時系列予測のための深層学習の新しい手法の考案

ICGI併催のチャレンジで優勝

the 13 th International Conference in Grammatical Inference

文法推論の国際会議

Sequence PredIction ChallengE (SPiCe) :

http://spice.lif.univ-mrs.fr

時系列(記号列)予測のコンテスト

多様なデータセット：

様々な国の自然言語や、ネットワークのアクセスログや通信データなど、様々

時期と規模：

コンテストの実施期間自体は、3月17日〜8月1日

登録者数約65チーム、 参加者数17チーム、最終提出者数12チーム

http://www.teu.ac.jp/information/2016.html?id=254

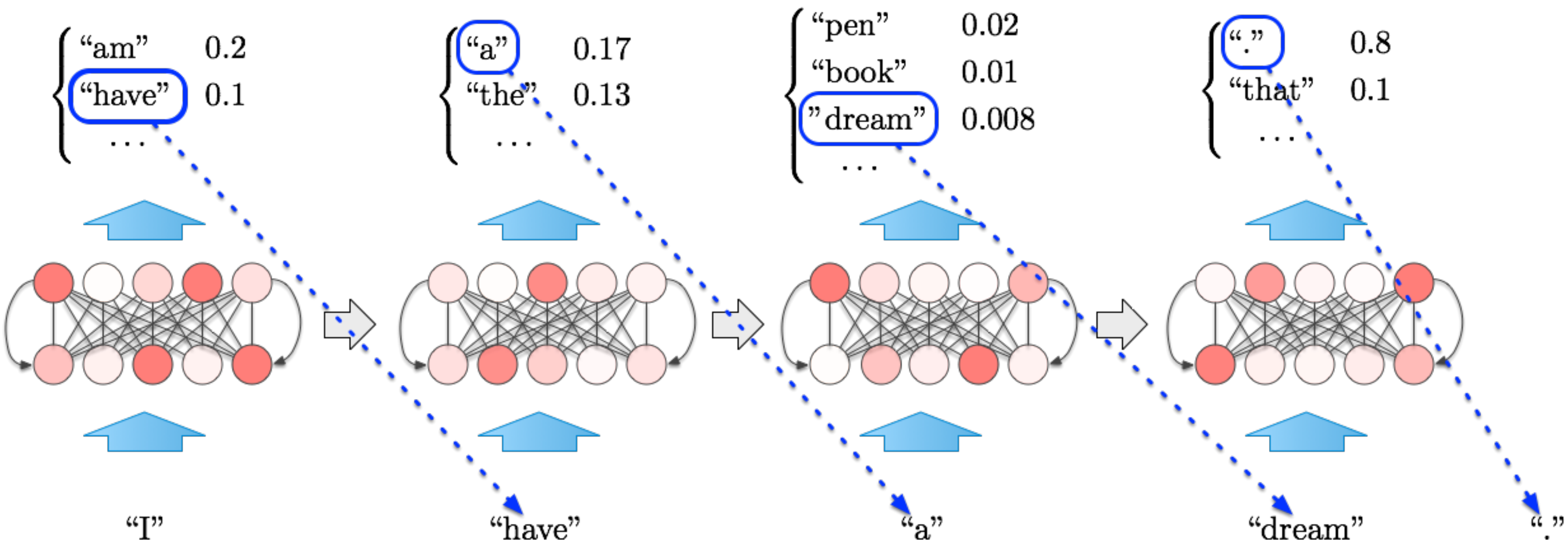
用いた手法は以下の組み合わせ:

LSTM ベース

SP-k 言語のベクトル表現

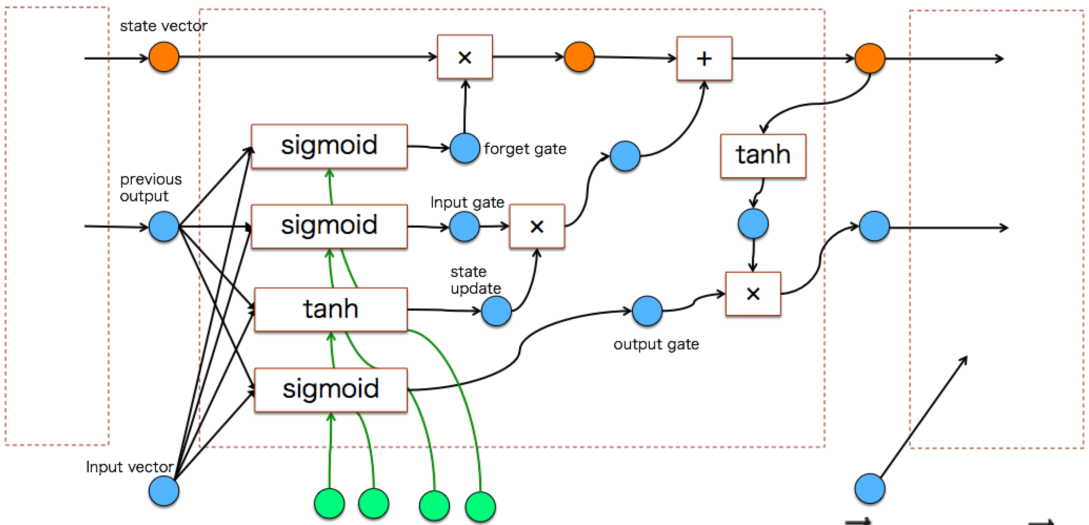
時系列予測とリカレントニューラルネット(RNN)

Recurrent neural networks(RNN) are empirically known to be useful when next elements in time-series data are required to be predicted. e.g.) sentence generation:



Long Short-term Memory (LSTM)

- LSTM is a kind of RNN designed to track long-term dependencies.
  - It has a redundant hidden output which is called “the state vector”.



$$\vec{f} = \sigma(W_f[x,h] + b_f)$$
$$\vec{i} = \sigma(W_i[x,h] + b_i)$$
$$\Delta\vec{c} = \tanh(W_o[x,h] + b_o)$$
$$\vec{o} = \sigma(W_o[x,h] + b_o)$$

$$\vec{c}_{\text{new}} = \vec{f} \odot \vec{c} + \vec{i} \odot \Delta\vec{c}$$
$$\vec{h}_{\text{new}} = \vec{o} \odot \tanh(\vec{c}_{\text{new}})$$

SP-k vector representation:

e.g.

Σ={a,b,c,d,e}
prefix: x = a d a c d

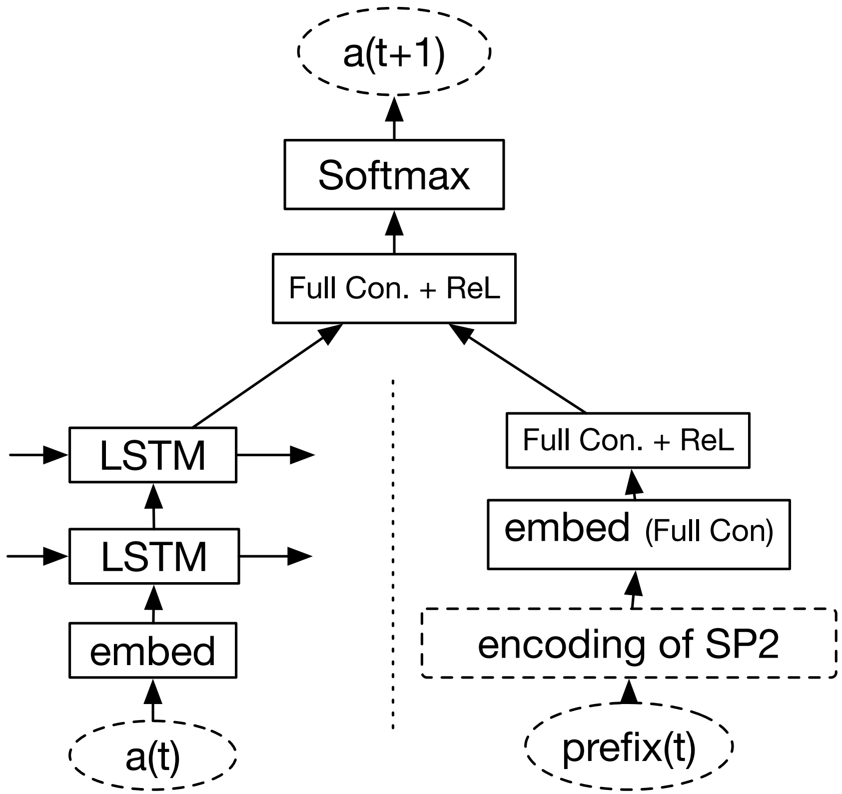
The SP-2 vector for prefix x: $\begin{array}{ccccc} a & b & c & d & e \\ \left[1\right. & 0 & 1 & 1 & \left.0\right] \end{array}$

The SP-3 vector (or tensor) for x: $$\begin{array}{c c} & \begin{array}{ccccc} a & b & c & d & e \end{array} \\ \begin{array}{c} a \\ b \\ c \\ d \\ e \end{array} & \begin{bmatrix} 1 & & 1 & 1 & \\ & & & & \\ & & & 1 & \\ 1 & & 1 & 1 & \\ & & & & \end{bmatrix} \end{array}$$

The num of 1s:
SP-2: $\mathcal{O}(\max\{|\Sigma|, |w|\})$
SP-3: $\mathcal{O}(\max\{|\Sigma|^2, |w|^2/2\})$

Strictly Piecewise Model

- We concatenate the output of the LSTM network and the SP-2 vector.
- SP-2 vector:
  - Encode the prefix into the SP-2 zero-one vector.
  - Embed into a real-valued vector through a fully connected layer
  - As an intermediate layer, use another fully connected layer with a non-linear activation function.
- LSTM network:
  - the same as in the basic model.



Results of Experiments

- The tables show the top-5 scores for each problem in the SPiCe challenge.

Table 1: comparison of scores with the dim. 400

	simple(400)	sp2(400)	bigram(400)
1	0.906(0.002)	**0.915**(0.000)	0.836(0.009)
2	0.920(0.000)	0.919(0.000)	0.878(0.003)
3	0.884(0.001)	0.884(0.001)	0.848(0.001)
4	0.615(0.001)	0.614(0.002)	**0.633**(0.002)
6	0.848(0.001)	**0.855**(0.001)	0.836(0.001)
7	0.717(0.001)	0.718(0.000)	**0.730**(0.001)
8	0.646(0.001)	0.646(0.001)	0.630(0.001)
9	0.959(0.001)	0.960(0.000)	0.958(0.000)
10	0.575(0.001)	0.577(0.001)	0.569(0.001)
11	0.529(0.001)	0.527(0.001)	–
12	0.782(0.002)	**0.796**(0.000)	0.727(0.003)
13	0.588(0.001)	0.588(0.001)	0.578(0.001)
14	0.344(0.001)	0.346(0.001)	0.332(0.001)
15	0.264(0.001)	0.265(0.001)	0.261(0.000)

Table 2: comparison of scores with the dim. 600

	simple(600)	sp2(600)	bigram(600)
1	0.909(0.002)	**0.915**(0.000)	0.769(0.003)
2	0.920(0.000)	0.920(0.000)	0.838(0.004)
3	0.888(0.001)	0.886(0.001)	0.831(0.001)
4	0.619(0.002)	0.616(0.002)	**0.634**(0.001)
6	0.863(0.001)	**0.867**(0.001)	0.828(0.002)
7	0.736(0.000)	0.736(0.001)	**0.747**(0.001)
8	0.645(0.001)	0.644(0.001)	0.614(0.001)
9	0.962(0.000)	0.962(0.000)	0.959(0.000)
10	0.574(0.001)	0.573(0.001)	0.570(0.002)
11	0.520(0.001)	0.519(0.001)	–
12	0.799(0.002)	**0.807**(0.001)	0.713(0.001)
13	0.592(0.001)	0.590(0.001)	0.581(0.000)
14	0.350(0.002)	0.351(0.002)	0.333(0.002)
15	0.263(0.001)	0.263(0.001)	0.258(0.001)

- Comparison proposed models to basic models:
  - SP-2 model : for some problems, is significantly better,
    and for other propems, has no significant difference.
  - Bigram model : for some problems, is significantly better as well,
    but for other problems, is significantly worse.