	1.370±1.203	2.049±0.280	**2.341**±0.161
	3	1.867±0.453	2.830±0.191	**2.970**±0.110	1.706±1.296	2.288±0.297	**2.619**±0.167
n=36	10	1.546±0.606	2.916±0.191	**2.920**±0.149	1.948±1.353	2.467±0.270	**2.738**±0.187
strollers	2	2.042±0.181	2.829±0.144	**2.928**±0.060	0.865±0.952	1.933±0.256	**2.202**±0.226
	3	1.814±0.264	2.958±0.146	**2.978**±0.077	1.172±1.063	2.181±0.297	**2.543**±0.254
n=40	10	1.328±0.544	**3.065**±0.162	2.910±0.140	1.702±1.334	2.480±0.304	**2.767**±0.336
media	2	3.221±0.066	3.309±0.055	**3.493**±0.051	0.372±0.286	**1.477**±0.128	1.336±0.101
	3	3.276±0.082	3.492±0.083	**3.712**±0.079	0.418±0.366	1.736±0.177	**1.762**±0.095
n=58	10	2.840±0.183	3.894±0.122	**3.924**±0.114	0.653±0.727	2.309±0.244	**2.524**±0.130
toys	2	3.543±0.047	3.454±0.091	**3.856**±0.044	0.597±0.480	1.731±0.182	**1.761**±0.133
	3	3.362±0.055	3.412±0.070	**3.736**±0.051	0.578±0.520	1.738±0.192	**1.802**±0.151
n=62	10	3.037±0.138	3.706±0.108	**3.859**±0.119	0.758±0.871	2.140±0.242	**2.330**±0.177
bedding	2	3.406±0.080	3.374±0.088	**3.620**±0.062	0.525±0.121	1.932±0.194	**2.001**±0.080
	3	3.648±0.106	3.564±0.083	**3.876**±0.081	2.499±0.972	2.250±0.269	**2.624**±0.066
n=100	10	3.355±0.161	3.799±0.144	**3.912**±0.082	**3.919**±0.045	2.578±0.358	3.157±0.091
apparel	2	3.560±0.094	3.527±0.046	**3.784**±0.059	0.268±0.109	**1.552**±0.141	1.513±0.191
	3	3.878±0.092	3.755±0.062	**4.140**±0.063	0.490±0.677	1.900±0.237	**2.225**±0.136
n=100	10	3.751±0.087	4.084±0.075	**4.425**±0.066	0.820±1.372	2.351±0.337	**2.967**±0.150

→ source code released & poster #98