# Subset Selection and Summarization in Sequential Data

**Ehsan Elhamifar**
e-mail: eelhami@ccs.neu.edu

**M. Clara De Paolis Kaluza**
e-mail: clara@ccs.neu.edu

College of Computer and Information Science, Northeastern University, Boston, USA

## Motivation

**Subset selection** finds a small a subset of most informative items from a large ground set to be used for summarization and other inference and learning tasks.

**Sequential data**, including time-series, such as video and speech, and ordered data, such as text, form a significant part of modern datasets.

- There exist structural dependencies among sequential data, imposed by the underlying dynamic model, that must play a vital role in summarization.
- Existing subset selection methods ignore dynamics, treating data as a bag of randomly permutable items.

Summary:

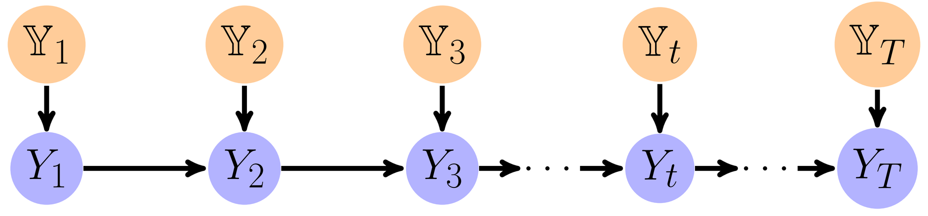
## Contributions

- Develop a **sequential subset selection** framework, incorporating dynamics.
- Form potentials to optimize encoding, cardinality and **coherency** of the summary.
- Propose a **binary optimization** over data assignments to representatives.
- Develop a **max-sum message passing** and an ADMM framework.

## Prior Work

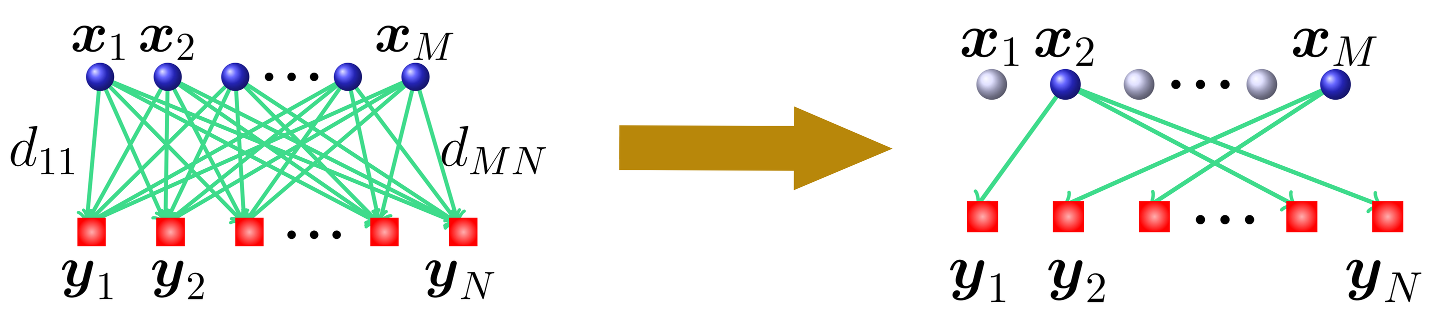
Determinantal Point Processes (DPPs) choose subset(s) $Y$ of data $\mathbb{Y}$

- Markov DPP [1]: successively selects items, diverse from previously selected items.
- SeqDPP [2]: divides a sequence into windows $\mathbb{Y}_t$ and selects sets $Y_t$ diverse within window and with respect to items selected in previous window.




- **Limitations**: i) do not consider dynamics of data; ii) single set summarization.

## Facility Location Review

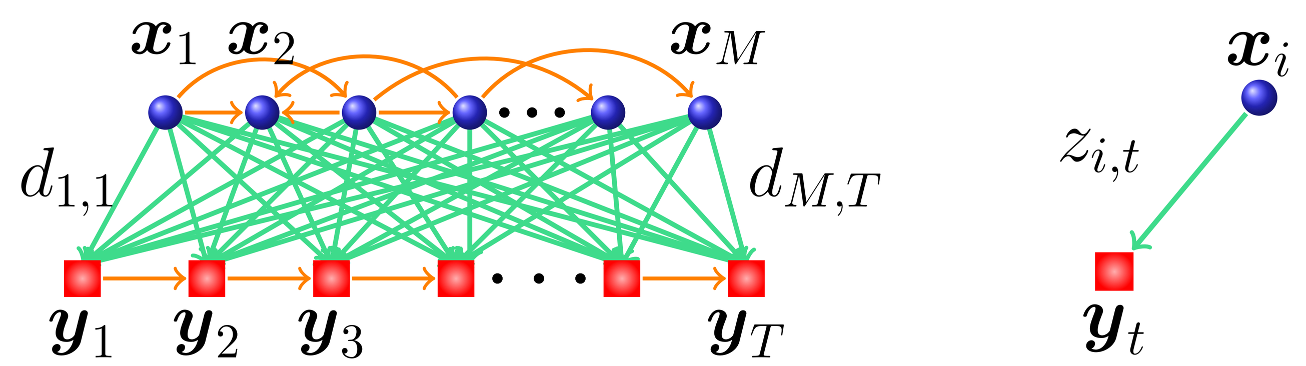



- Given: source set $\mathbb{X}$, target set $\mathbb{Y}$, and pairwise dissimilarity $d_{ij}$.
- $d_{ij}$: how well $x_i$ represents $y_j$, smaller means better.
- Goal: find a small subset $\mathcal{S} \subseteq \mathbb{X}$ to represent every item $y_j \in \mathbb{Y}$.
- Minimize cardinality plus encoding quality of the representative set:

$$\min_{\mathcal{S}\subseteq\mathbb{X}} \lambda|\mathcal{S}| + \sum_{j\in\mathbb{Y}} \min_{i\in\mathcal{S}} d_{ij}$$

## Sequential Facility Location

- **Approach:** Introduce transition model among source set items $p(\boldsymbol{x}_{i'}|\boldsymbol{x}_{i_1},\ldots,\boldsymbol{x}_{i_n})$
  - Target set has a sequential structure $\mathbb{Y} = (\boldsymbol{y}_1,\ldots,\boldsymbol{y}_T)$.
  - $\boldsymbol{x}_{r_t}$ denotes the representative of $\boldsymbol{y}_t$, for $t \in \{1,\ldots,T\}$.




  - Maximize potential function over representative assignments $(r_1,\ldots,r_T) \subseteq \{1,\ldots,M\}^T$.

$$\Psi(r_1,\ldots,r_T) \triangleq \Phi_{\text{enc}}(r_1,\ldots,r_T) \times \Phi_{\text{card}}(r_1,\ldots,r_T) \times \Phi_{\text{dyn}}(r_1,\ldots,r_T)$$

Encoding: $$\Phi_{\text{enc}}(r_1,\ldots,r_T) = \prod_{t=1}^{T} \phi_{\text{enc},t}(r_t) = \prod_{t=1}^{T} e^{-d_{r_t,t}}$$

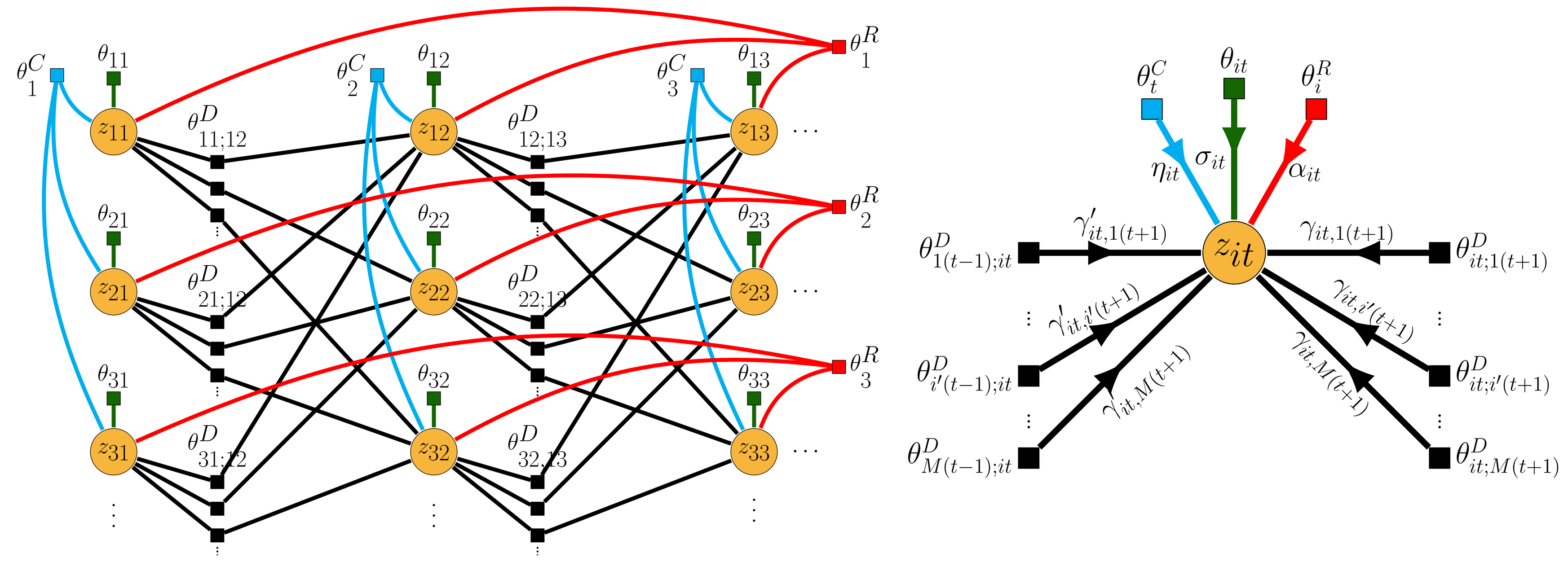
Cardinality: $$\Phi_{\text{card}}(r_1,\ldots,r_T) = \exp(-\lambda|\{r_1,\ldots,r_T\}|)$$

Dynamics: $$\Phi_{\text{dyn}}(r_1,\ldots,r_T) = \left(\prod_t p_t(\boldsymbol{x}_{r_t}|\boldsymbol{x}_{r_{t-1}},\ldots,\boldsymbol{x}_{r_{t-n}})\right)^{\beta}$$

- **Integer Binary Optimization Formulation:**
  - Binary assignment variable $z_{i,t} \in \{0,1\}$, indicates if $\boldsymbol{x}_i$ is a representative of $\boldsymbol{y}_t$.
  - Consider first-order Markov model and maximize $\log \Psi$

$$\max_{\{z_{i,t}\}} \sum_{t=1}^{T}\sum_{i=1}^{M} -z_{i,t}d_{i,t} - \lambda\sum_{i=1}^{M} \| [z_{i,1} \cdots z_{i,T}] \|_\infty + \beta(\sum_{i=1}^{M} z_{i,1}\log p_1(\boldsymbol{x}_i))$$
$$+ \sum_{t=2}^{T}\sum_{i,i'=1}^{M} z_{i,t-1}z_{i',t}\log p(\boldsymbol{x}_{i'}|\boldsymbol{x}_i)) \quad \text{s. t.} \quad z_{i,t}\in\{0,1\}, \quad \sum_{i=1}^{M} z_{i,t} = 1, \ \forall\, i,t.$$

- **Optimization via Max-Sum Message Passing:** cast the optimization as a MAP inference on binary random variables.



Encoding
$$\theta_t(z_{i,t}) \triangleq \begin{cases} -z_{i,t}(d_{i,t} - \log p_1(\boldsymbol{x}_i)), & t = 1 \\ -z_{i,t}d_{i,t}, & \text{otherwise} \end{cases}$$

Cardinality
$$\theta_i^R(\boldsymbol{z}_{i,:}) \triangleq \begin{cases} -\lambda, & \|\boldsymbol{z}_{i,:}\|_\infty > 0 \\ 0, & \text{otherwise} \end{cases}$$

Constraint
$$\theta_t^C(\boldsymbol{z}_{:,t}) \triangleq \begin{cases} 0, & \sum_{i=1}^{M} z_{i,t} = 1 \\ -\infty, & \text{otherwise} \end{cases}$$

Dynamic
$$\theta_t^D(z_{i,t}) \triangleq \log p(\boldsymbol{x}_{i'}|\boldsymbol{x}_i) z_{i,t-1}z_{i',t}$$

## Experiments

- **Synthetic Experiments:** $\mathbb{X}$: means of 50 Gaussians, $\mathbb{Y}$: from a Markov model

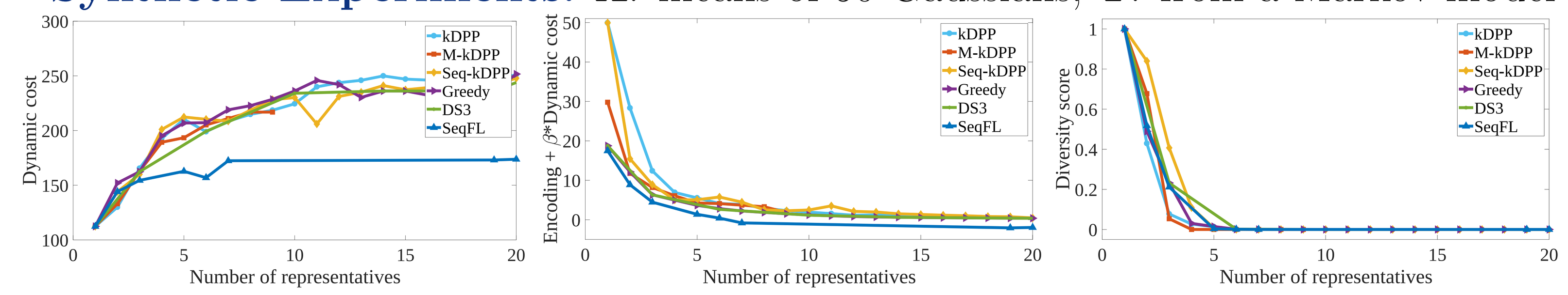



**Figure 1:** Dynamic cost, total cost and diversity score as a function of the number of representatives.

- SeqFL achieves **lower costs** and **higher diversity** than DPP methods.

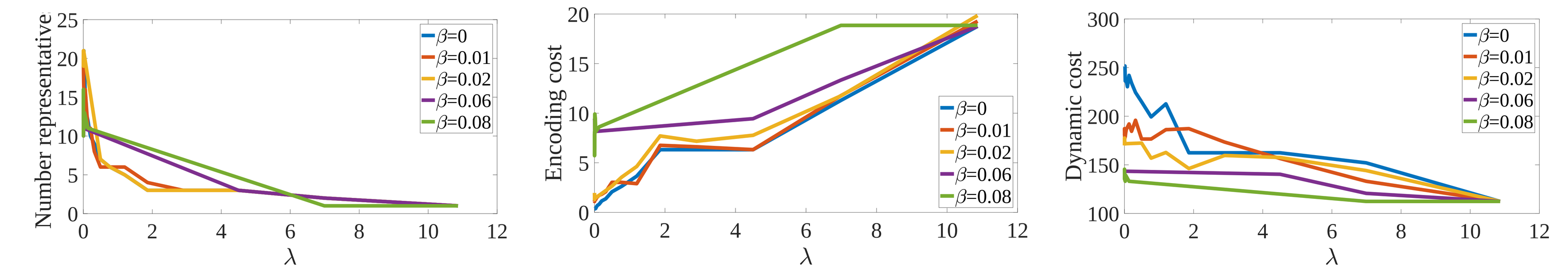



**Figure 2:** Number of representatives, encoding cost and dynamic cost of SeqFL as a function of the parameters $(\beta, \lambda)$.

- **Instructional Video Summarization**:
  Use the Instructional Video dataset [5]: 5 tasks, 30 videos per task available.
  - Fit HMM to training data to construct transition model.
  - Use SeqFL to choose representative HMM states for each test video.
  - Assign labels to states based on training set nearest neighbors.
  - Align sequences of representatives from all test videos to form final summary.

| Task | | kDPP[4] | M-kDPP[1] | Seq-kDPP[2] | DS3[3] | SeqFL |
|---|---|---|---|---|---|---|
| Change tire | (P, R) | (0.56, 0.50) | (0.55, 0.60) | (0.44, 0.40) | (0.56, 0.50) | (0.60, 0.60) |
| | F-score | 0.53 | 0.57 | 0.42 | 0.53 | **0.60** |
| Make coffee | (P, R) | (0.38, 0.33) | (0.50, 0.44) | (0.63, 0.56) | (0.50, 0.56) | (0.50, 0.56) |
| | F-score | 0.35 | 0.47 | **0.59** | 0.53 | 0.53 |
| CPR | (P, R) | (0.71, 0.71) | (0.71, 0.71) | (0.71, 0.71) | (0.71, 0.71) | (0.83, 0.71) |
| | F-score | 0.71 | 0.71 | 0.71 | 0.71 | **0.77** |
| Jump car | (P, R) | (0.50, 0.50) | (0.56, 0.50) | (0.56, 0.50) | (0.50, 0.50) | (0.60, 0.60) |
| | F-score | 0.50 | 0.53 | 0.53 | 0.50 | **0.60** |
| Repot plant | (P, R) | (0.57, 0.67) | (0.60, 0.50) | (0.57, 0.67) | (0.57, 0.67) | (0.80, 0.67) |
| | F-score | 0.62 | 0.55 | 0.62 | 0.62 | **0.73** |
| All tasks | (P, R) | (0.54, 0.54) | (0.58, 0.55) | (0.58, 0.57) | (0.57, 0.59) | (0.67, 0.63) |
| | F-score | 0.54 | 0.57 | 0.57 | 0.58 | **0.65** |

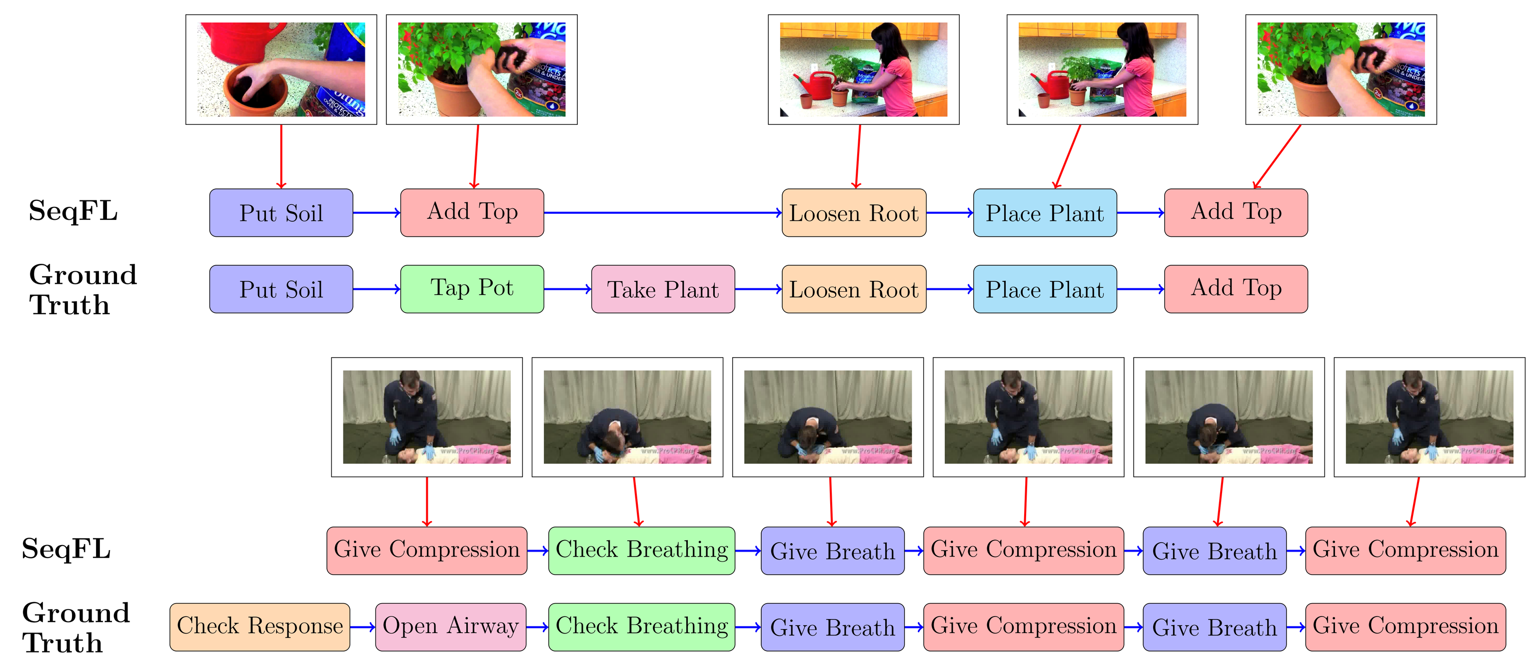



**Figure 3:** Summaries obtained for the task of repotting a plant (top) and performing CPR (bottom).

[1] R. H. Affandi, A. Kulesza, and E. B. Fox, Markov determinantal point processes, UAI, 2012.
[2] B. Gong, et al., Diverse sequential subset selection for supervised video summarization, NIPS, 2014
[3] E. Elhamifar, G. Sapiro, S. Sastry. Dissimilarity-based sparse subset selection, PAMI,2016.
[4] A. Kulesza and B. Taskar, K-DPPs: Fixed-size determinantal point processes, ICML, 2011.
[5] J.-B. Alayrac, et al., Unsupervised learning from narrated instruction videos, CVPR, 2016.