Spectral Learning of Sequence Taggers over Continuos Sequences

A.Recasens A.Quattoni

Universitat Politècnica de Catalunya

ECML PKKD 2013, Prague

Tagging Continuous Sequences

Setting:

- We are given pairs $< [x_1, x_2 \dots x_T], [y_1, y_2 \dots y_T] >$ of aligned sequences.
- $\Phi(x_i) \in \mathbb{R}^k$, a real feature representation of x_i.
- and $y_i \in \Sigma$, where Σ is some discrete set of Tags.

Goal :

- We want to learn a model of $\mathbb{P}(x, y)$.
- We will use it to make predictions: $argmax_y \mathbb{P}(x, y)$.
- **Contribution: A Spectral Algorithm for this task.**

Examples of Tagging Problems:

- Discret tagging problems:

He	reckons	the	current	account	deficit	will	narrow	significantly
[PRP]	[VB]	[DT]	[JJ]	[NN]	[NN]	[MD]	[VB]	[RB]

Examples of Tagging Problems:

- Discret tagging problems:

He reckons the current account deficit will narrow significantly

[PRP] [VB] [DT] [JJ] [NN] [NN] [MD] [VB] [RB]

- Gesture Recognition:

[HTF] [HTF] [HTF] [HOF] [HOF] [HOS]

Examples of Tagging Problems:

- Discret tagging problems:

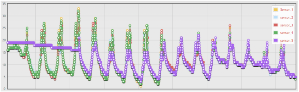
He reckons the current account deficit will narrow significantly

[PRP] [VB] [DT] [JJ] [NN] [NN] [MD] [VB] [RB]

- Gesture Recognition:

[HTF] [HTF] [HTF] [HOF] [HOF] [HOS]

- Robot navegation:



MR MR LT LT ML ML MR RT RT ML ML

Outline

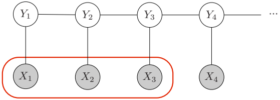
- Spectral Learning of Sequence Models background.
- A Model for Tagging Continuous Sequences.
- Spectral Learning Algorithm for Continuous Tagging.

Outline

- **Spectral Learning of Sequence Models background.**
- A Model for Tagging Continuous Sequences.
- Spectral Learning Algorithm for Continuous Tagging.

A Simple Spectral Method [HKZ09]

Discrete Homogeneous Hidden Markov Model



- m states – $S_t \in \{1, \ldots, m\}$.
- k symbols – $x_t \in \{\sigma_1, \ldots, \sigma_k\}$.
- Forward-backward equations with $A_\sigma \in \mathbb{R}^{m \times m}$:
$$\mathbb{P}(x) = \alpha_1^\top A_{x_1} \cdots A_{x_t} \vec{1}$$

- Probabilities arranged into matrices $H, H_{\sigma_1}, \ldots, H_{\sigma_k} \in \mathbb{R}^{k \times k}$.
$$H(i,j) = \mathbb{P}(x_{t-1} = \sigma_i,\ x_t = \sigma_j)$$
$$H_\sigma(i,j) = \mathbb{P}(x_{t-1} = \sigma_i,\ x_t = \sigma,\ x_{t+1} = \sigma_j)$$

- Spectral learning algorithm:
  1. Compute SVD $H = UDV^\top$ and take top m right singular vectors V_m.
  2. $A_\sigma = (HV_m)^+ H_\sigma V_m$.

Sequence Tagging

Discrete FST:

- Input alphabet: Δ.
- Output Alphabet: Σ.
- Operators: $A_\delta^\sigma \in \mathbb{R}^{m\times m}$, depend on input and output.
- $\mathbb{P}(x, y) = \alpha_1^\top A_{x_1}^{y_1} \cdots A_{x_T}^{y_T} \alpha_\infty$
- Balle et al, [ECML 2011] developed a Spectral Algorithm.

We can try to solve our problem by...

- $\mathbb{P}(x_{1:T}, y_{1:T}) = \alpha_1^T A_{\phi(x_1)}^{y_1} \ldots A_{\phi(x_T)}^{y_T} \alpha_\infty \Rightarrow$ Infinite operators!
- A discretisation of the $\mathcal{X}$ space and use [Balle et al, 2011].
- **Generalizing the FST to continuous inputs.**

Outline

- Spectral Learning of Sequence Models background.
- A Model for Tagging Continuous Sequences.
- Spectral Learning Algorithm for Continuous Tagging.

Outline

- Spectral Learning of Sequence Models background.
- **A Model for Tagging Continuous Sequences.**
- Spectral Learning Algorithm for Continuous Tagging.

Continuous Sequence Taggers (CFST)

A *CFST* over $(\Phi(\mathcal{X}) \times \Sigma)^*$ with m states is a tuple:
$A = \langle \Phi, \alpha_1, \alpha_\infty, O_l^\sigma \rangle$ where:

- Φ is a set of k feature functions.
  $\phi_l(x) : \mathcal{X} \rightarrow \mathbb{R}$, the real feature representation of $\mathcal{X}$.
- $\alpha_1, \alpha_\infty \in \mathbb{R}^m$ are starting and ending parameters.
- $O_l^\sigma \in \mathbb{R}^{m \times m}$ are the $k \times |\Sigma|$ operators.
  There is one operator for each output symbol and input feature.
- $A(\Phi(x_t), y_t) = \sum_{l=1}^{k} \phi_l(x_t) O_l^{y_t}$
  The operator function is a combination of the feature operators.

Continuous Sequence Taggers

CFST

The function f_A realised by the CFST is defined by:

$$\begin{aligned} f_A(x,y) &= \alpha_1^\top A(\Phi(x_1),y_1)\cdots A(\Phi(x_T),y_T)\,\alpha_\infty \\ &= \alpha_1^\top \left(\sum_{l=1}^{k}\phi_l(x_1)O_l^{y_1}\right)\cdots\left(\sum_{l=1}^{k}\phi_l(x_T)O_l^{y_T}\right)\alpha_\infty \end{aligned}$$

Main Idea:

The operators A_δ^σ have become $A_x^\sigma = \sum_{l=1}^{k}\phi_l(x)O_l^\sigma$.

Examples

Discrete FST A as CFST A'

- For each input δ we define $\phi_\delta(x) = \mathbb{I}_\delta(x)$.
- $\Phi(x = \sigma) = [0 \ldots 1 \ldots 0]$.
  A real vector $\in \mathbb{R}^k$ of zeros with a 1 at position σ.
- Set $O_l^\sigma = A_l^\sigma$, $\alpha'_1 = \alpha_1$ and $\alpha'_\infty = \alpha_\infty$.
- Finally: $A(\Phi(\delta), \sigma) = A_\delta^\sigma$.

Transitions as Mixture Models:

- $\mathbb{P}(x, y) = \sum_h \mathbb{P}(h_0) \prod_{t=1}^{T-1} \mathbb{P}(h_{t+1}, x_t, y_t \mid h_t)$.
- $\mathbb{P}(h_{t+1}, x_t, y_t \mid h_t) = \sum_{l=1}^{k} \mathbb{P}_l(h_{t+1}, y_t \mid h_t)\mathbb{P}(z = l, x_t)$.
- $\phi(x) = [\mathbb{P}(z = 1, x) \ldots \mathbb{P}(z = k, x)]$.
- $O_l^y = \mathbb{P}_l(h_{t+1}, y \mid h_t)$.

Outline

- Spectral Learning of Sequence Models background.
- A Model for Tagging Continuous Sequences.
- Spectral Learning Algorithm for Continuous Tagging.

Outline

- Spectral Learning of Sequence Models background.
- A Model for Tagging Continuous Sequences.
- **Spectral Learning Algorithm for Continuous Tagging.**

Spectral Learning Algorithm

Observable Statistics:

- $H_1 \in \mathbb{R}^k$, where $H_1(i) = \mathbb{E}_{\mathbb{P}}[\phi_i(x_t)]$.
  Input unigram feature expectations.
- $H_2 \in \mathbb{R}^{k \times k}$, where $H_2(i,j) = \mathbb{E}_{\mathbb{P}}[\phi_i(x_t)\phi_j(x_{t+1})]$.
  Input bigram feature expectations.
- $H_l^\sigma \in \mathbb{R}^{k \times k}$, where
  $H_l^\sigma(i,j) = \mathbb{E}_{\mathbb{P}y_t=\sigma}(\phi_i(x_{t-1})\phi_l(x_t)\phi_j(x_{t+1}))$.
  Input trigram feature expectations conditioned on y_t.
- $C \in \mathbb{R}^{k \times k}$, where $C(i,j) = \mathbb{E}_{\mathbb{P}}[\phi_i(x_t)\phi_j(x_t)]$.

Duality between CFST and factorizations of H_2

Theorem: Minimal CFST $A \iff$ Rank factorization of H_2.

Remarks $\Rightarrow$:

- Hypothesis: minimal CFST.
- H_2 can be written as $H_2 = FB$ where $F \in \mathbb{R}^{k\times m}$ and $B \in \mathbb{R}^{m\times k}$,
- $H_l^\sigma = F \sum_{i=1}^k O_i^\sigma C(l,i) B$ and $H_1 = F\alpha_\infty = \alpha_1^\top B$

Remarks $\Leftarrow$:

- Hypothesis: $H_2 = FB$, a rank factorization.
- $A = \langle \Phi, \alpha_1, \alpha_\infty, O_l^\sigma \rangle$ can be defined as:
  - $\alpha_\infty = F^+ H_1 \quad \alpha_1^\top = H_1 B^+ \quad Q_l^\sigma = F^+ H_l^\sigma B^+$.
  - $[O_1^\sigma(i,j), \dots, O_k^\sigma(i,j)]^\top = C^{-1}[Q_1^\sigma(i,j), \dots, Q_k^\sigma(i,j)]$.
- Then, A computes f.

Spectral algorithm

`LearnCFST`$(\mathcal{X}, \Phi, \Sigma, S, m)$

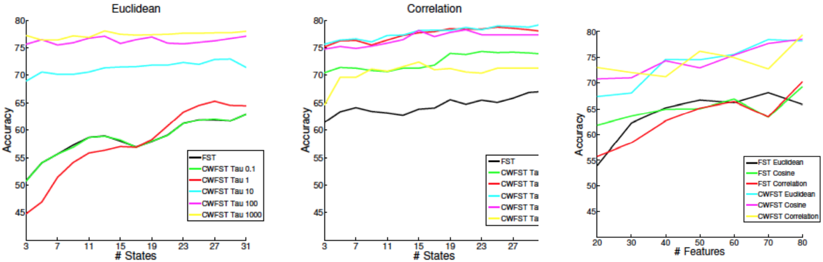
---

1. For every pair of sequences (x, y) in S and every index $1 < t < |x|$compute $\phi(x_t) = [\phi_1(x_t), \dots, \phi_k(x_t)]$
2. Use S to estimate matrix statistics $\widehat{H}_1 \in \mathbb{R}^k$, $\widehat{H}_2 \in \mathbb{R}^{k \times k}$, $\widehat{H}_l^\sigma \in \mathbb{R}^{k \times k}$ and $\widehat{C} \in \mathbb{R}^{k \times k}$.
3. Compute the m rank compact SVD of $\widehat{H}_2 = (U\Lambda)V^\top$.
4. Compute the inverse of $\widehat{C}$ and $Q_l^\sigma = (\widehat{H}_2 V)^+ \widehat{H}_l^\sigma V$.
5. Compute the start and ending parameters of the CWFST as: $\alpha_1^\top = \widehat{H}_1 V \quad \alpha_\infty = (\widehat{H}_2 V)^+ \widehat{H}_1$
6. Compute the transition matrices O_l^σ:

$$\begin{bmatrix} O_1^\sigma(i,j) \\ \vdots \\ O_k^\sigma(i,j) \end{bmatrix} = \widehat{C}^{-1} \begin{bmatrix} Q_1^\sigma(i,j) \\ \vdots \\ Q_k^\sigma(i,j) \end{bmatrix}$$

The Experiment

- Task Robot Navegation
  - Input: Sequence of Sensor Readings.
  - Output: Sequence of optimal Actions.
- Features
  - Select $\{z_1 \dots z_k\}$ points in $\mathbb{R}^k$ (e.g. via kmeans).
  - Define $\phi_l(x) = e^{-\frac{D(z_l, x)}{\tau}}$.



Want to know more?

Poster Stand Number 15