CS 581

Advanced Artificial Intelligence

March 06, 2024

Announcements / Reminders

- **Please follow the Week 09 To Do List instructions (if you haven't already)**
- **Next week: Spring Break! No office hours.**
- **Programming Assignment #02: will be posted soon**
- **Written Assignment #03: will be posted soon**

Plan for Today

- **Decision Networks**
- **Probabilistic Reasoning over Time**
  - **Hidden Markov Model**

Value of Perfect Information

The value/utility of best action α without additional evidence (information) is :

$$\mathrm{MEU}(\alpha) = \max_{a} \sum_{s'} \mathrm{P}(Result(a) = s') * \mathrm{U}(s')$$

If we include new evidence/information ($\mathrm{E_j} = \mathrm{e_j}$) given by some variable Ej, value/utility of best action α becomes:

$$\mathrm{MEU}(a_{e_j} \mid e_j) = \max_{a} \sum_{s'} \mathrm{P}(Result(a) = s' \mid e_j) * \mathrm{U}(s')$$

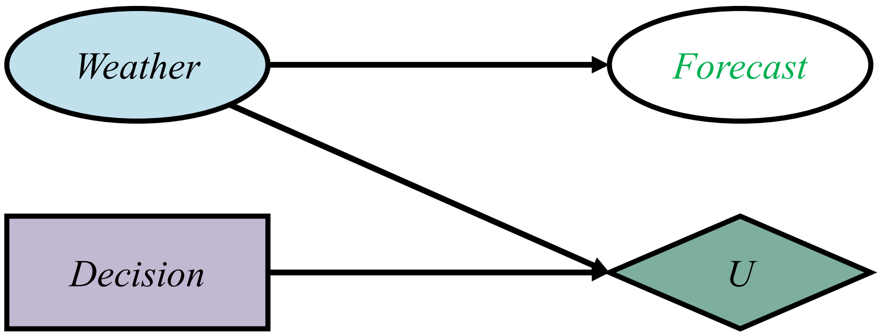
The value of additional evidence/information from Ej is:

$$\mathrm{VPI}(E_j) = \left(\sum_{e_j} \mathrm{P}(E_j = e_j) * \mathrm{MEU}(a_{e_j} \mid E_j = e_j)\right) - MEU(a)$$

using our current beliefs about the world.

Decision Network: Example

Decision network



The value of best action α without additional evidence

$$\mathrm{MEU}(\alpha) = \mathrm{MEU}(\mathrm{leave}) = 70$$

With evidence information ($E_j = e_j$) given by Forecast:

$$\mathrm{MEU}(a_{e_1} \mid e_1) = \mathrm{MEU}(\mathrm{take} \mid F = rain) = 53$$

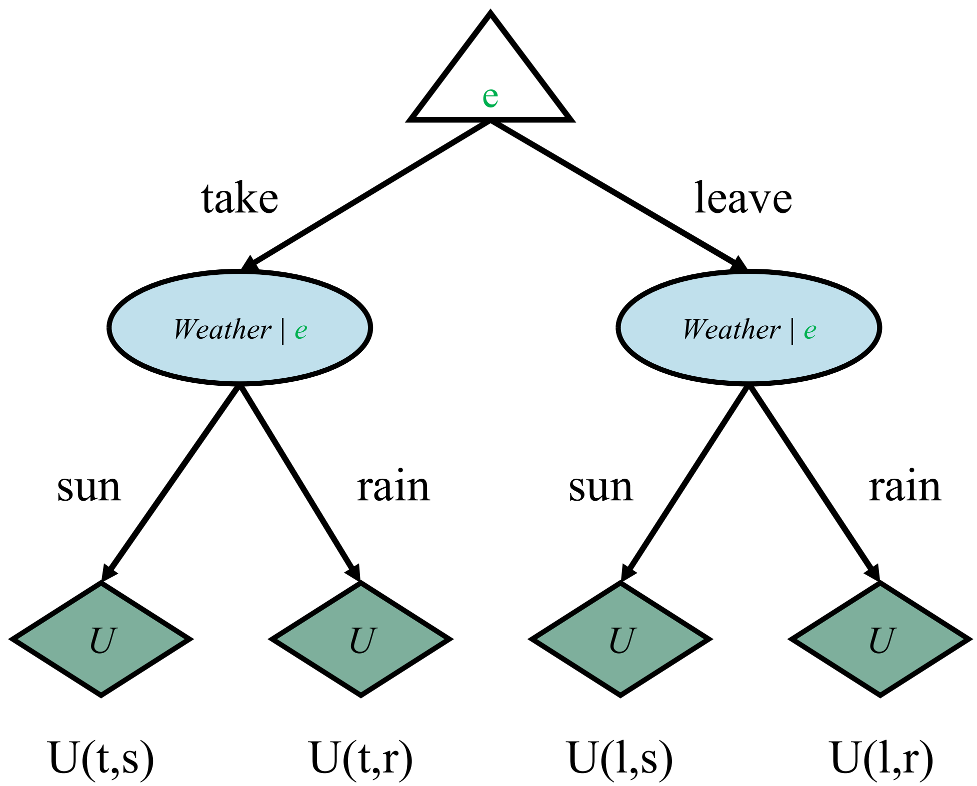
$$\mathrm{MEU}(a_{e_2} \mid e_2) = \mathrm{MEU}(\mathrm{leave} \mid F = sun) = 95$$

The value of additional evidence / information from F is:

$$\mathrm{VPI}(E_j) = \left(\sum_{e_j} \mathrm{P}(E_j = e_j) * \mathrm{MEU}(a_{e_j} \mid E_j = e_j)\right) - MEU(a)$$

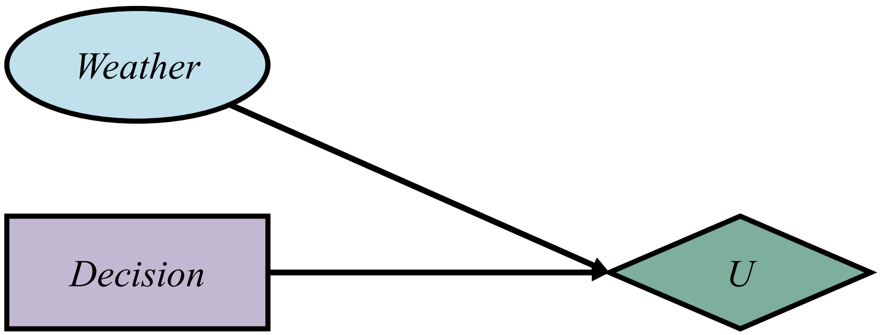
$$\mathrm{VPI}(F) = (\mathrm{P}(F = rain) * \mathrm{MEU}(take \mid F = rain) + \mathrm{P}(F = sun) * \mathrm{MEU}(\mathrm{leave} \mid F = sun)) - MEU(leave) = (0.41 * 53 + 0.59 * 95) - 70 = 7.78$$

Outcome tree



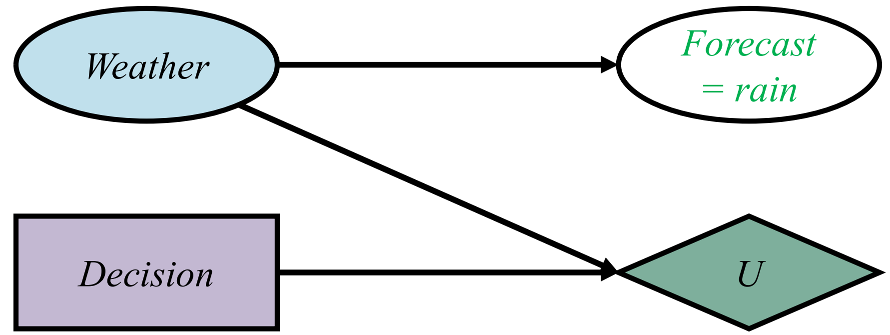
Decision Networks: Example

Decision: leave umbrella



$$\text{EU}(\text{leave}) = 70$$

Decision: take umbrella given rain



$$\text{EU}(\text{take given rain forecast}) = 53$$

The value of best action α without additional evidence

$$\text{MEU}(\alpha) = \text{MEU}(\text{leave}) = 70$$

With evidence information ($\mathrm{E_j} = \mathrm{e_j}$) given by Forecast:

$$\text{MEU}(a_{e_1} \mid e_1) = \text{MEU}(\text{take} \mid F = rain) = 53$$

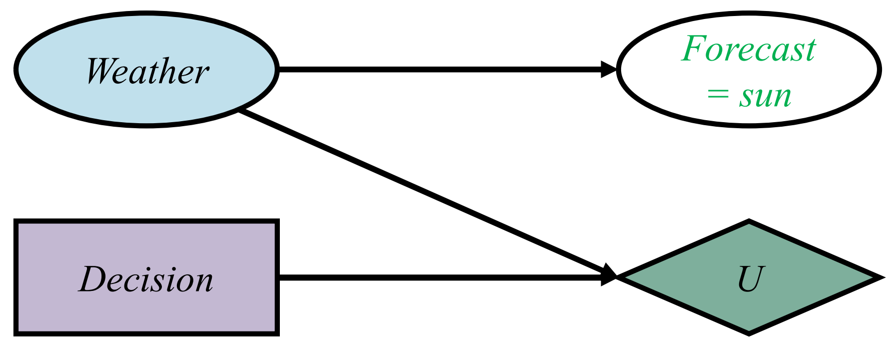
$$\text{MEU}(a_{e_2} \mid e_2) = \text{MEU}(\text{leave} \mid F = sun) = 95$$

The value of additional evidence / information from F is:

$$\text{VPI}(E_j) = \left(\sum_{e_j} \text{P}(E_j = e_j) * \text{MEU}(a_{e_j} \mid E_j = e_j)\right) - MEU(\alpha)$$

$$\text{VPI}(F) = (\text{P}(F = rain) * \text{MEU}(take \mid F = rain) + \text{P}(F = sun) * \text{MEU}(\text{leave} \mid F = sun)) - MEU(leave) = (0.41 * 53 + 0.59 * 95) - 70 = 7.78$$

Decision: leave umbrella given sun



$$\text{EU}(\text{leave given sun forecast}) = 95$$

VPI Properties

Given a decision network with possible observations E_j (sources of new information / evidence):

- **The expected value of information is nonnegative:**

$$\forall_\mathrm{j}\, \mathrm{VPI}(E_j) \geq 0$$

- **VPI is not additive:**

$$\mathrm{VPI}(E_j, E_k) \neq \mathrm{VPI}(E_j) + \mathrm{VPI}(E_k)$$

- **VPI is order-independent:**

$$\mathrm{VPI}(E_j,\ E_k) = VPI(E_j) + VPI(\ E_k \mid E_j) = VPI(E_k) + VPI(\ E_j \mid E_k) = VPI(E_k\ ,\ E_j\ )$$

Information Gathering Agent

function INFORMATION-GATHERING-AGENT(*percept*) **returns** an *action*
  persistent: D, a decision network

  integrate *percept* into D
  $j \leftarrow$ the value that maximizes $VPI(E_j)\ /\ C(E_j)$
  if $VPI(E_j)\ >\ C(E_j)$
    then return $Request(E_j)$
  else return the best action from D

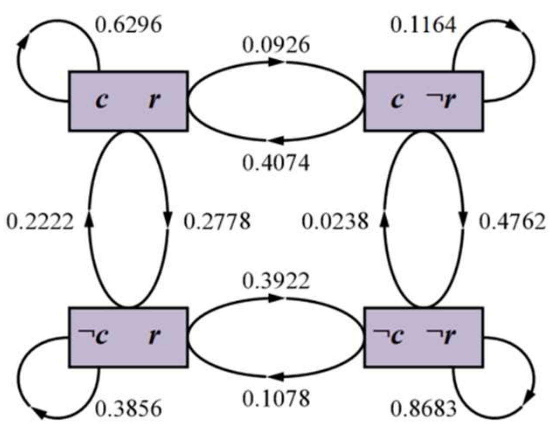
Inference in Temporal Models

Agents and Belief State

Sensor Model -> Belief State

S1: A = 4, B = 7, C = 0 → Plan X
S2: A = 4, B = 7, C = 1 → Plan Y
S3: A = 4, B = 7, C = 2 → Plan X
S4: A = 4, B = 7, C = 3 → Plan Z

Transition Model



Agent

Sensor

Agent can consult its internal representation of the world / environment to choose action

Actuator

A = 4 B = 7

ACTION(S)

Partially observable

A = 4

B = 7

C = 0

Environment

Assume: $D_C = \{0,1,2,3\}$

State Space and Transition Model

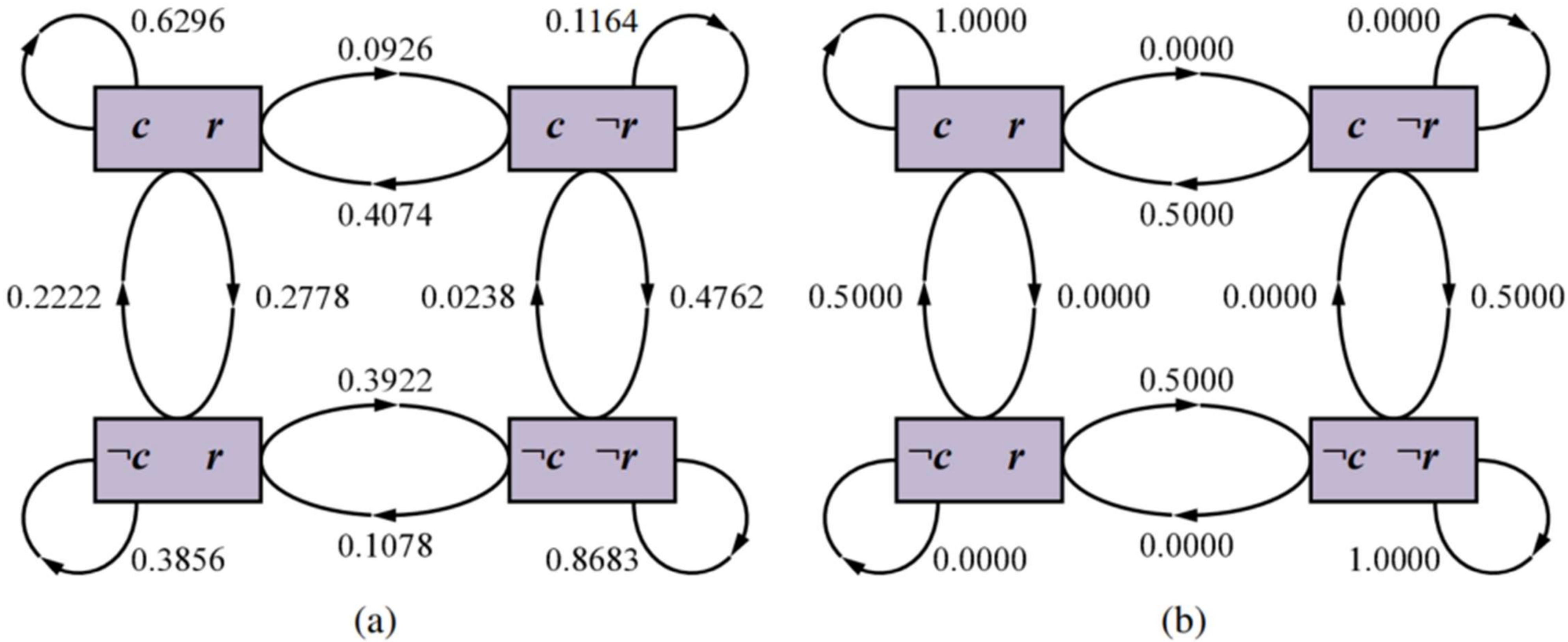


Figure 13.21 (a) The states and transition probabilities of the Markov chain for the query $\mathbf{P}(Rain \mid Sprinkler = true, WetGrass = true)$. Note the self-loops: the state stays the same when *either* variable is chosen and then resamples the same value it already has. (b) The transition probabilities when the CPT for *Rain* constrains it to have the same value as *Cloudy*.

Transition and Emission Probabilities

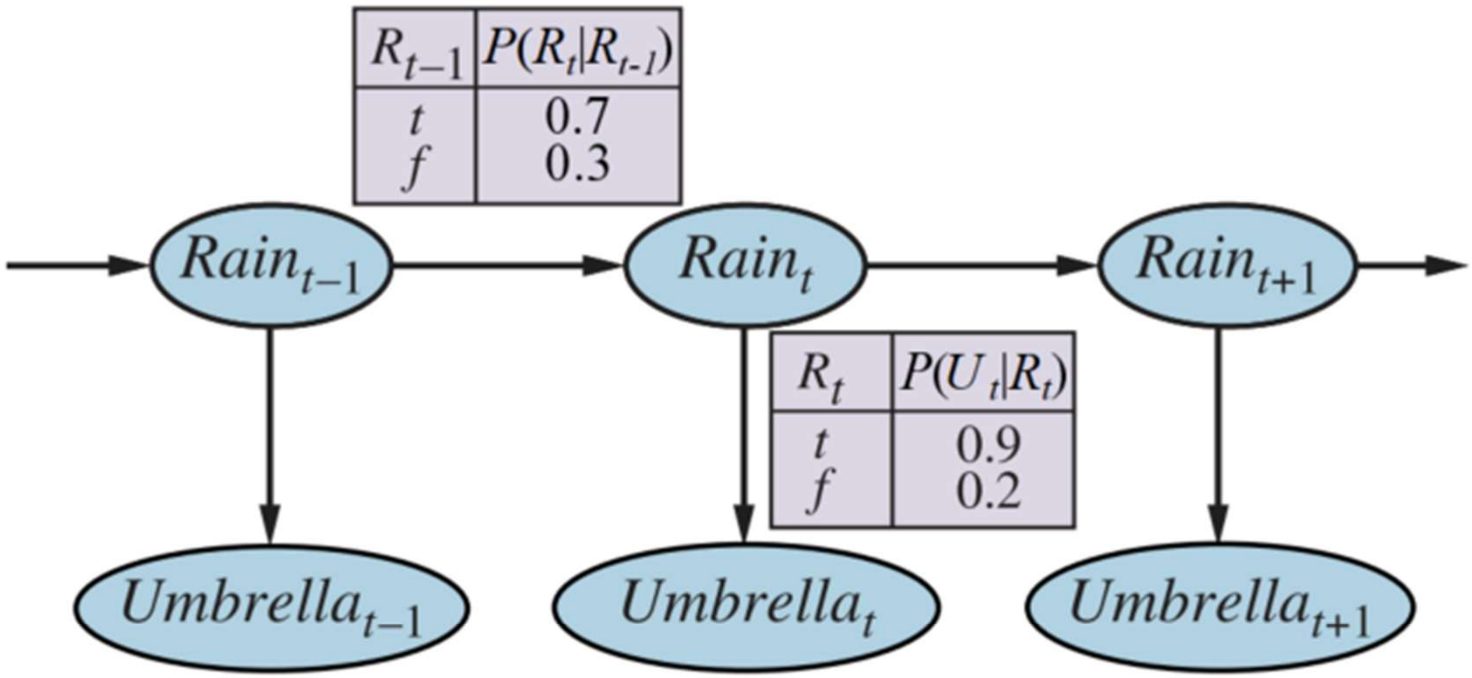


Figure 14.2 Bayesian network structure and conditional distributions describing the umbrella world. The transition model is $\mathbf{P}(Rain_t \mid Rain_{t-1})$ and the sensor model is $\mathbf{P}(Umbrella_t \mid Rain_t)$.

Inference in Temporal Models

- **Filtering / State Estimation**
  - **Current Belief State given Evidence/Percept so far**
- **Prediction**
  - **Future Belief State given Evidence/Percept so far**
- **Smoothing**
  - **Past Belief State given Evidence/Percept so far**
- **Most likely explanation:**
  - **Use sequence of observations to find sequence of states that generated them**
- **Learning:**
  - **Learn the transition and sensor models based on observations (“emissions”)**

Andrey Markov's Work (early 1900s)

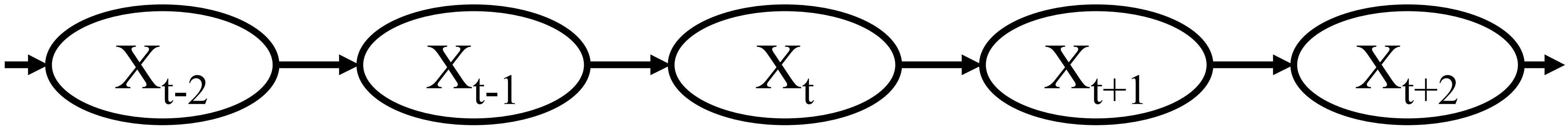
Sources: https://en.wikipedia.org/wiki/Markov_model
https://en.wikipedia.org/wiki/Andrey_Markov

Andrey Andreyevich Markov was a Russian mathematician **best known for his work on stochastic processes**. A primary subject of his research later became known as the **Markov chain**.

A **Markov model** is a model used to model systems in which **future states depend only on the current state, not on the events that occurred before it**.

Markov Process / Chain

Markov Process (Chain) is a random process that generates a sequence of states



Markov Process / Chain

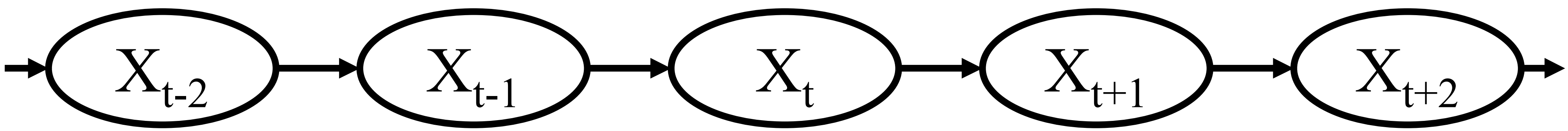
Markov Process (Chain) is a random process that generates a sequence of states

$$\rightarrow X_{t-2} \rightarrow X_{t-1} \rightarrow X_t \rightarrow X_{t+1} \rightarrow X_{t+2} \rightarrow$$

Bayesian Network?? Anyone?

Markov Process / Chain

Markov Process (Chain) is a random process that generates a sequence of states



Bayesian Network?? Anyone? Indeed!

$$P(X_{t+1} \mid X_t, X_{t-2}, X_{t-2}) = P(X_t \mid Parents(X_t)) = P(X_{t+1} \mid X_t)$$

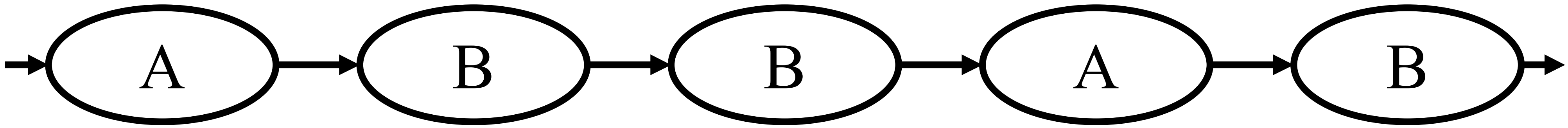
$$P(X_{t+1} \mid X_t, X_{t-2}, X_{t-2}) = P(X_{t+1} \mid X_t)$$

Markov ASSUMPTION

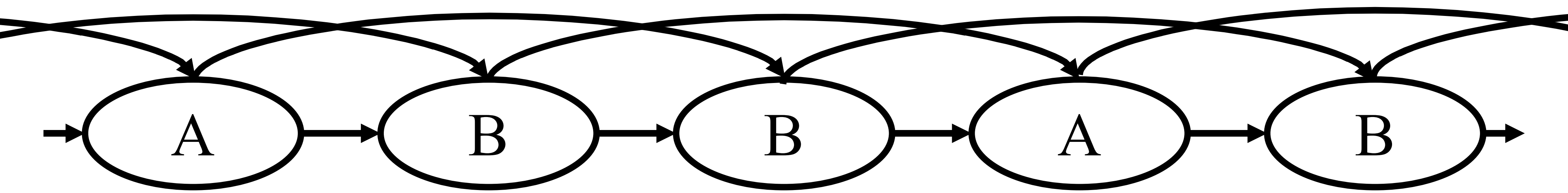
Markov Process / Chain: Possible?

Markov Process (Chain) is a random process that generates a sequence of states

First-order Markov Assumption:

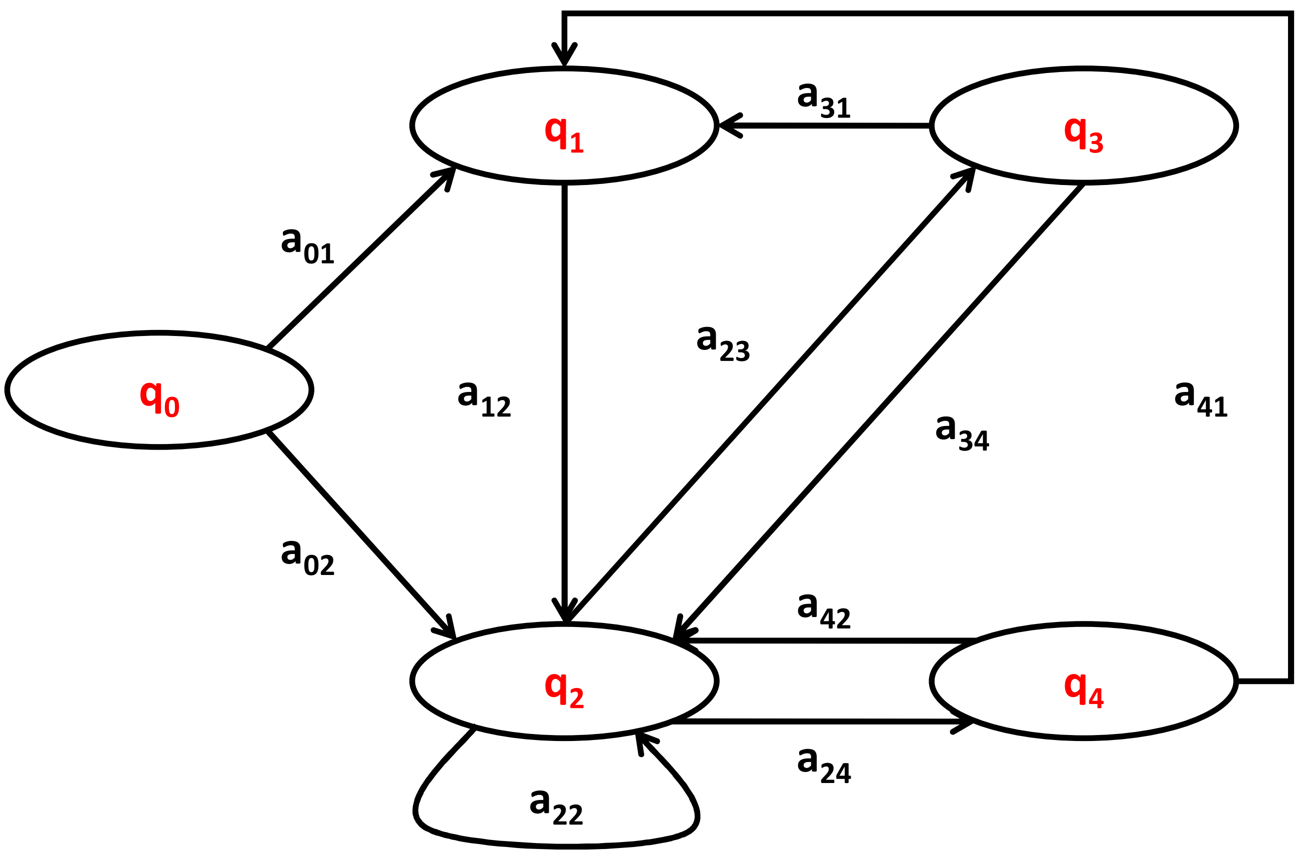


Second-order Markov Assumption:



Hidden Markov Model

Hidden Markov Model



Transition probability matrix A						
	q_0	q_1	q_2	q_3	q_4	Notes
q_0	$a_{0,0}$	$a_{0,1}$	$a_{0,2}$	$a_{0,3}$	$a_{0,4}$	row sum = 1
q_1	$a_{1,0}$	$a_{1,1}$	$a_{1,2}$	$a_{1,3}$	$a_{1,4}$	row sum = 1
q_2	$a_{2,0}$	$a_{2,1}$	$a_{2,2}$	$a_{2,3}$	$a_{2,4}$	row sum = 1
q_3	$a_{3,0}$	$a_{3,1}$	$a_{3,2}$	$a_{3,3}$	$a_{3,4}$	row sum = 1
q_4	$a_{4.,0}$	$a_{4,1}$	$a_{4,2}$	$a_{4,3}$	$a_{4,4}$	row sum = 1

HMMs are specified with:

- **A set of N states:**

  $Q = \{q_1, q_2, \ldots, q_N\}$

- **A transition probability matrix A, where each $a_{i,j}$ represents the probability of moving from state q_i to state q_j**

- **A sequence of T observations O:**

  $O = o_1, o_2, \ldots, o_T$

- **A sequence of observation likelihoods (emission probabilities): probability of observation o_t being generated by a state q_i:**

  $B = b_i(o_t)$

- **Special start (<s>) and end (final: not here) states**

  q_0 **and** q_F

Hidden Markov Models: Decoding

The task of determining which sequence of variables is the underlying source of some sequence of observations is called the decoding:

Given as input an HMM $\alpha = (A, B)$ *and a sequence of observations* $o_1, o_2, \ldots, o_T$ *find the most probable sequence of states* $q_1, q_2, \ldots, q_T$.

or in our case:

Given as input an HMM $\alpha = (A, B)$ *and a sequence of* ***words*** $w_1, w_2, \ldots, w_T$ *find the most probable sequence of* ***tags/states*** $C_1, C_2, \ldots, C_T$.

A - transition probabilities matrix

B - emission probabilities matrix

HMM and Viterbi Algorithm: POS Tagging Example

POS Tagging: General Approach

Given a sequence of words (a "sentence"):

$$w_1, w_2, w_3, \ldots, w_T$$

there is going to be a corresponding sequence of lexical categories:

$$C_1, C_2, C_3, \ldots, C_T$$

What is most likely sequence of categories?

POS Tagging: General Approach

The probability we are looking for

$$P(C_1, C_2, C_3, ..., C_T \mid w_1, w_2, w_3, ..., w_T)$$

will require a lot of data, which we most likely won't have.

We can use Bayes' Theorem:

$$\boldsymbol{P(C_1, C_2, C_3, \ldots, C_T \mid w_1, w_2, w_3, \ldots, w_T) =}$$

$$\boldsymbol{= \frac{P(w_1, w_2, w_3, \ldots, w_T \mid C_1, C_2, C_3, \ldots, C_T) * P(C_1, C_2, C_3, \ldots, C_T)}{P(w_1, w_2, w_3, \ldots, w_T)}}$$

POS Tagging: General Approach

Maximizing :

$$P(C_1, C_2, C_3, ..., C_T \mid w_1, w_2, w_3, ..., w_T)$$

in practice means maximizing the numerator:

$$\frac{P(w_1, w_2, w_3, ..., w_T \mid C_1, C_2, C_3, ..., C_T) * P(C_1, C_2, C_3, ..., C_T)}{P(w_1, w_2, w_3, ..., w_T)}$$

as denominator $P(w_1, w_2, w_3, ..., w_T)$ will not change:

POS Tagging: General Approach

Estimating:

$$P(w_1, w_2, w_3, \ldots, w_T \mid C_1, C_2, C_3, \ldots, C_T) * P(C_1, C_2, C_3, \ldots, C_T)$$

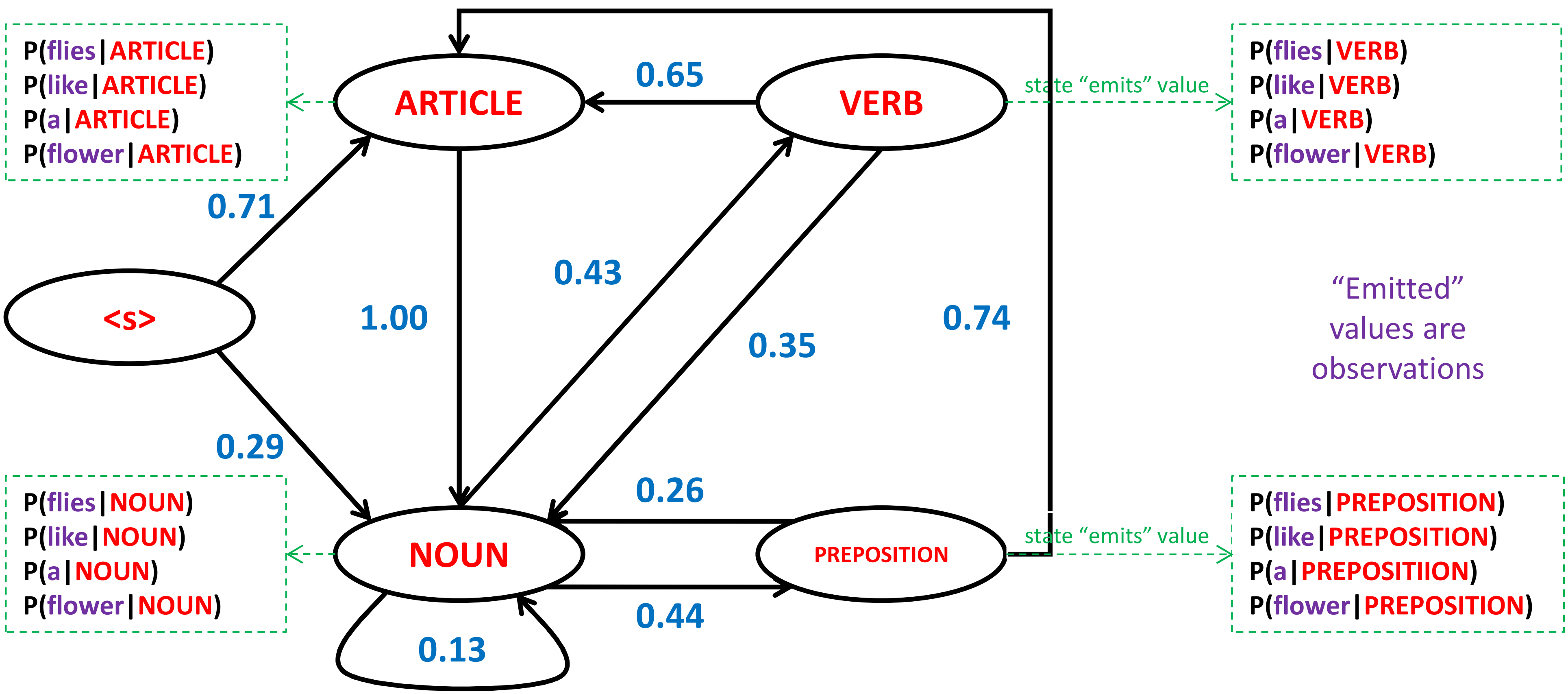
Approximate it with N-grams (here bigrams):

$$P(C_1, C_2, C_3, \ldots, C_T) = \prod_{i=1}^{T} P(C_i \mid all\ categories\ preceding\ C_i)$$

$$P(C_1, C_2, C_3, \ldots, C_T) \cong \prod_{i=1}^{T} P(C_i \mid C_{i-1})$$

$$P(w_1, w_2, w_3, \ldots, w_T \mid C_1, C_2, C_3, \ldots, C_T) \cong \prod_{i=1}^{T} P(w_i \mid C_i)$$

Hidden Markov Model



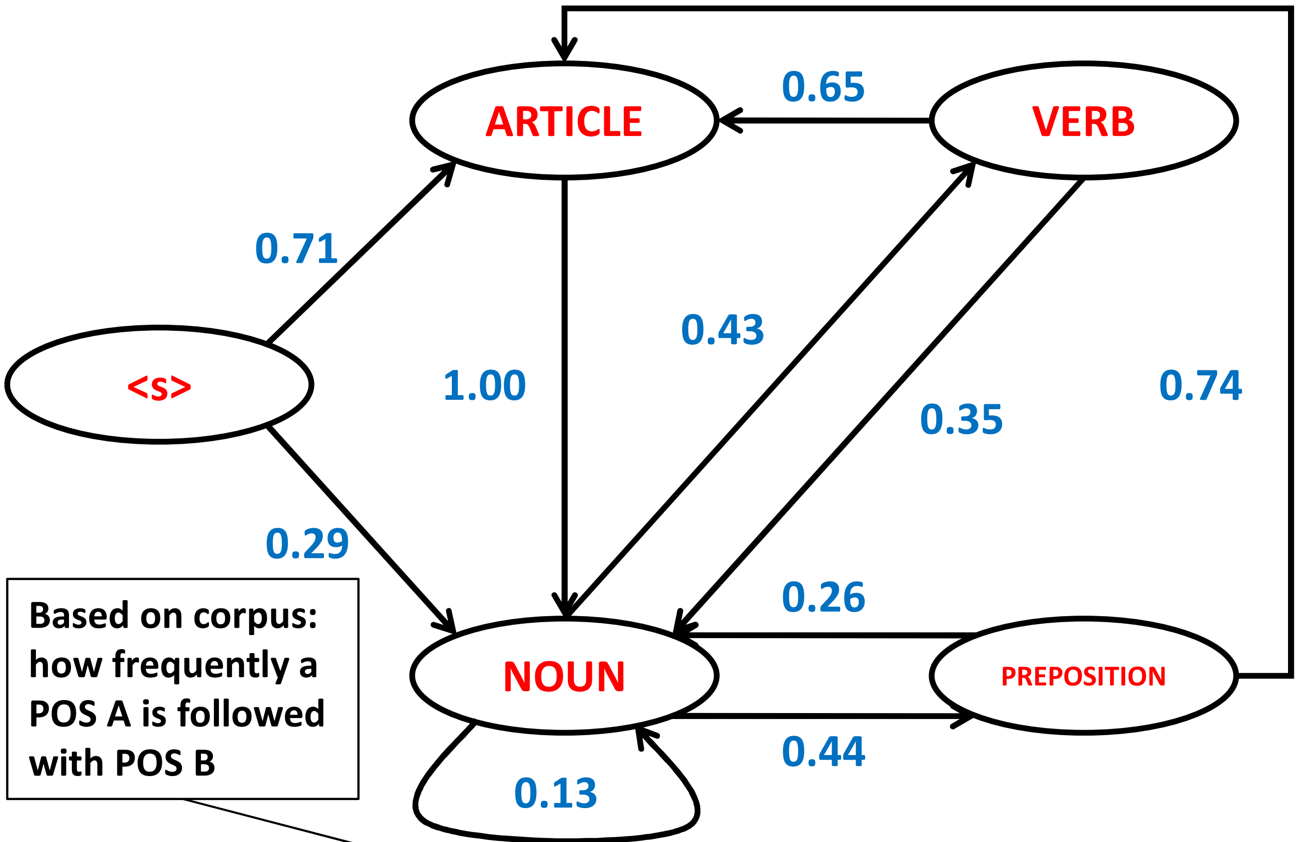
Transition probability matrix

	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Emission probability matrix

	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Hidden Markov Model



Based on corpus: how frequently a POS A is followed with POS B

Based on corpus: how frequently a word X is tagged with Y

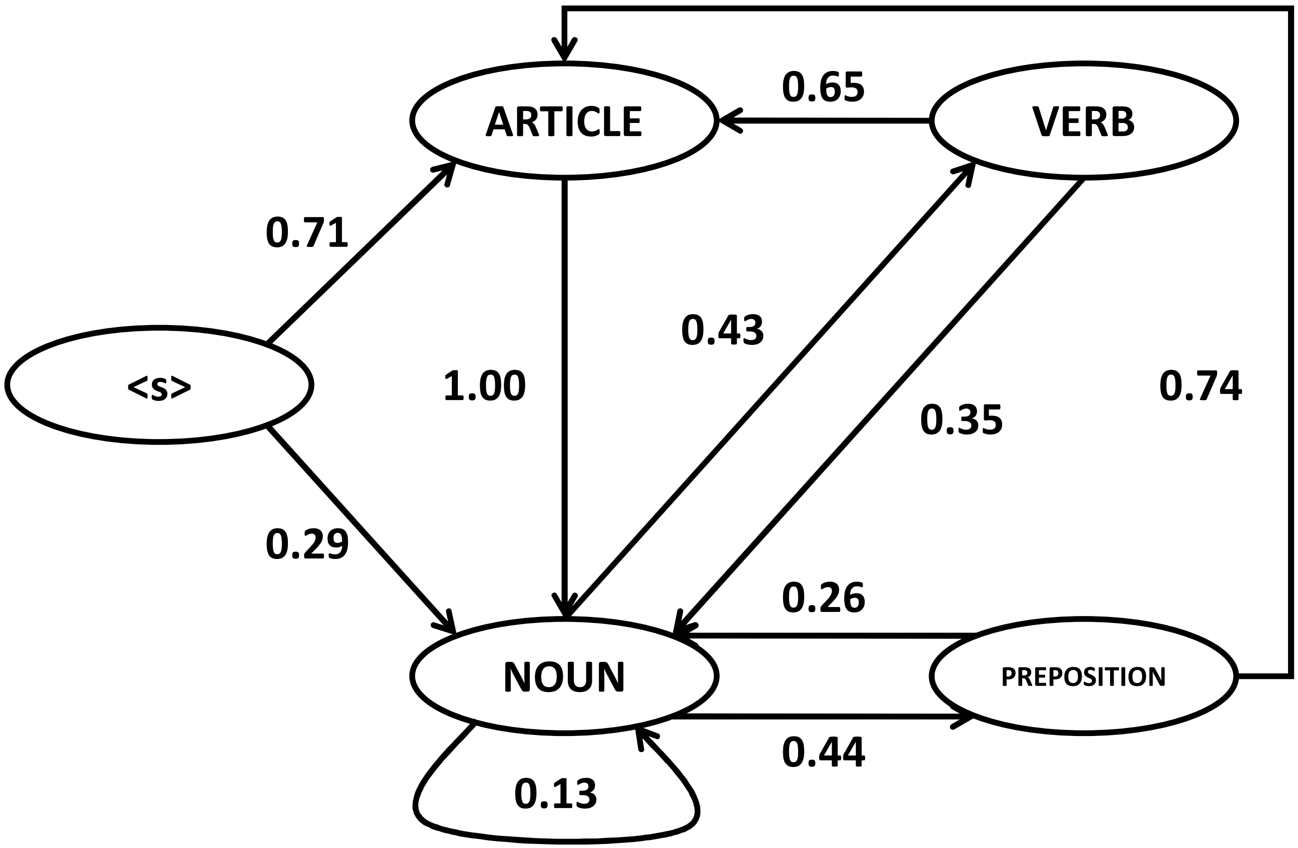
Transition probability matrix

	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Emission probability matrix

	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Hidden Markov Model



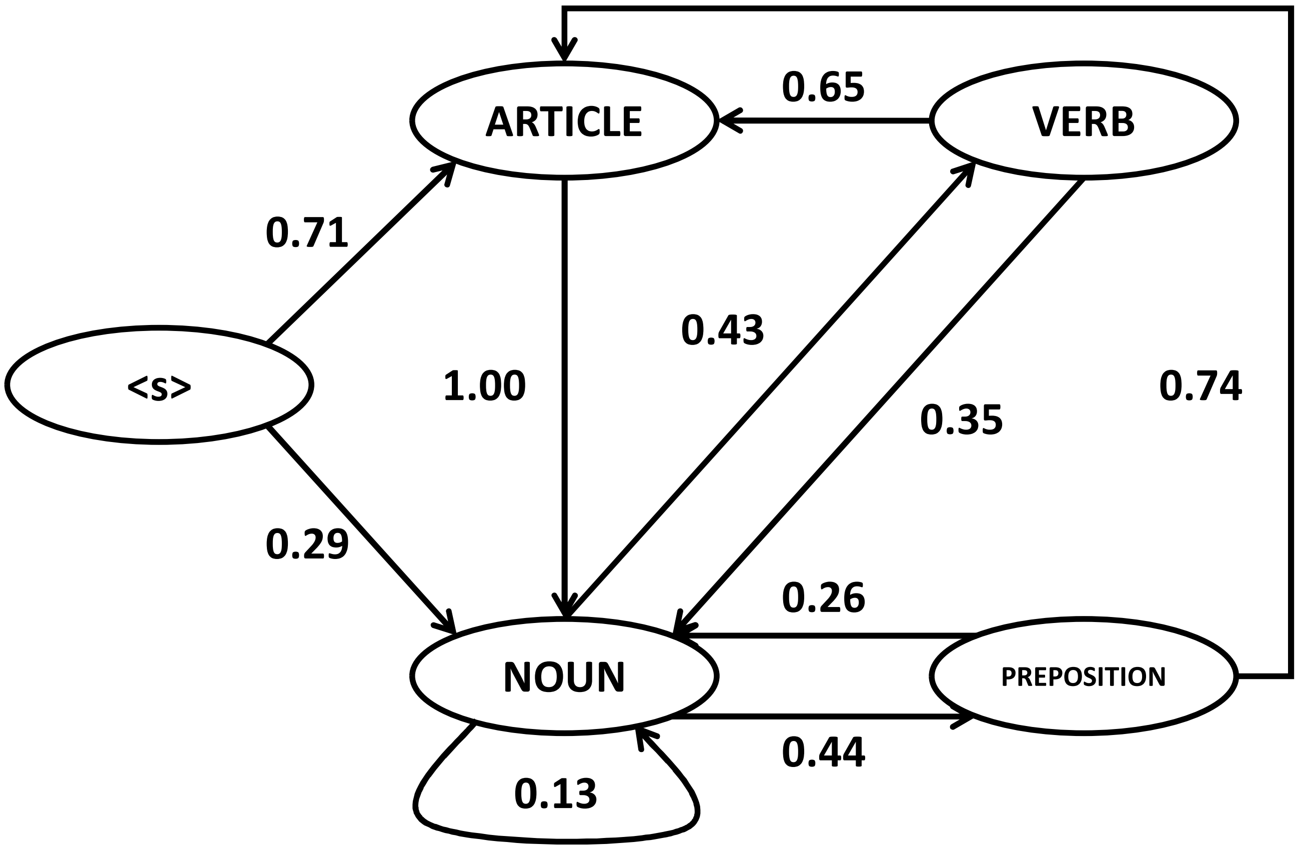
P(Bigram)	Estimate
P(ARTICLE|<S>)	0.71
P(NOUN|<S>)	0.29
P(NOUN|ARTICLE)	1.00
P(VERB|NOUN)	0.43
P(NOUN|NOUN)	0.13
P(PREPOSITION|NOUN)	0.44
P(NOUN|VERB)	0.35
P(ARTICLE|VERB)	0.65
P(ARTICLE|PREPOSITION)	0.74
P(NOUN|PREPOSITION)	0.26

Consider a following sequence of categories (tags):

<s>, ARTICLE, NOUN, VERB, NOUN

What's the probability of its occurence in our synthetic corpus?

Hidden Markov Model (HMM)

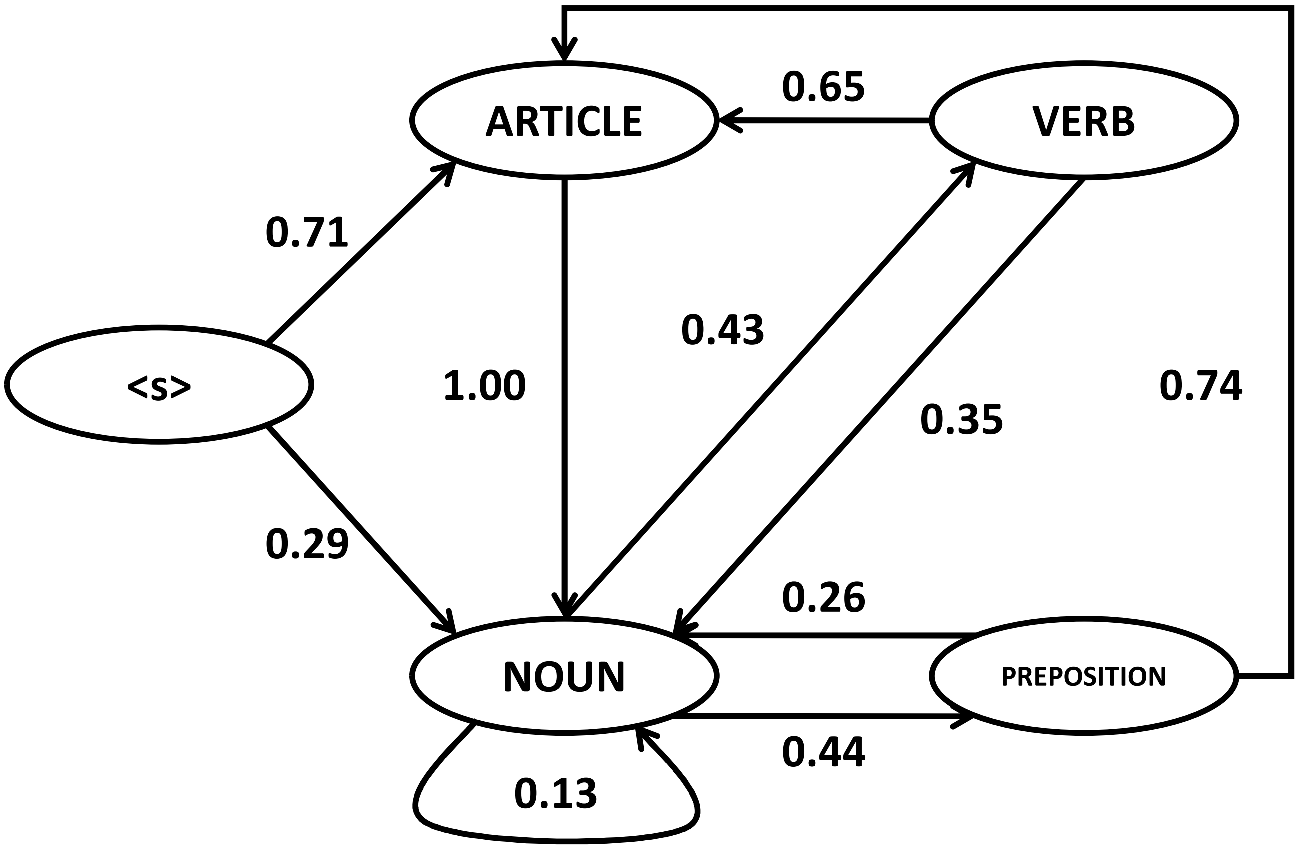


P(Bigram)	Estimate
P(ARTICLE|<S>)	0.71
P(NOUN|<S>)	0.29
P(NOUN|ARTICLE)	1.00
P(VERB|NOUN)	0.43
P(NOUN|NOUN)	0.13
P(PREPOSITION|NOUN)	0.44
P(NOUN|VERB)	0.35
P(ARTICLE|VERB)	0.65
P(ARTICLE|PREPOSITION)	0.74
P(NOUN|PREPOSITION)	0.26

The word "Hidden" in Hidden Markov Model means that for a specific sequence (of words) it is unclear what state the model is in.

The word *flies* could be generated from state NOUN and state VERB.

Hidden Markov Model

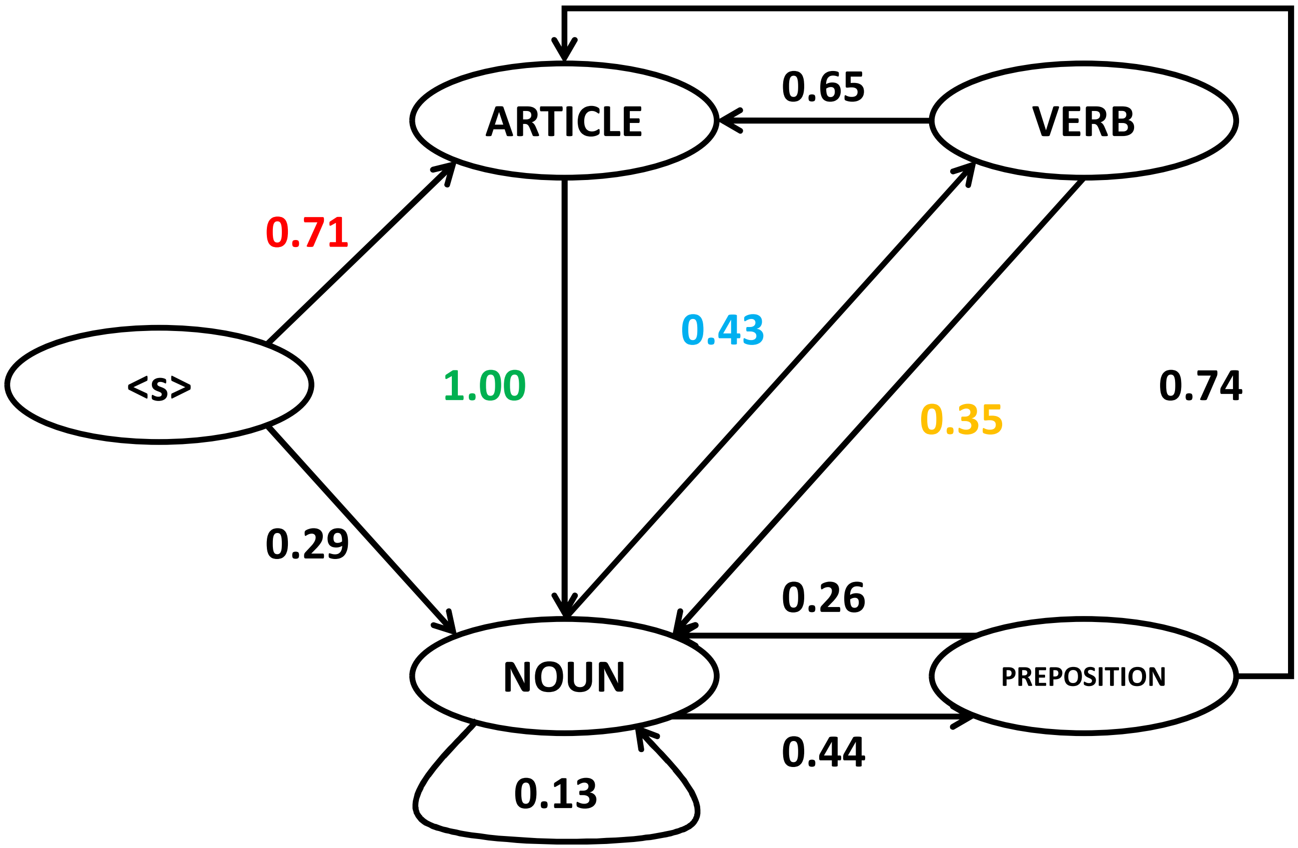


P(Bigram)	Estimate
P(ARTICLE|<S>)	0.71
P(NOUN|<S>)	0.29
P(NOUN|ARTICLE)	1.00
P(VERB|NOUN)	0.43
P(NOUN|NOUN)	0.13
P(PREPOSITION|NOUN)	0.44
P(NOUN|VERB)	0.35
P(ARTICLE|VERB)	0.65
P(ARTICLE|PREPOSITION)	0.74
P(NOUN|PREPOSITION)	0.26

Probability of occurrence of a sequence of categories (tags):

$$P(C_1, C_2, C_3, \ldots, C_T) \cong \prod_{i=1}^{T} P(C_i \mid C_{i-1})$$

Hidden Markov Model



P(Bigram)	Estimate
P(ARTICLE|<S>)	0.71
P(NOUN|<S>)	0.29
P(NOUN|ARTICLE)	1.00
P(VERB|NOUN)	0.43
P(NOUN|NOUN)	0.13
P(PREPOSITION|NOUN)	0.44
P(NOUN|VERB)	0.35
P(ARTICLE|VERB)	0.65
P(ARTICLE|PREPOSITION)	0.74
P(NOUN|PREPOSITION)	0.26

Probability of occurrence of a sequence of categories (tags):

$$P(\langle s\rangle, \text{ARTICLE}, \text{NOUN}, \text{VERB}, \text{NOUN}) =$$

$$\cong P(\text{ART}|\langle s\rangle) * P(\text{N}|\text{ART}) * P(\text{V}|\text{N}) * P(\text{N}|\text{V}) = 0.71 * 1.00 * 0.43 * 0.35 = 0.107$$

Example

Given our synthetic corpus, what is the most like sequence of categories (tags) corresponding to a sentence:

Flies like a flower

We need to maximize:

$$P(w_1, w_2, w_3, \ldots, w_T \mid C_1, C_2, C_3, \ldots, C_T) * P(C_1, C_2, C_3, \ldots, C_T) \cong$$

$$\cong \prod_{i=1}^{T} P(w_i \mid C_i) * P(C_i \mid C_{i-1})$$

POS Tagging: Simple Tagset

Let's assume we have a simple tagset:

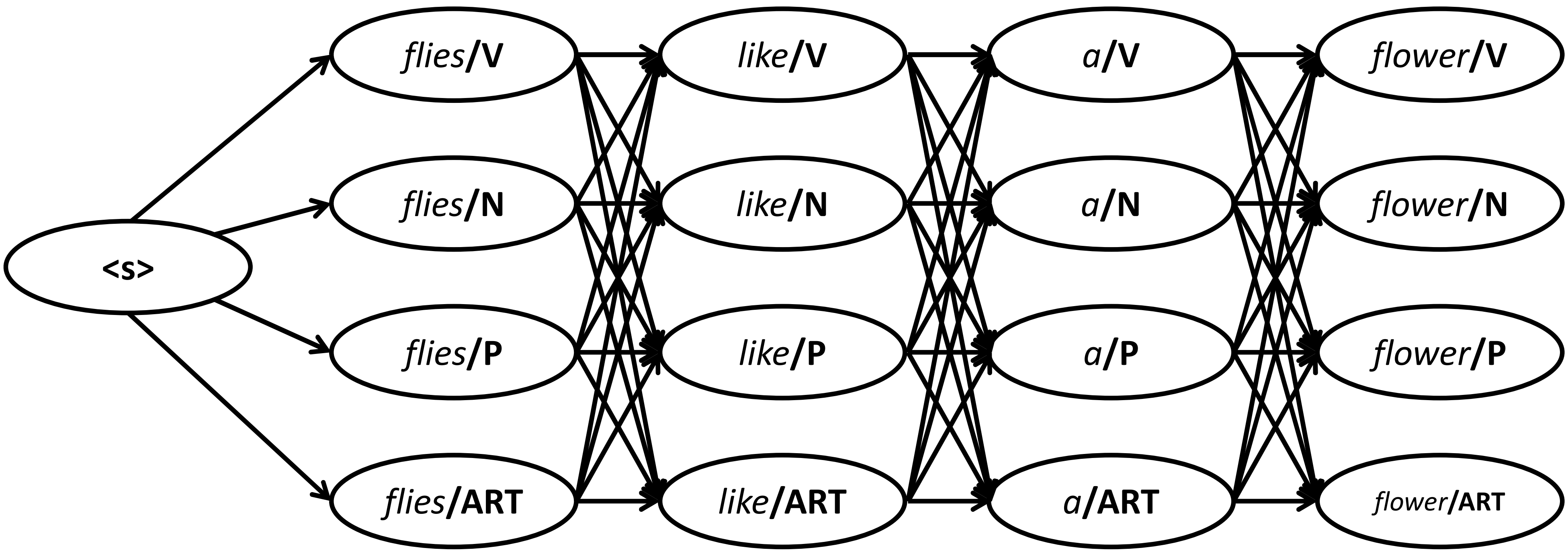
- **N - NOUN**
- **V - VERB**
- **ART - ARTICLE**
- **P - PREPOSITION**

and some synthetic corpus.

Example sentence:

Flies like a flower

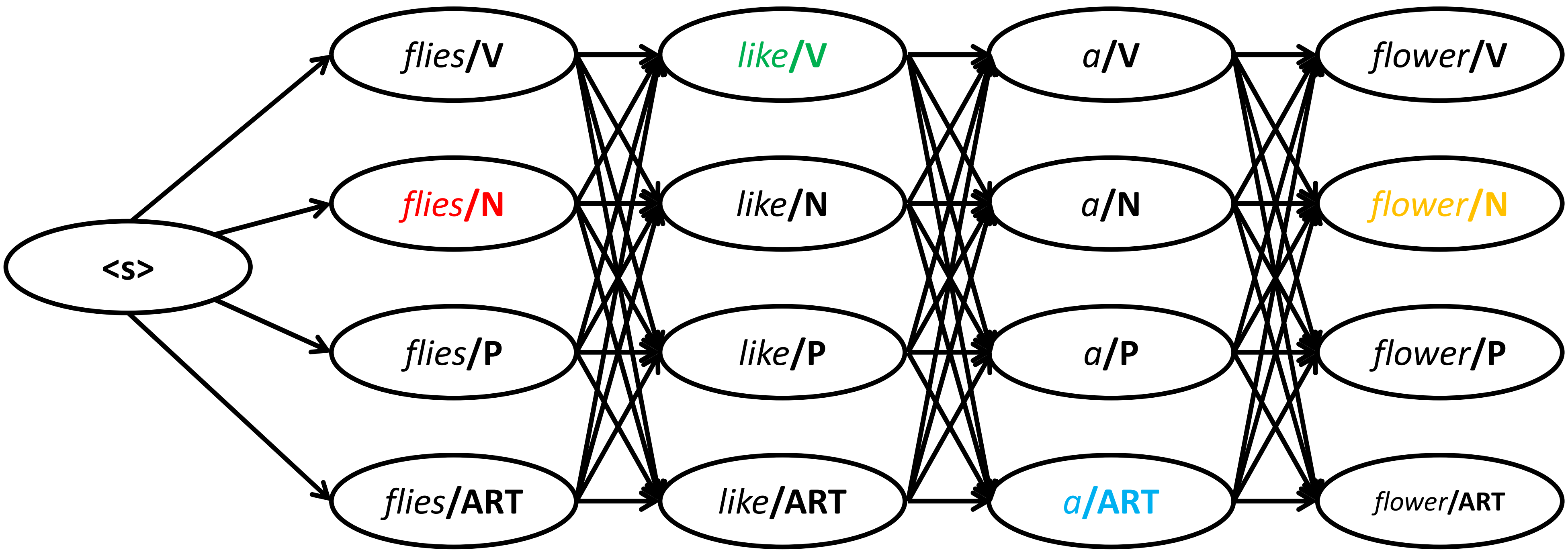
Example: All Possible Sequences



Every sequence can be assigned a probability:

$$P(w_1, w_2, w_3, \ldots, w_T \mid C_1, C_2, C_3, \ldots, C_T) \cong \prod_{i=1}^{T} P(w_i \mid C_i)$$

Example: All Possible Sequences



Every sequence can be assigned a probability:

$$\prod_{i=1}^{T} P(w_i \mid C_i) = P(flies|N) * P(like|V) * P(a|ART) * P(flower|N)$$

POS Tagging: General Approach

Estimations with corpus counts:

$$P(C_i = VERB \mid C_{i-1} = NOUN) = \frac{Count\ (NOUN\ at\ position\ i-1\ and\ VERB\ at\ i)}{Count(NOUN\ at\ position\ i-1)}$$

Sample bigram probabilities from our synthetic corpus:

Category	Count at i	Pair	Count at i,i+1	P(Bigram)	Estimate
<s>	300	<s>, ARTICLE	213	P(ARTICLE|<S>)	0.71
<s>	300	<s>, NOUN	87	P(NOUN|<S>)	0.29
ARTICLE	558	ARTICLE, NOUN	558	P(NOUN|ARTICLE)	1.00
NOUN	833	NOUN, VERB	358	P(VERB|NOUN)	0.43
NOUN	833	[illegible]	108	P(NOUN|NOUN)	0.13
NOUN	833	NOUN, PREPOSITION	366	P(PREPOSITION|NOUN)	0.44
VERB	300	VERB, NOUN	75	P(NOUN|VERB)	0.35
VERB	300	VERB, ARTICLE	194	P(ARTICLE|VERB)	0.65
PREPOSITION	307	PREPOSITION, ARTICLE	226	P(ARTICLE|PREPOSITION)	0.74
PREPOSITION	307	PREPOSITION, NOUN	81	P(NOUN|PREPOSITION)	0.26

Transition probabilities

Synthetic Corpus: Word/Tag Counts

Summary of selected word counts in the synthetic corpus:

Word/Tag	N	V	ART	P	TOTAL
flies	21	23	0	0	44
fruit	49	5	1	0	55
like	10	30	0	21	61
a	1	0	201	0	202
the	1	0	300	2	303
flower	53	15	0	0	68
flowers	42	16	0	0	58
birds	64	1	0	0	65
others	592	210	56	284	1142
TOTAL	833	300	558	307	1998

From the table we can calculate lexical generation probabilities P(w|C) estimates:

P(*the*|ART) = 300/558 = 0.54

P(*flies*|N) = 21/833 = 0.025

P(*flies*|V) = 23/300 = 0.076

P(*like*|V) = 30/300 = 0.1

P(*like*|P) = 21/307 = 0.068

P(*a*|ART) = 201/558 = 0.36

P(*a*|N) = 1/833 = 0.001

P(*flower*|N) = 53/833 = 0.063

P(*flower*|V) = 15/300 = 0.05

P(*like*|N) = 10/833 = 0.012

Emission probabilities

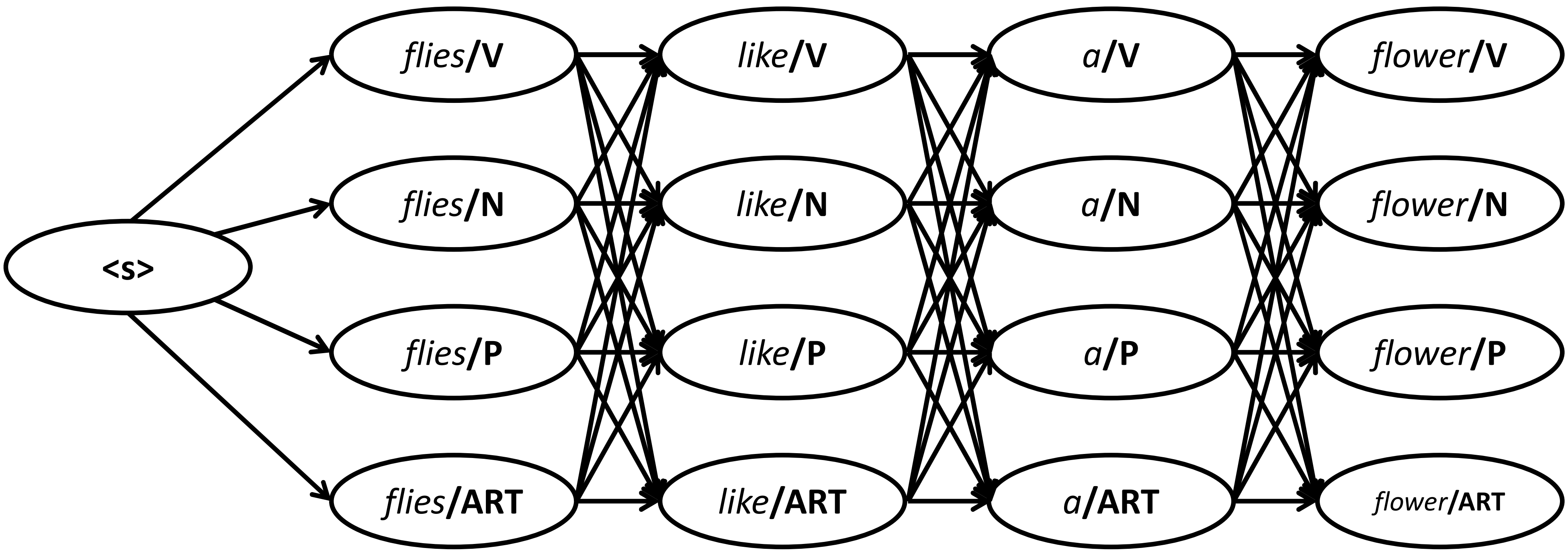
Viterbi Algorithm

Sample Tagged Sentence

There/PRO were/VERB 70/NUM children/NOUN there/ADV ./PUNC

Preliminary/ADJ findings/NOUN were/AUX reported/VERB in/ADP today/NOUN ’s/PART New/PROPN England/PROPN Journal/PROPN of/ADP Medicine/PROPN

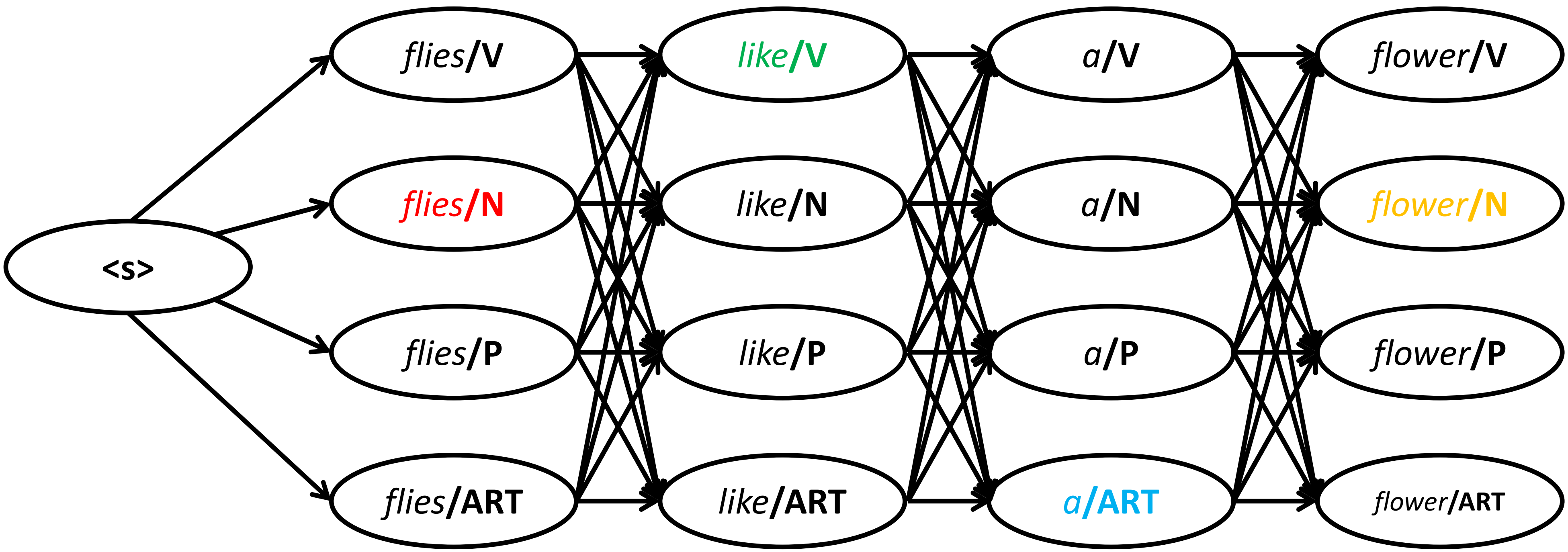
Example: All Possible Sequences



Every sequence can be assigned a probability:

$$P(w_1, w_2, w_3, \ldots, w_T \mid C_1, C_2, C_3, \ldots, C_T) \cong \prod_{i=1}^{T} P(w_i \mid C_i)$$

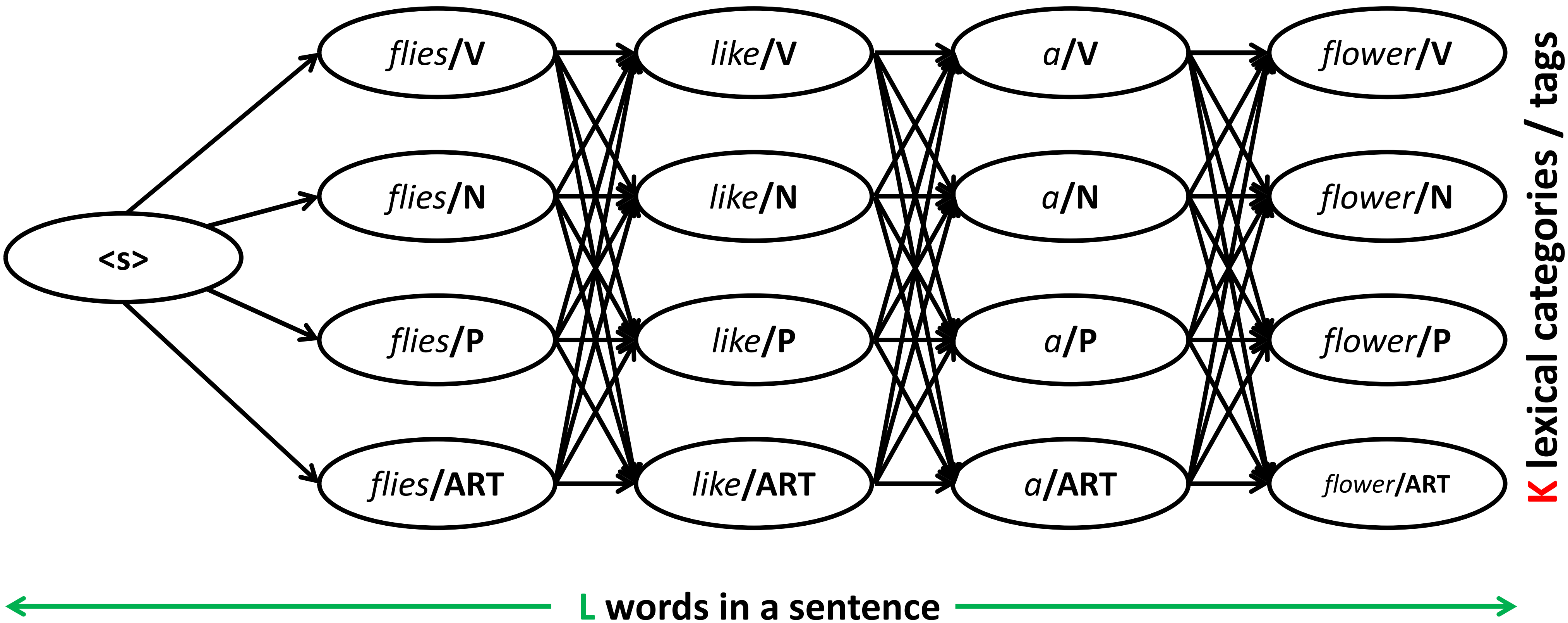
Example: Best Option



Best option will be:

$$\prod_{i=1}^{T} P(w_i \mid C_i) = P(flies|N) * P(like|V) * P(a|ART) * P(flower|N)$$

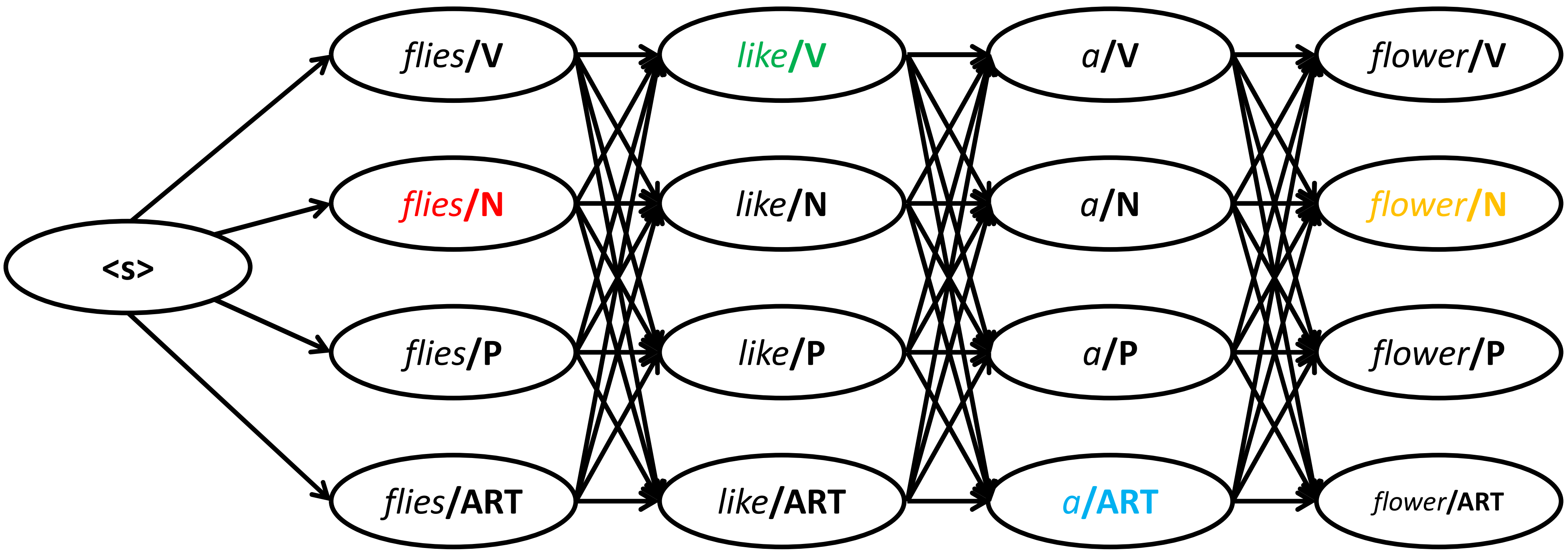
Example: All Possible Sequences



Brute force approach time complexity: O(K^L)

$K = 20$, $L = 10 \rightarrow 20^{10} = 10240000000000$

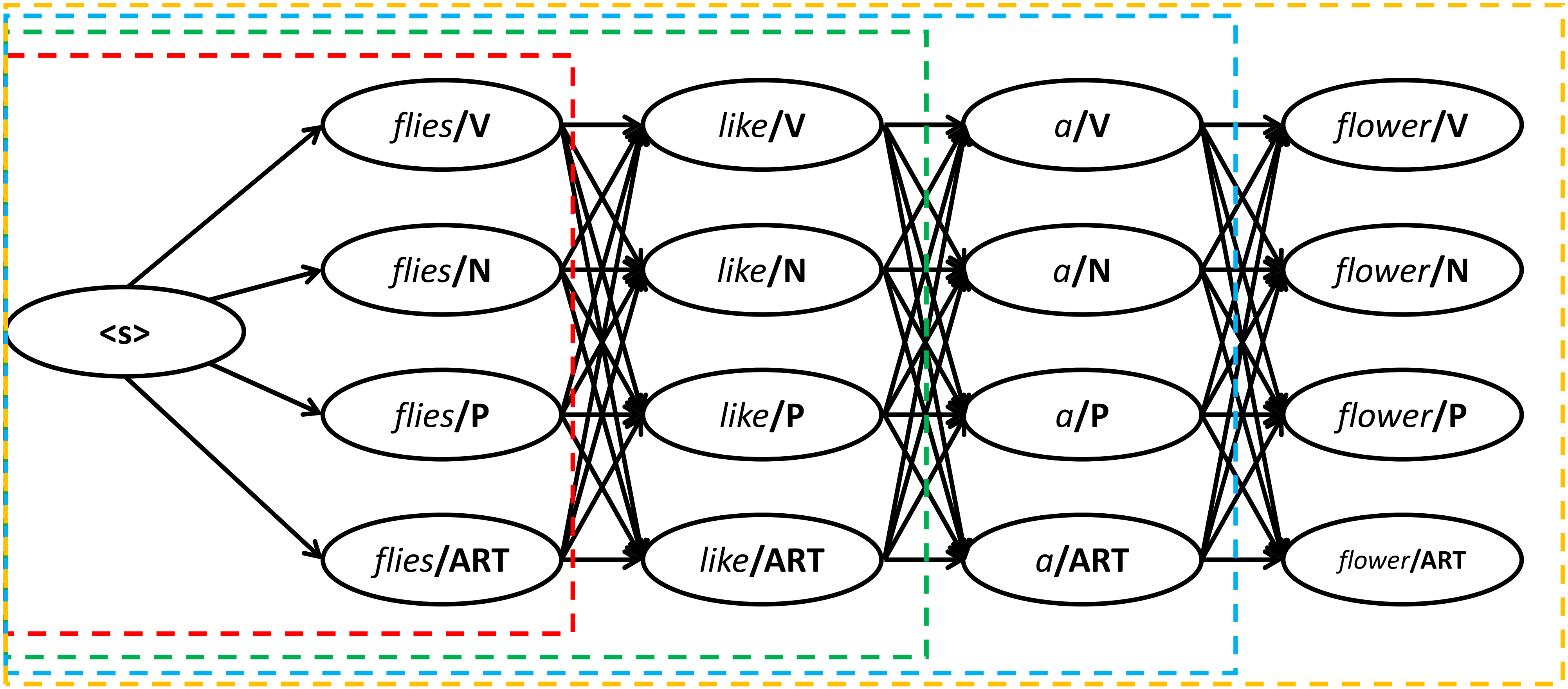
Example: Best Option



How can we efficiently find:

$$\prod_{i=1}^{T} P(w_i \mid C_i) = P(flies|N) * P(like|V) * P(a|ART) * P(flower|N)$$

Viterbi Algorithm: the Idea



Maximizing [ ] means maximizing [ ], [ ], and [ ].

In other words: maximize P() for all "sub-sentences".

Viterbi Algorithm: Pseudocode

Input sentence

function VITERBI(*observations* of len T, *state-graph* of len N) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state s **from** 1 **to** N **do** ; initialization step
$\quad viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
$\quad backpointer[s,1] \leftarrow 0$
for each time step t **from** 2 **to** T **do** ; recursion step
$\quad$ **for** each state s **from** 1 **to** N **do**

$$viterbi[s,t] \leftarrow \max_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$$

$$backpointer[s,t] \leftarrow \operatorname*{argmax}_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$$

$$bestpathprob \leftarrow \max_{s=1}^{N} viterbi[s,T]$$; termination step

$$bestpathpointer \leftarrow \operatorname*{argmax}_{s=1}^{N} viterbi[s,T]$$; termination step

bestpath $\leftarrow$ the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Algorithm: Pseudocode

Hidden Markov Model

function VITERBI(*observations* of len *T*, *state-graph* of len *N*) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state *s* **from** 1 **to** *N* **do** ; initialization step
    $viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
    $backpointer[s,1] \leftarrow 0$
for each time step *t* **from** 2 **to** *T* **do** ; recursion step
  for each state *s* **from** 1 **to** *N* **do**
    $viterbi[s,t] \leftarrow \max_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
    $backpointer[s,t] \leftarrow \operatorname{argmax}_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
$bestpathprob \leftarrow \max_{s=1}^{N} viterbi[s,T]$; termination step
$bestpathpointer \leftarrow \operatorname{argmax}_{s=1}^{N} viterbi[s,T]$; termination step
bestpath ← the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Algorithm: Pseudocode

function VITERBI(*observations* of len T,*state-graph* of len N) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state s **from** 1 **to** N **do** ; initialization step
    $viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
    $backpointer[s,1] \leftarrow 0$
for each time step t **from** 2 **to** T **do** ; recursion step
  for each state s **from** 1 **to** N **do**
    $viterbi[s,t] \leftarrow \max_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
    $backpointer[s,t] \leftarrow \operatorname{argmax}_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
$bestpathprob \leftarrow \max_{s=1}^{N} viterbi[s,T]$; termination step
$bestpathpointer \leftarrow \operatorname{argmax}_{s=1}^{N} viterbi[s,T]$; termination step
bestpath ← the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Matrix (N x T): Structure

Rows: 1, 2, 3, ..., N-2, N-1, N

States: $Q = q_1, q_2, q_3, \ldots, q_{N-2}, q_{N-1}, q_N$

N	q_N	$V_{col}^{(row)}$	$V_{col}^{(row)}$	$V_{col}^{(row)}$	...	$V_{col}^{(row)}$	$V_{col}^{(row)}$	$V_{col}^{(row)}$
N-1	q_{N-1}	$V_{col}^{(row)}$	$V_{col}^{(row)}$	$V_{col}^{(row)}$	...	$V_{col}^{(row)}$	$V_{col}^{(row)}$	$V_{col}^{(row)}$
N-2	q_{N-2}	$V_{col}^{(row)}$	$V_{col}^{(row)}$	$V_{col}^{(row)}$	...	$V_{col}^{(row)}$	$V_{col}^{(row)}$	$V_{col}^{(row)}$
...	...	...	...	...	...	...	...	...
3	q_3	$V_{col}^{(row)}$	$V_{col}^{(row)}$	$V_{col}^{(row)}$	...	$V_{col}^{(row)}$	$V_{col}^{(row)}$	$V_{col}^{(row)}$
2	q_2	$V_{col}^{(row)}$	$V_{col}^{(row)}$	$V_{col}^{(row)}$	...	$V_{col}^{(row)}$	$V_{col}^{(row)}$	$V_{col}^{(row)}$
1	q_1	$V_{col}^{(row)}$	$V_{col}^{(row)}$	$V_{col}^{(row)}$	...	$V_{col}^{(row)}$	$V_{col}^{(row)}$	$V_{col}^{(row)}$
	q_0	o_1	o_2	o_3	...	o_{T-2}	o_{T-1}	o_T
		1	2	3		T-2	T-1	T

Observations:
$O = o_1, o_2, o_3, \ldots, o_{T-2}, o_{T-1}, o_T$

Columns / Time steps: 1, 2, 3, ..., T-2, T-1, T

---> Time

Viterbi Matrix (N x T): Structure

Rows: 1, 2, 3, …, N-2, N-1, N

States: $Q = q_1, q_2, q_3, \ldots, q_{N-2}, q_{N-1}, q_N$

N	q_N	$V_1^{(N)}$	$V_2^{(N)}$	$V_3^{(N)}$	…	$V_{T-2}^{(N)}$	$V_{T-1}^{(N)}$	$V_T^{(N)}$
N-1	q_{N-1}	$V_1^{(N-1)}$	$V_2^{(N-1)}$	$V_3^{(N-1)}$	…	$V_{T-2}^{(N-1)}$	$V_{T-1}^{(N-1)}$	$V_T^{(N-1)}$
N-2	q_{N-2}	$V_1^{(N-2)}$	$V_2^{(N-2)}$	$V_3^{(N-2)}$	…	$V_{T-2}^{(N-2)}$	$V_{T-1}^{(N-2)}$	$V_T^{(N-2)}$
…	…	…	…	…	…	…	…	…
3	q_3	$V_1^{(3)}$	$V_2^{(3)}$	$V_3^{(3)}$	…	$V_{T-2}^{(3)}$	$V_{T-1}^{(3)}$	$V_T^{(3)}$
2	q_2	$V_1^{(2)}$	$V_2^{(2)}$	$V_3^{(2)}$	…	$V_{T-2}^{(2)}$	$V_{T-1}^{(2)}$	$V_T^{(2)}$
1	q_1	$V_1^{(1)}$	$V_2^{(1)}$	$V_3^{(1)}$	…	$V_{T-2}^{(1)}$	$V_{T-1}^{(1)}$	$V_T^{(1)}$
	q_0	o_1	o_2	o_3	…	o_{T-2}	o_{T-1}	o_T
		1	2	3		T-2	T-1	T

Observations:
$O = o_1, o_2, o_3, \ldots, o_{T-2}, o_{T-1}, o_T$

Columns / Time steps: 1, 2, 3, …, T-2, T-1, T

---> Time

Viterbi Matrix (N x T): Structure

Rows: 1, 2, 3, ..., N-2, N-1, N

States: $Q = q_1, q_2, q_3, \ldots, q_{N-2}, q_{N-1}, q_N$

N	q_N	$V_1^{(N)}$	$V_2^{(N)}$	$V_3^{(N)}$	...	$V_{T-2}^{(N)}$	$V_{T-1}^{(N)}$	$V_T^{(N)}$
N-1	q_{N-1}	$V_1^{(N-1)}$	$V_2^{(N-1)}$	$V_3^{(N-1)}$	...	$V_{T-2}^{(N-1)}$	$V_{T-1}^{(N-1)}$	$V_T^{(N-1)}$
N-2	q_{N-2}	$V_1^{(N-2)}$	$V_2^{(N-2)}$	$V_3^{(N-2)}$	...	$V_{T-2}^{(N-2)}$	$V_{T-1}^{(N-2)}$	$V_T^{(N-2)}$
...	...	...	...	...	...	...	...	...
3	q_3	$V_1^{(3)}$	$V_2^{(3)}$	$V_3^{(3)}$	...	$V_{T-2}^{(3)}$	$V_{T-1}^{(3)}$	$V_T^{(3)}$
2	q_2	$V_1^{(2)}$	$V_2^{(2)}$	$V_3^{(2)}$	...	$V_{T-2}^{(2)}$	$V_{T-1}^{(2)}$	$V_T^{(2)}$
1	q_1	$V_1^{(1)}$	$V_2^{(1)}$	$V_3^{(1)}$	...	$V_{T-2}^{(1)}$	$V_{T-1}^{(1)}$	$V_T^{(1)}$
	q_0	o_1	o_2	o_3	...	o_{T-2}	o_{T-1}	o_T
		1	2	3		T-2	T-1	T

Observations:
$O = o_1, o_2, o_3, \ldots, o_{T-2}, o_{T-1}, o_T$

Columns / Time steps: 1, 2, 3, ..., T-2, T-1, T

---> Time

POS Tagging: Simple Tagset

Let's assume we have a simple tagset:

- **N - NOUN**
- **V - VERB**
- **ART - ARTICLE**
- **P - PREPOSITION**

and some synthetic corpus.

Example sentence:

Flies like a flower

Viterbi Matrix (4 x 4): Structure

Rows: 1, 2, 3, 4

States: $Q = q_1, q_2, q_3, q_4$

4	q_4	$V_1^{(4)}$	$V_2^{(4)}$	$V_3^{(4)}$	$V_4^{(4)}$
3	q_3	$V_1^{(3)}$	$V_2^{(3)}$	$V_3^{(3)}$	$V_4^{(3)}$
2	q_2	$V_1^{(2)}$	$V_2^{(2)}$	$V_3^{(2)}$	$V_4^{(2)}$
1	q_1	$V_1^{(1)}$	$V_2^{(1)}$	$V_3^{(1)}$	$V_4^{(1)}$
	q_0	o_1	o_2	o_3	o_4
		Observations: $O = o_1, o_2, o_3, o_4$			
		1	2	3	4
		Columns / Time steps: 1, 2, 3, 4			

---------------------------------> Time

Viterbi Matrix (4 x 4): Structure

Rows: 1, 2, 3, 4		States: $Q = q_1, q_2, q_3, q_4$					
	4		PREPOSITION	$V_1^{(4)}$	$V_2^{(4)}$	$V_3^{(4)}$	$V_4^{(4)}$
	3		VERB	$V_1^{(3)}$	$V_2^{(3)}$	$V_3^{(3)}$	$V_4^{(3)}$
	2		NOUN	$V_1^{(2)}$	$V_2^{(2)}$	$V_3^{(2)}$	$V_4^{(2)}$
	1		ARTICLE	$V_1^{(1)}$	$V_2^{(1)}$	$V_3^{(1)}$	$V_4^{(1)}$
			<s>	flies	like	a	flower
				Observations: $O = o_1, o_2, o_3, o_4$			
				1	2	3	4
				Columns / Time steps: 1, 2, 3, 4			

---------------------------------> Time

Viterbi Algorithm: Example

q_4	$V_1^{(4)}$	$V_2^{(4)}$	$V_3^{(4)}$	$V_4^{(4)}$
q_3	$V_1^{(3)}$	$V_2^{(3)}$	$V_3^{(3)}$	$V_4^{(3)}$
q_2	$V_1^{(2)}$	$V_2^{(2)}$	$V_3^{(2)}$	$V_4^{(2)}$
q_1	$V_1^{(1)}$	$V_2^{(1)}$	$V_3^{(1)}$	$V_4^{(1)}$
q_0	o_1	o_2	o_3	o_4

Transition probability matrix

	q_0	q_1	q_2	q_2	q_4
q_0	$a_{0,0}$	$a_{0,1}$	$a_{0,2}$	$a_{0,3}$	$a_{0,4}$
q_1	$a_{1,0}$	$a_{1,1}$	$a_{1,2}$	$a_{1,3}$	$a_{1,4}$
q_2	$a_{2,0}$	$a_{2,1}$	$a_{2,2}$	$a_{2,3}$	$a_{2,4}$
q_3	$a_{3,0}$	$a_{3,1}$	$a_{3,2}$	$a_{3,3}$	$a_{3,4}$
q_4	$a_{4.,0}$	$a_{4,1}$	$a_{4,2}$	$a_{4,3}$	$a_{4,4}$

Hidden Markov Model

Emission probability matrix

	o_1	o_2	o_3	o_4
q_0	$b_0(o_1)$	$b_0(o_2)$	$b_0(o_3)$	$b_0(o_4)$
q_1	$b_1(o_1)$	$b_1(o_2)$	$b_1(o_3)$	$b_1(o_4)$
q_2	$b_2(o_1)$	$b_2(o_2)$	$b_2(o_3)$	$b_2(o_4)$
q_3	$b_3(o_1)$	$b_3(o_3)$	$b_3(o_3)$	$b_3(o_4)$
q_4	$b_4(o_1)$	$b_4(o_3)$	$b_4(o_3)$	$b_4(o_4)$

Viterbi Algorithm: Example

PREPOSITION	$V_1^{(4)}$	$V_2^{(4)}$	$V_3^{(4)}$	$V_4^{(4)}$
VERB	$V_1^{(3)}$	$V_2^{(3)}$	$V_3^{(3)}$	$V_4^{(3)}$
NOUN	$V_1^{(2)}$	$V_2^{(2)}$	$V_3^{(2)}$	$V_4^{(2)}$
ARTICLE	$V_1^{(1)}$	$V_2^{(1)}$	$V_3^{(1)}$	$V_4^{(1)}$
<s>	flies	like	a	flower

Transition probability matrix

	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Pseudocode

function VITERBI(*observations* of len T,*state-graph* of len N) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state s **from** 1 **to** N **do** ; initialization step
  $viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
  $backpointer[s,1] \leftarrow 0$

First column
t = 1

for each time step t **from** 2 **to** T **do** ; recursion step
  for each state s **from** 1 **to** N **do**
    $viterbi[s,t] \leftarrow \max_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
    $backpointer[s,t] \leftarrow \operatorname{argmax}_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
$bestpathprob \leftarrow \max_{s=1}^{N} viterbi[s,T]$; termination step
$bestpathpointer \leftarrow \operatorname{argmax}_{s=1}^{N} viterbi[s,T]$; termination step
bestpath ← the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Algorithm: Example

PREPOSITION	$V_1^{(4)}$	$V_2^{(4)}$	$V_3^{(4)}$	$V_4^{(4)}$
VERB	$V_1^{(3)}$	$V_2^{(3)}$	$V_3^{(3)}$	$V_4^{(3)}$
NOUN	$V_1^{(2)}$	$V_2^{(2)}$	$V_3^{(2)}$	$V_4^{(2)}$
ARTICLE	$V_1^{(1)}$	$V_2^{(1)}$	$V_3^{(1)}$	$V_4^{(1)}$
<s>	flies	like	a	flower

$$V_1^{(s)} = \text{viterbi}[s, \text{observation}_1] = \text{viterbi}[s, o_1] = \pi_s * b_s(o_1) = a_{0,s} * b_s(o_1) = P(s \mid q_0) * P(o_1 \mid s)$$

$$V_1^{(1)} = \text{viterbi}[\text{state}_1, \text{observation}_1] = \text{viterbi}[q_1, o_1] = \pi_1 * b_1(o_1) = a_{0,1} * b_1(o_1) = P(q_1 \mid q_0) * P(o_1 \mid q_1)$$

$$V_1^{(2)} = \text{viterbi}[\text{state}_2, \text{observation}_1] = \text{viterbi}[q_2, o_1] = \pi_2 * b_2(o_1) = a_{0,2} * b_2(o_1) = P(q_2 \mid q_0) * P(o_1 \mid q_2)$$

$$V_1^{(3)} = \text{viterbi}[\text{state}_3, \text{observation}_1] = \text{viterbi}[q_3, o_1] = \pi_3 * b_3(o_1) = a_{0,3} * b_3(o_1) = P(q_3 \mid q_0) * P(o_1 \mid q_3)$$

$$V_1^{(4)} = \text{viterbi}[\text{state}_4, \text{observation}_1] = \text{viterbi}[q_4, o_1] = \pi_4 * b_4(o_1) = a_{0,4} * b_4(o_1) = P(q_4 \mid q_0) * P(o_1 \mid q_4)$$

Transition probability matrix

	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Pseudocode

function VITERBI(*observations* of len T,*state-graph* of len N) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state s **from** 1 **to** N **do** ; initialization step
$\quad viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
$\quad backpointer[s,1] \leftarrow 0$

First column
t = 1

for each time step t **from** 2 **to** T **do** ; recursion step
$\quad$ **for** each state s **from** 1 **to** N **do**

$$viterbi[s,t] \leftarrow \max_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$$

$$backpointer[s,t] \leftarrow \operatorname*{argmax}_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$$

$$bestpathprob \leftarrow \max_{s=1}^{N} viterbi[s,T]$$; termination step

$$bestpathpointer \leftarrow \operatorname*{argmax}_{s=1}^{N} viterbi[s,T]$$; termination step

bestpath $\leftarrow$ the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Algorithm: Example

PREPOSITION	$V_1^{(4)}$	$V_2^{(4)}$	$V_3^{(4)}$	$V_4^{(4)}$
VERB	$V_1^{(3)}$	$V_2^{(3)}$	$V_3^{(3)}$	$V_4^{(3)}$
NOUN	$V_1^{(2)}$	$V_2^{(2)}$	$V_3^{(2)}$	$V_4^{(2)}$
ARTICLE	$V_1^{(1)}$	$V_2^{(1)}$	$V_3^{(1)}$	$V_4^{(1)}$
<s>	flies	like	a	flower

$$V_1^{(s)} = \text{viterbi[state } s, \text{observation}_1] = \text{viterbi}[s, o_1] = \pi_s * b_s(o_1) = a_{0,s} * b_s(o_1) = P(s \mid q_0) * P(o_1 \mid s)$$

Transition probability matrix

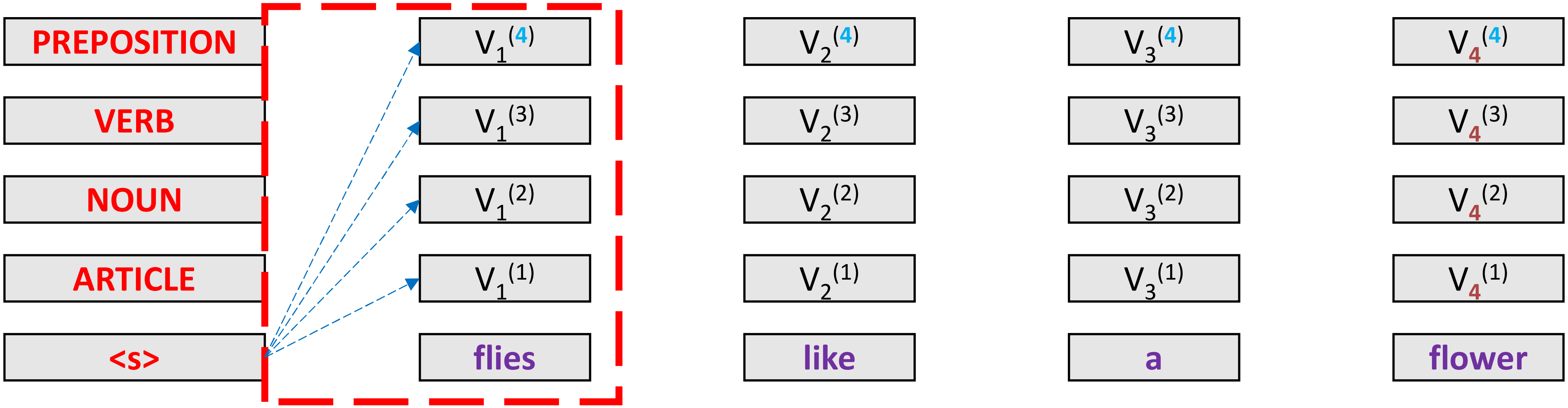
	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Example



$V_1^{(s)}$ = viterbi[state s, observation$_1$] = viterbi[s, o_1] = $\pi_s * b_s(o_1) = a_{0,s} * b_s(o_1) = P(s \mid q_0) * P(o_1 \mid s)$

$V_1^{(1)}$ = viterbi[state$_1$, observation$_1$] = viterbi[q_1, o_1] = $\pi_1 * b_1(o_1) = a_{0,1} * b_1(o_1) = P(q_1 \mid q_0) * P(o_1 \mid q_1)$

$V_1^{(2)}$ = viterbi[state$_2$, observation$_1$] = viterbi[q_2, o_1] = $\pi_2 * b_2(o_1) = a_{0,2} * b_2(o_1) = P(q_2 \mid q_0) * P(o_1 \mid q_2)$

$V_1^{(3)}$ = viterbi[state$_3$, observation$_1$] = viterbi[q_3, o_1] = $\pi_3 * b_3(o_1) = a_{0,3} * b_3(o_1) = P(q_3 \mid q_0) * P(o_1 \mid q_3)$

$V_1^{(4)}$ = viterbi[state$_4$, observation$_1$] = viterbi[q_4, o_1] = $\pi_4 * b_4(o_1) = a_{0,4} * b_4(o_1) = P(q_4 \mid q_0) * P(o_1 \mid q_4)$

Transition probability matrix

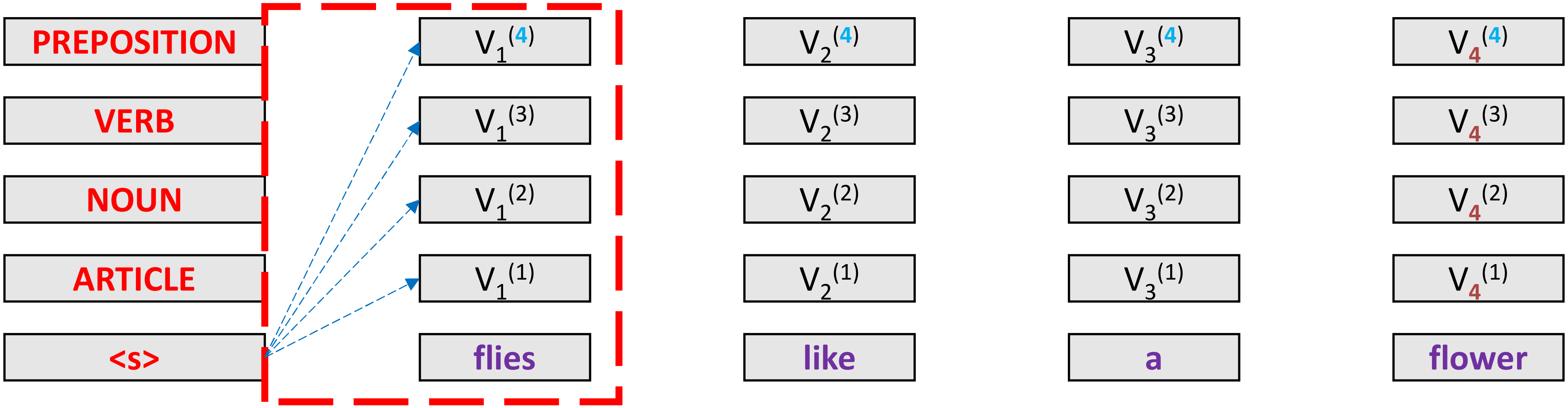
	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Example



$V_1^{(s)}$ = viterbi[state s, observation$_1$] = viterbi[s, o_1] = π_s * $b_s(o_1)$ = $a_{0,s}$ * $b_s(o_1)$ = P(s | q_0) * P(o_1 | s)

$V_1^{(1)}$ = viterbi[**ARTICLE**, **flies**] = $a_{0,1}$ * b_1(**flies**) = P(**ARTICLE** | **<s>**) * P(**flies** | **ARTICLE**)

$V_1^{(2)}$ = viterbi[**NOUN**, **flies**] = $a_{0,2}$ * b_2(**flies**) = P(**NOUN** | **<s>**) * P(**flies** | **NOUN**)

$V_1^{(3)}$ = viterbi[**VERB**, **flies**] = $a_{0,3}$ * b_3(**flies**) = P(**VERB** | **<s>**) * P(**flies** | **VERB**)

$V_1^{(4)}$ = viterbi[**PREPOSITION**, **flies**] = $a_{0,4}$ * b_4(**flies**) = P(**PREPOSITION** | **<s>**) * P(**flies** | **PREPOSITION**)

Transition probability matrix

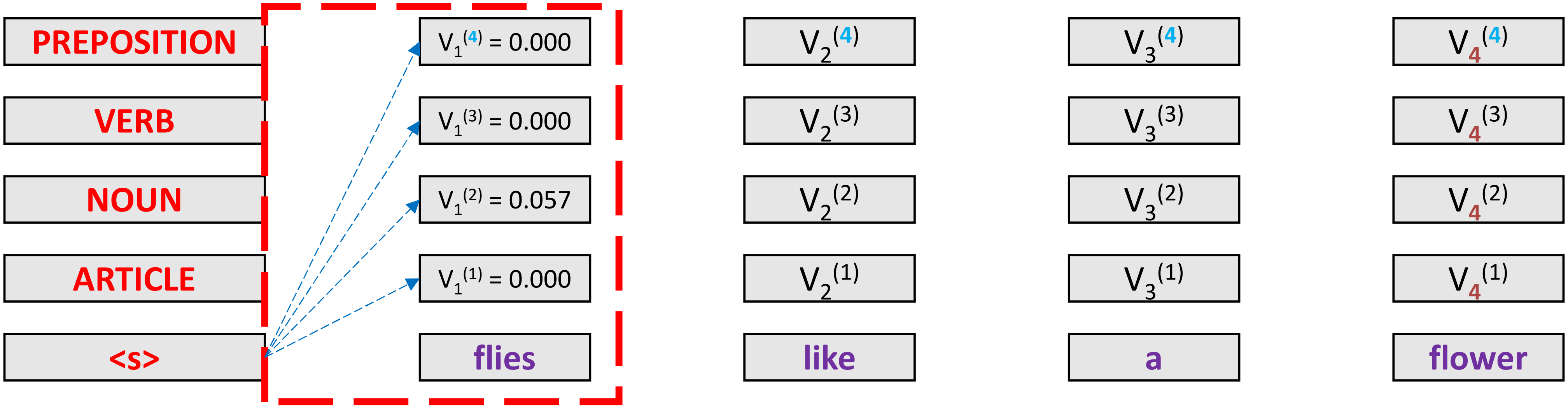
	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Example



$V_1^{(s)}$ = viterbi[state s, observation$_1$] = viterbi[s, o_1] = $\pi_s * b_s(o_1) = a_{0,s} * b_s(o_1) = P(s \mid q_0) * P(o_1 \mid s)$

$V_1^{(1)}$ = viterbi[ARTICLE, flies] = P(ARTICLE | <s>) * P(flies | ARTICLE) = 0.71 * 0.000 = 0.000

$V_1^{(2)}$ = viterbi[NOUN, flies] = P(NOUN | <s>) * P(flies | NOUN) = 0.29 * 0.025 = 0.057

$V_1^{(3)}$ = viterbi[VERB, flies] = P(VERB | <s>) * P(flies | VERB) = 0.00 * 0.076 = 0.000

$V_1^{(4)}$ = viterbi[PREPOSITION, flies] = P(PREPOSITION | <s>) * P(flies | PREPOSITION) = 0.00 * 0.000 = 0.000

Transition probability matrix

	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Pseudocode

function VITERBI(*observations* of len T,*state-graph* of len N) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*

for each state s **from** 1 **to** N **do** ; initialization step

$\quad viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$

$\quad \boxed{backpointer[s,1] \leftarrow 0}$

First column

t = 1

for each time step t **from** 2 **to** T **do** ; recursion step

$\quad$ **for** each state s **from** 1 **to** N **do**

$$viterbi[s,t] \leftarrow \max_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$$

$$backpointer[s,t] \leftarrow \operatorname{argmax}_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$$

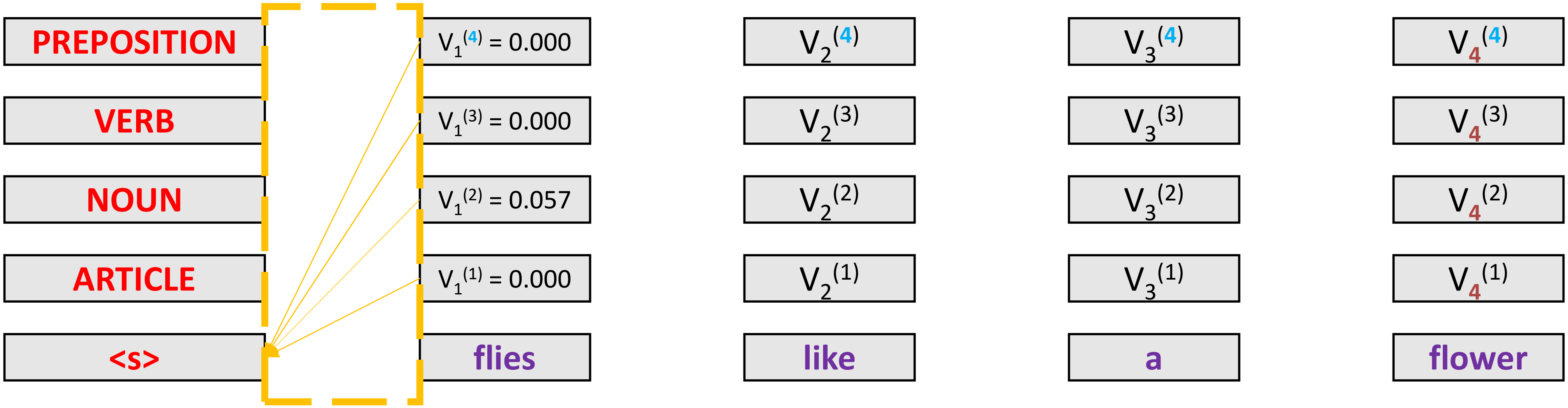
$bestpathprob \leftarrow \max_{s=1}^{N} viterbi[s,T]$; termination step

$bestpathpointer \leftarrow \operatorname{argmax}_{s=1}^{N} viterbi[s,T]$; termination step

bestpath ← the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time

return *bestpath*, *bestpathprob*

Viterbi Algorithm: Example



$$V_1^{(s)} = \text{viterbi}[s, \text{observation}_1] = \text{viterbi}[s, o_1] = \pi_s * b_s(o_1) = a_{0,s} * b_s(o_1) = P(s \mid q_0) * P(o_1 \mid s)$$

$V_1^{(1)}$ = viterbi[**ARTICLE**, **flies**] = P(**ARTICLE** | **<s>**) * P(**flies** | **ARTICLE**) = 0.71 * 0.000 = 0.000

$V_1^{(2)}$ = viterbi[**NOUN**, **flies**] = P(**NOUN** | **<s>**) * P(**flies** | **NOUN**) = 0.29 * 0.025 = 0.057

$V_1^{(3)}$ = viterbi[**VERB**, **flies**] = P(**VERB** | **<s>**) * P(**flies** | **VERB**) = 0.00 * 0.076 = 0.000

$V_1^{(4)}$ = viterbi[**PREPOSITION**, **flies**] = P(**PREPOSITION** | **<s>**) * P(**flies** | **PREPOSITION**) = 0.00 * 0.000 = 0.000

Transition probability matrix

	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Pseudocode

function VITERBI(*observations* of len T,*state-graph* of len N) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state s **from** 1 **to** N **do** ; initialization step
    $viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
    $backpointer[s,1] \leftarrow 0$
for each time step t **from** 2 **to** T **do** ; recursion step
  for each state s **from** 1 **to** N **do**
    $$viterbi[s,t] \leftarrow \max_{s'=1}^{N} \; viterbi[s',t-1] * a_{s',s} * b_s(o_t)$$
    $$backpointer[s,t] \leftarrow \operatorname{argmax}_{s'=1}^{N} \; viterbi[s',t-1] * a_{s',s} * b_s(o_t)$$

t = 2
Second column

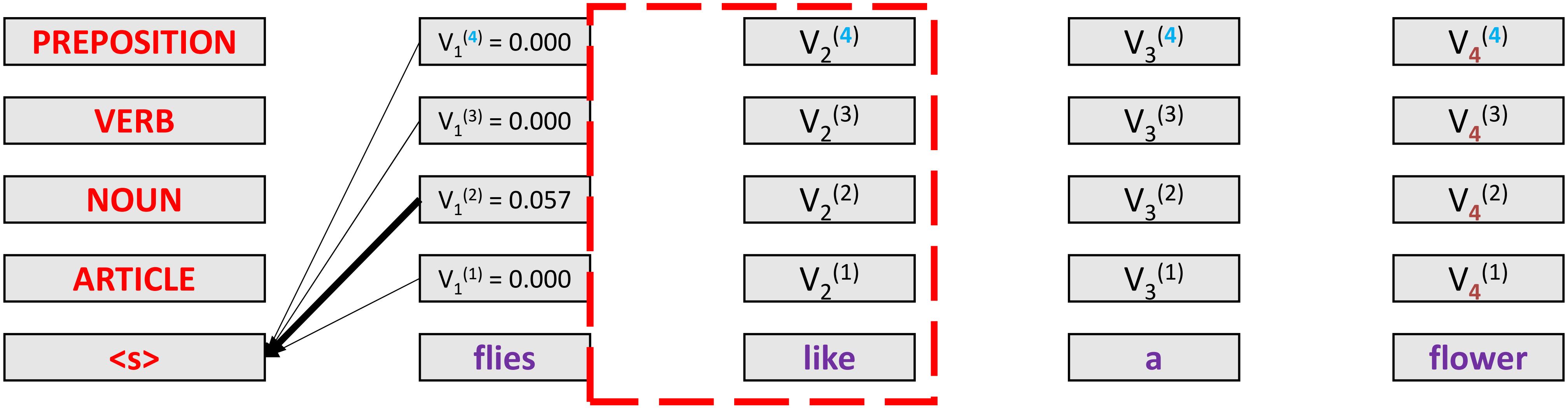
$$bestpathprob \leftarrow \max_{s=1}^{N} \; viterbi[s,T]$$; termination step
$$bestpathpointer \leftarrow \operatorname{argmax}_{s=1}^{N} \; viterbi[s,T]$$; termination step
bestpath $\leftarrow$ the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Algorithm: Pseudocode

function VITERBI(*observations* of len *T*,*state-graph* of len *N*) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state *s* **from** 1 **to** *N* **do** ; initialization step
  $viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
  $backpointer[s,1] \leftarrow 0$
for each time step *t* **from** 2 **to** *T* **do** ; recursion step
  for each state *s* **from** 1 **to** *N* **do**
   $viterbi[s,t] \leftarrow \max_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
   $backpointer[s,t] \leftarrow \operatorname{argmax}_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
$bestpathprob \leftarrow \max_{s=1}^{N} viterbi[s,T]$; termination step
$bestpathpointer \leftarrow \operatorname{argmax}_{s=1}^{N} viterbi[s,T]$; termination step
bestpath ← the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Algorithm: Example



Transition probability matrix

	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

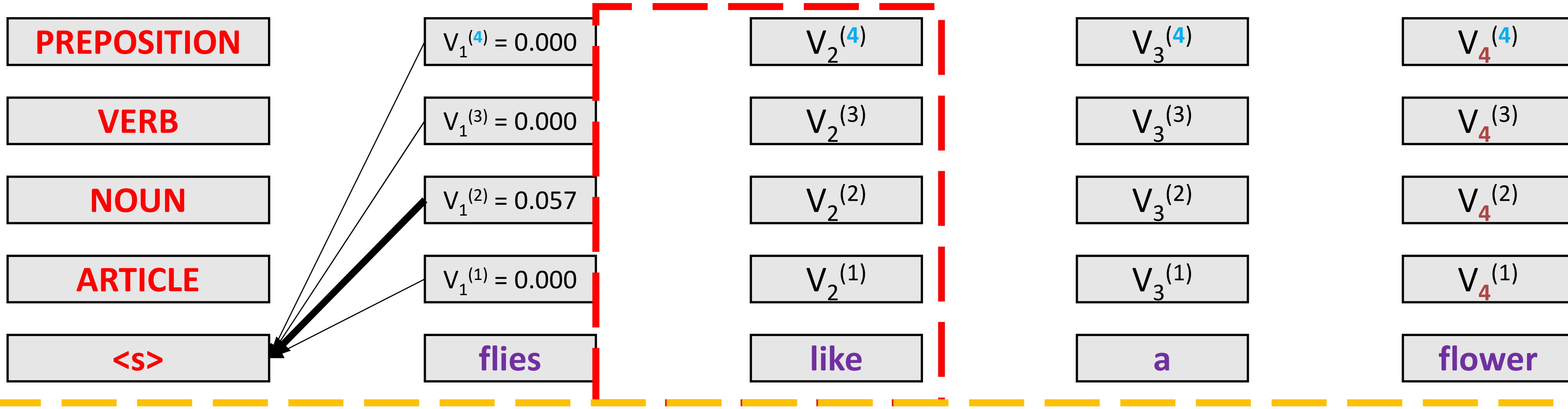
	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Pseudocode

function VITERBI(*observations* of len T,*state-graph* of len N) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state s **from** 1 **to** N **do** ; initialization step
    $viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
    $backpointer[s,1] \leftarrow 0$
for each time step t **from** 2 **to** T **do** ; recursion step
    for each state s **from** 1 **to** N **do**
        $viterbi[s,t] \leftarrow \max_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
        $backpointer[s,t] \leftarrow \operatorname{argmax}_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
$bestpathprob \leftarrow \max_{s=1}^{N} viterbi[s,T]$; termination step
$bestpathpointer \leftarrow \operatorname{argmax}_{s=1}^{N} viterbi[s,T]$; termination step
bestpath ← the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Algorithm: Example



$V_t^{(s)}$ = viterbi[state s, observation$_t$] = viterbi[s, o_t] = $\max_{s'}$(viterbi[state s', observation$_t$] * $a_{s',s}$ * $b_s(o_t)$)

Transition probability matrix

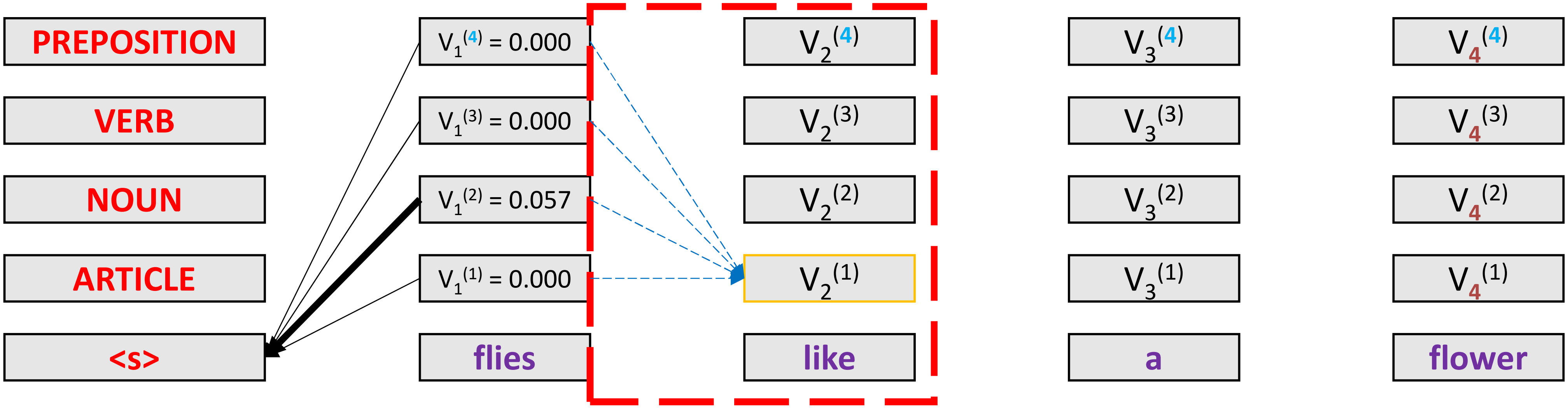
	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Example



$V_2^{(1)}$ = viterbi[ARTICLE, like] = $\max_{s'}$(viterbi[state s', like] * $a_{s',1}$ * b_1(like)) = max(

$V_1^{(1)}$ * P(ARTICLE | ARTICLE) * P(like | ARTICLE) = 0.000 * 0.00 * 0.000 = 0.000
$V_1^{(2)}$ * P(ARTICLE | NOUN) * P(like | ARTICLE) = 0.057 * 0.00 * 0.000 = 0.000
$V_1^{(3)}$ * P(ARTICLE | VERB) * P(like | ARTICLE) = 0.000 * 0.65 * 0.000 = 0.000
$V_1^{(4)}$ * P(ARTICLE | PREPOSITION) * P(like | ARTICLE) = 0.000 * 0.74 * 0.000 = 0.000
)

Transition probability matrix

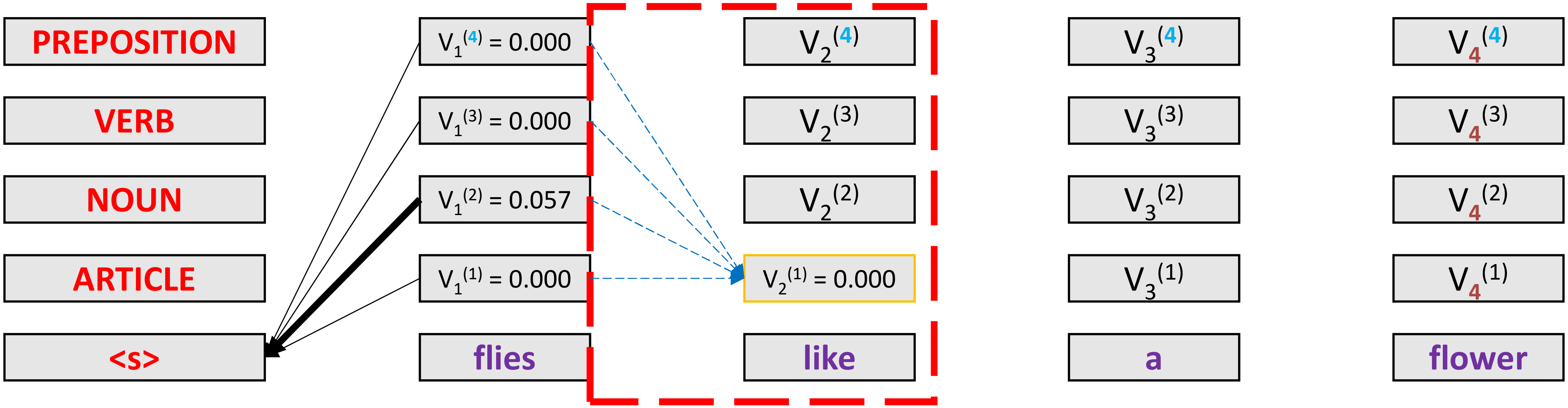
	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Example



$V_2^{(1)}$ = viterbi[ARTICLE, like] = $\max_{s'}$(viterbi[state s', like] * $a_{s',1}$ * b_1(like)) = 0.000

Transition probability matrix

	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

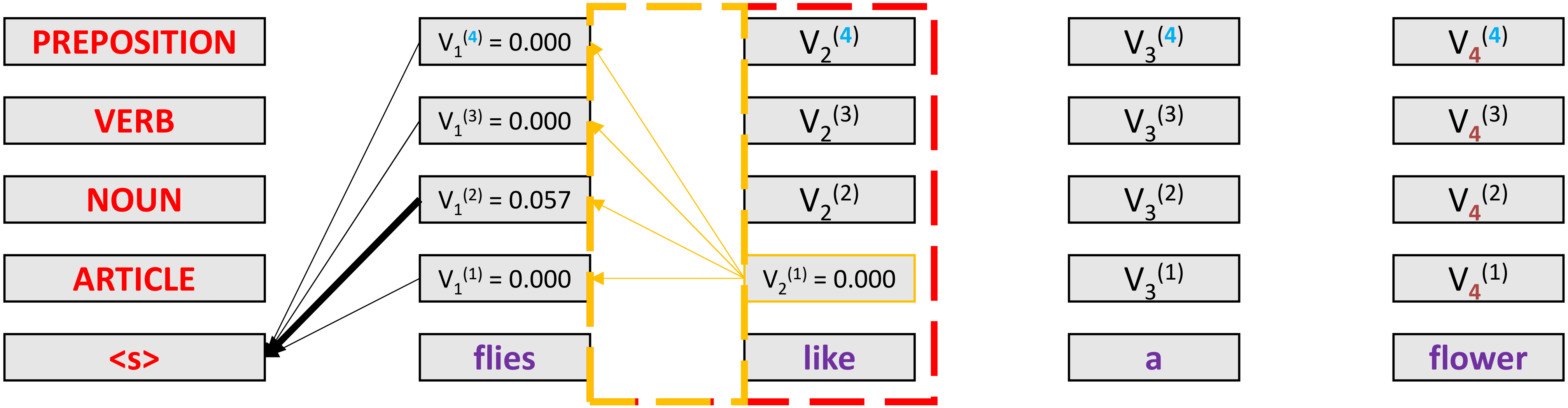
	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Pseudocode

function VITERBI(*observations* of len T,*state-graph* of len N) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state s **from** 1 **to** N **do** ; initialization step
    $viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
    $backpointer[s,1] \leftarrow 0$
for each time step t **from** 2 **to** T **do** ; recursion step
    for each state s **from** 1 **to** N **do**
        $viterbi[s,t] \leftarrow \max_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
        $backpointer[s,t] \leftarrow \operatorname{argmax}_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
$bestpathprob \leftarrow \max_{s=1}^{N} viterbi[s,T]$; termination step
$bestpathpointer \leftarrow \operatorname{argmax}_{s=1}^{N} viterbi[s,T]$; termination step
bestpath ← the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Algorithm: Example



$V_2^{(1)}$ = viterbi[ARTICLE, like] = $\max_{s'}$(viterbi[state s', like] * $a_{s',1}$ * b_1(like)) = 0.000

Transition probability matrix

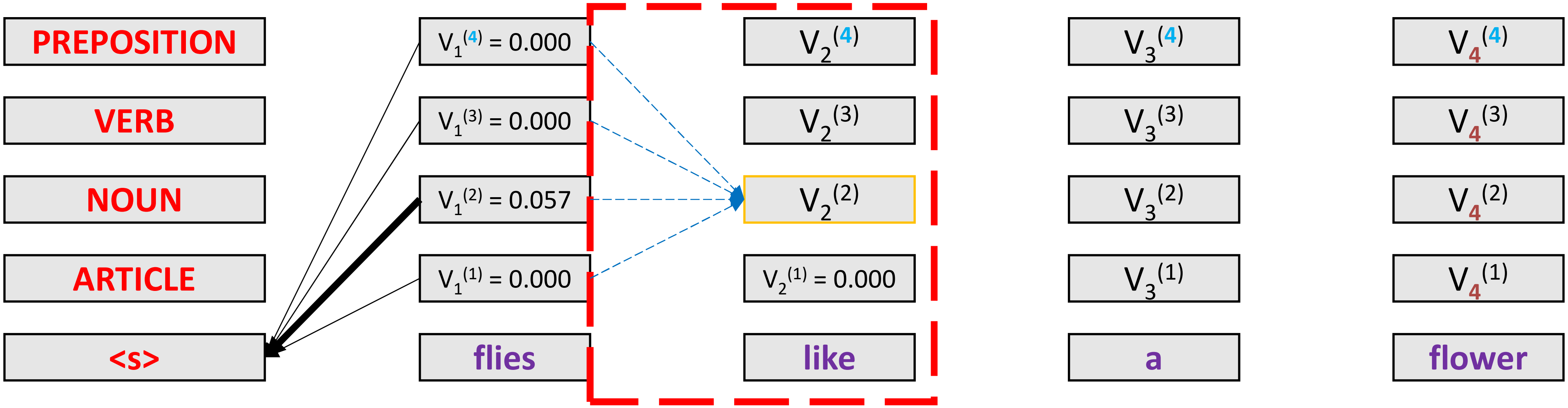
	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Example



$V_2^{(2)}$ = viterbi[NOUN, like] = $\max_{s'}$(viterbi[state s', like] * $a_{s',2}$ * b_2(like)) = max(

$V_1^{(1)}$ * P(NOUN | ARTICLE) * P(like | NOUN) = 0.000 * 1.00 * 0.012 = 0.000

$V_1^{(2)}$ * P(NOUN | NOUN) * P(like | NOUN) = 0.057 * 0.13 * 0.012 = <u>8.892E-5</u>

$V_1^{(3)}$ * P(NOUN | VERB) * P(like | NOUN) = 0.000 * 0.35 * 0.012 = 0.000

$V_1^{(4)}$ * P(NOUN | PREPOSITION) * P(like | NOUN) = 0.000 * 0.26 * 0.012 = 0.000

)

Transition probability matrix

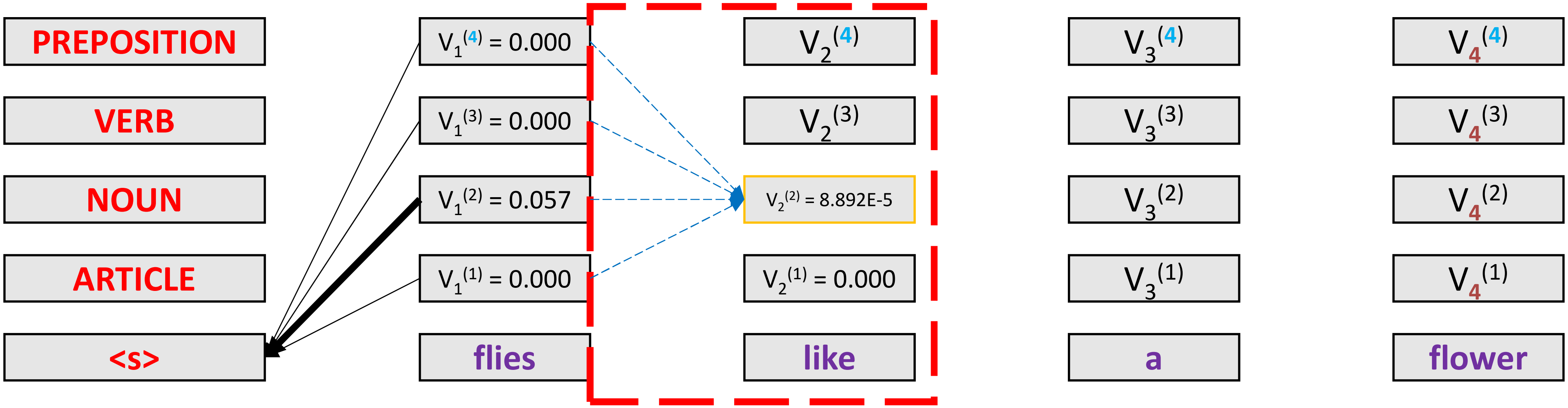
	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Example



$V_2^{(2)}$ = viterbi[NOUN, like] = $\max_{s'}$(viterbi[state s', like] * $a_{s',2}$ * b_2(like)) = 8.892E-5

Transition probability matrix

	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

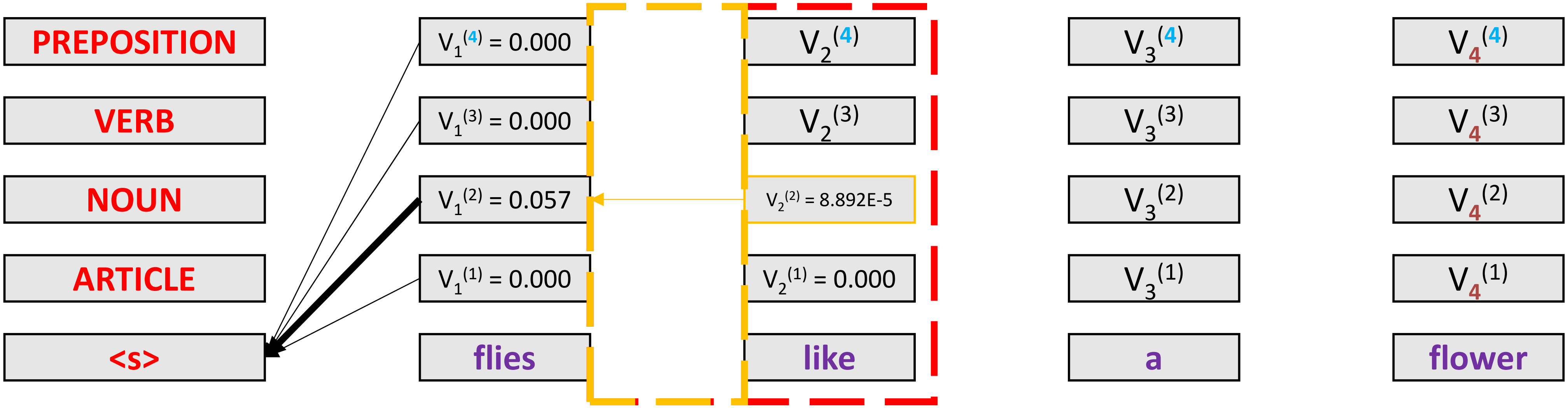
	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Pseudocode

function VITERBI(*observations* of len T,*state-graph* of len N) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state s **from** 1 **to** N **do** ; initialization step
$viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
$backpointer[s,1] \leftarrow 0$
for each time step t **from** 2 **to** T **do** ; recursion step
for each state s **from** 1 **to** N **do**
$viterbi[s,t] \leftarrow \max_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
$backpointer[s,t] \leftarrow \operatorname{argmax}_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
$bestpathprob \leftarrow \max_{s=1}^{N} viterbi[s,T]$; termination step
$bestpathpointer \leftarrow \operatorname{argmax}_{s=1}^{N} viterbi[s,T]$; termination step
bestpath $\leftarrow$ the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Algorithm: Example



$V_2^{(2)}$ = viterbi[NOUN, like] = $\max_{s'}$(viterbi[state s', like] * $a_{s',2}$ * b_2(like)) = 8.892E-5

Transition probability matrix

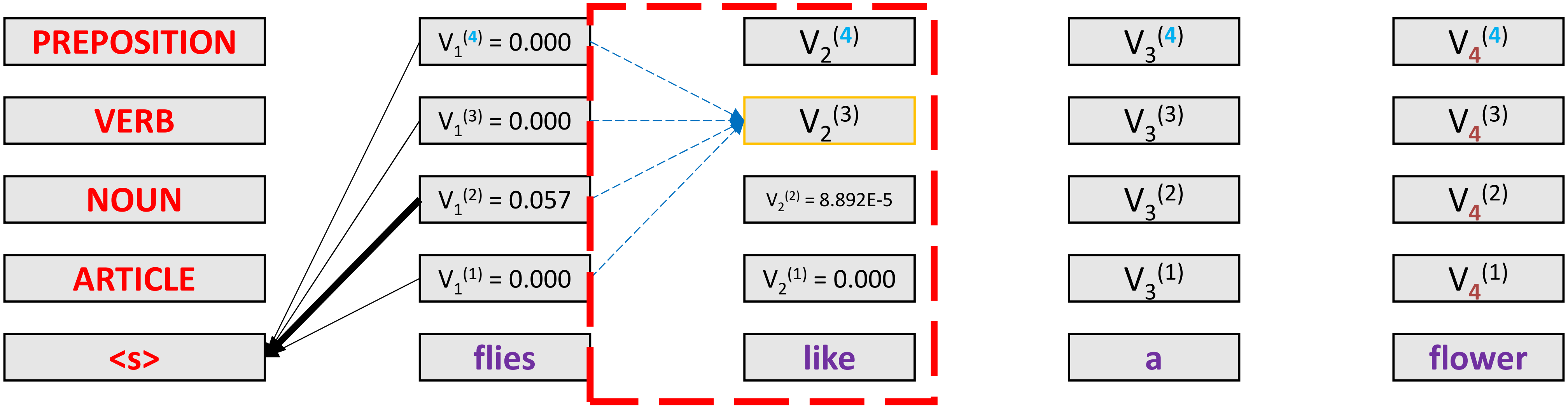
	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Example



$V_2^{(3)}$ = viterbi[**VERB**, **like**] = $\max_{s'}$(viterbi[state s', **like**] * $a_{s',3}$ * b_3(**like**)) = max(

$V_1^{(1)}$ * P(**VERB** | **ARTICLE**) * P(**like** | **VERB**) = 0.000 * 0.00 * 0.100 = 0.000

$V_1^{(2)}$ * P(**VERB** | **NOUN**) * P(**like** | **VERB**) = 0.057 * 0.43 * 0.100 = <u>0.002451</u>

$V_1^{(3)}$ * P(**VERB** | **VERB**) * P(**like** | **VERB**) = 0.000 * 0.00 * 0.100 = 0.000

$V_1^{(4)}$ * P(**VERB** | **PREPOSITION**) * P(**like** | **VERB**) = 0.000 * 0.00 * 0.100 = 0.000

)

Transition probability matrix

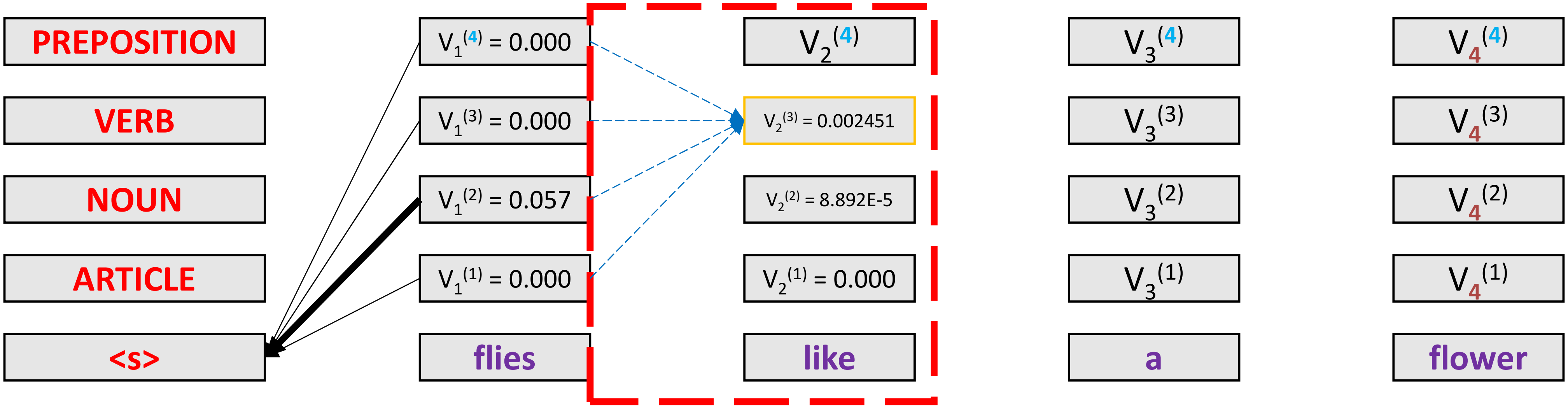
	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Example



$V_2^{(3)}$ = viterbi[VERB, like] = $\max_{s'}$(viterbi[state s', like] * $a_{s',3}$ * b_3(like)) = 0.002451

Transition probability matrix

	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

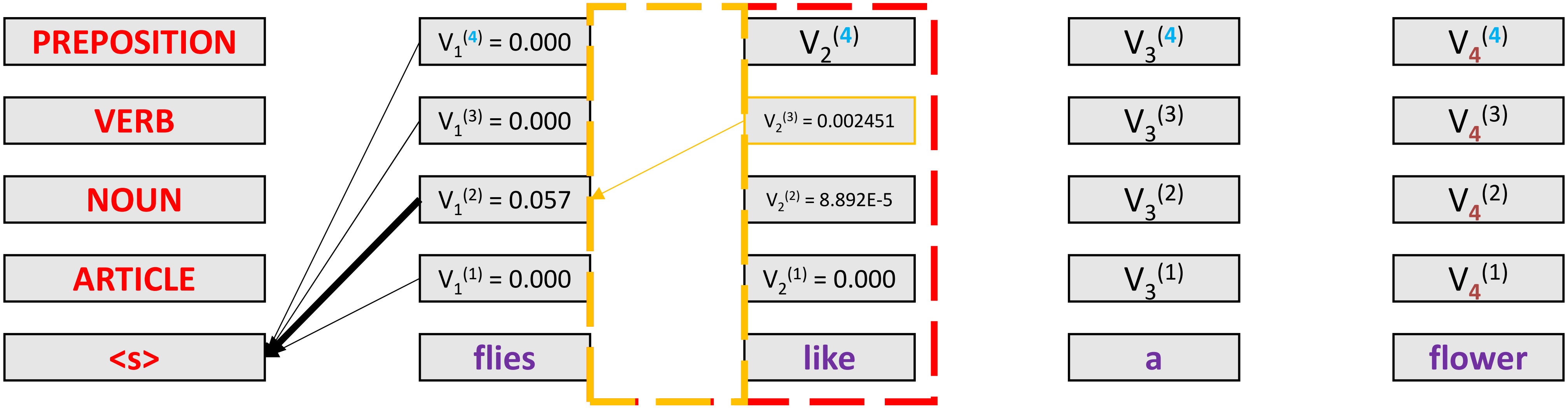
	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Pseudocode

function VITERBI(*observations* of len T,*state-graph* of len N) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state s **from** 1 **to** N **do** ; initialization step
$\quad viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
$\quad backpointer[s,1] \leftarrow 0$
for each time step t **from** 2 **to** T **do** ; recursion step
$\quad$ **for** each state s **from** 1 **to** N **do**
$$viterbi[s,t] \leftarrow \max_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$$
$$backpointer[s,t] \leftarrow \operatorname{argmax}_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$$
$bestpathprob \leftarrow \max_{s=1}^{N} viterbi[s,T]$; termination step
$bestpathpointer \leftarrow \operatorname{argmax}_{s=1}^{N} viterbi[s,T]$; termination step
bestpath $\leftarrow$ the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Algorithm: Example



$V_2^{(3)}$ = viterbi[VERB, like] = $\max_{s'}$(viterbi[state s', like] * $a_{s',3}$ * b_3(like)) = 0.002451

Transition probability matrix

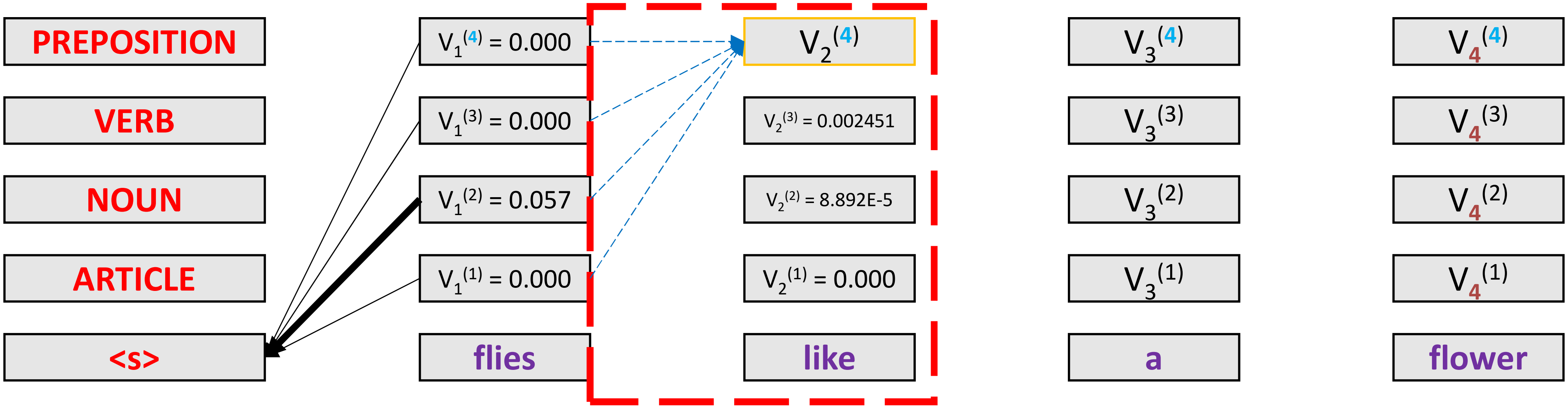
	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Example



$V_2^{(4)}$ = viterbi[PREPOSITION, like] = $\max_{s'}$(viterbi[state s', like] * $a_{s',4}$ * b_4(like)) = max(

$V_1^{(1)}$ * P(PREPOSITION | ARTICLE) * P(like | PREPOSITION) = 0.000 * 0.00 * 0.068 = 0.000
$V_1^{(2)}$ * P(PREPOSITION | NOUN) * P(like | PREPOSITION) = 0.057 * 0.44 * 0.068 = <u>0.00170544</u>
$V_1^{(3)}$ * P(PREPOSITION | VERB) * P(like | PREPOSITION) = 0.000 * 0.00 * 0.068 = 0.000
$V_1^{(4)}$ * P(PREPOSITION | PREPOSITION) * P(like | PREPOSITION) = 0.000 * 0.00 * 0.068 = 0.000
)

Transition probability matrix

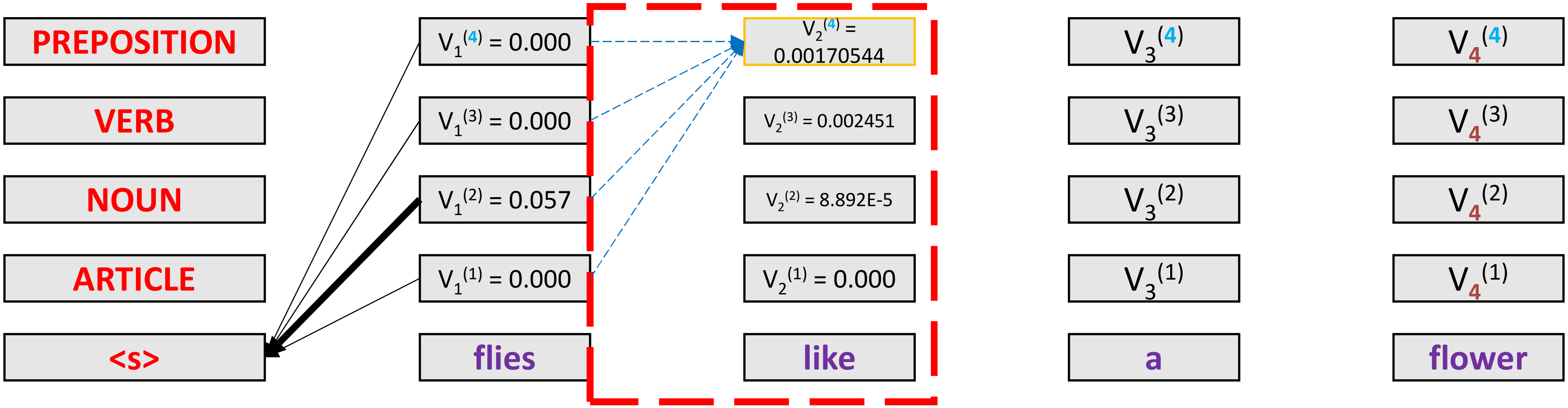
	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Example



$V_2^{(4)}$ = viterbi[PREPOSITION, like] = $\max_{s'}$(viterbi[state s', like] * $a_{s',4}$ * b_4(like)) = 0.00170544

Transition probability matrix

	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

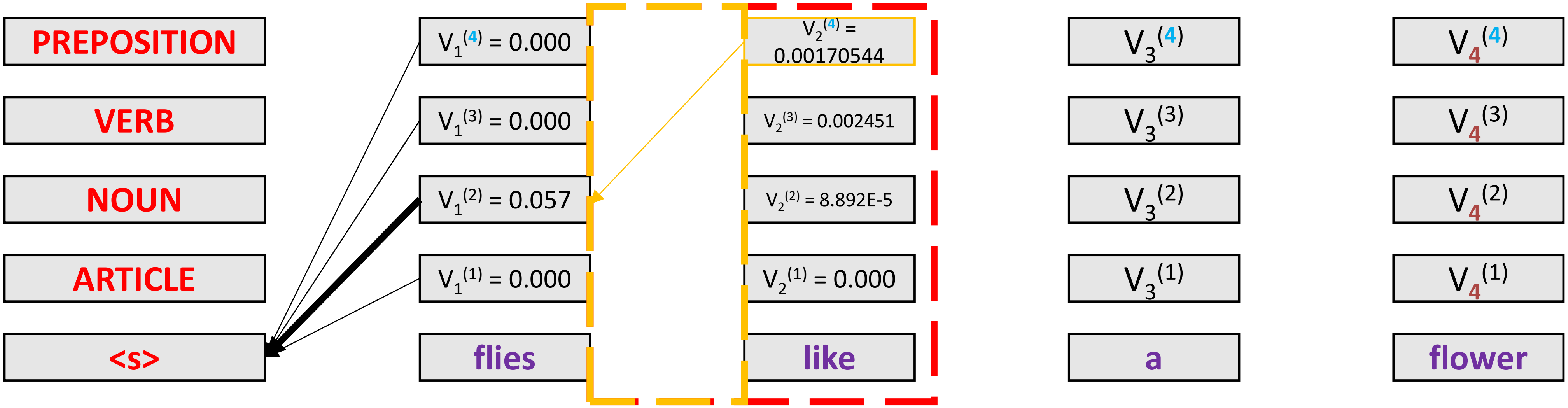
	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Pseudocode

function VITERBI(*observations* of len T,*state-graph* of len N) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state s **from** 1 **to** N **do** ; initialization step
  $viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
  $backpointer[s,1] \leftarrow 0$
for each time step t **from** 2 **to** T **do** ; recursion step
  for each state s **from** 1 **to** N **do**
    $viterbi[s,t] \leftarrow \max_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
    $backpointer[s,t] \leftarrow \operatorname{argmax}_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
$bestpathprob \leftarrow \max_{s=1}^{N} viterbi[s,T]$; termination step
$bestpathpointer \leftarrow \operatorname{argmax}_{s=1}^{N} viterbi[s,T]$; termination step
bestpath $\leftarrow$ the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Algorithm: Example



$V_2^{(4)}$ = viterbi[PREPOSITION, like] = $\max_{s'}$(viterbi[state s', like] * $a_{s',4}$ * b_4(like)) = 0.00170544

Transition probability matrix

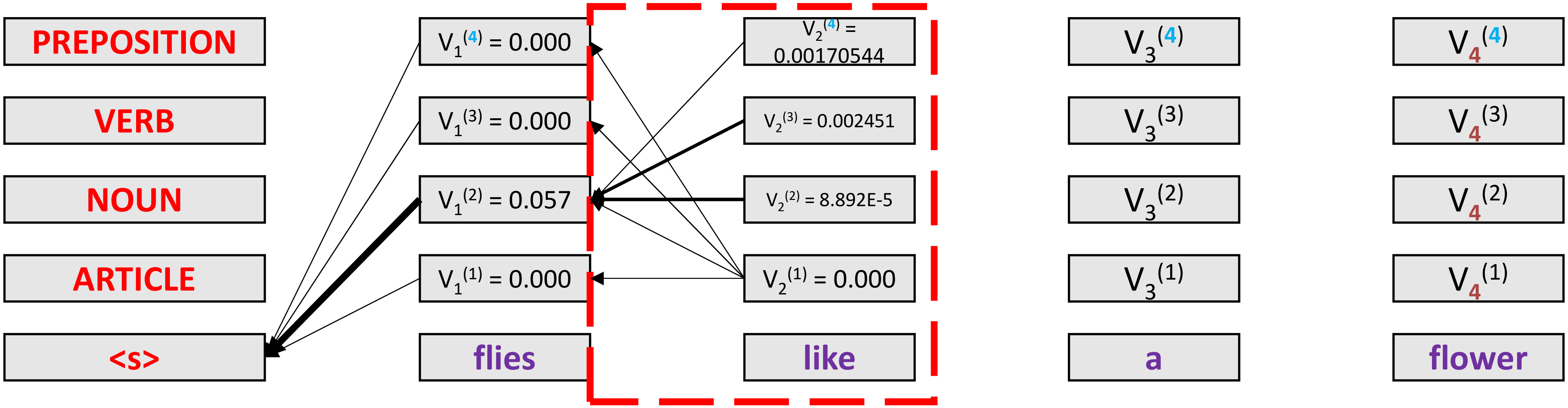
	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Example



Transition probability matrix

	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Pseudocode

function VITERBI(*observations* of len T,*state-graph* of len N) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state s **from** 1 **to** N **do** ; initialization step
$viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
$backpointer[s,1] \leftarrow 0$
for each time step t **from** 2 **to** T **do** ; recursion step
for each state s **from** 1 **to** N **do**

$$viterbi[s,t] \leftarrow \max_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$$

$$backpointer[s,t] \leftarrow \operatorname{argmax}_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$$

t = 3
Third column

$$bestpathprob \leftarrow \max_{s=1}^{N} viterbi[s,T]$$; termination step

$$bestpathpointer \leftarrow \operatorname{argmax}_{s=1}^{N} viterbi[s,T]$$; termination step

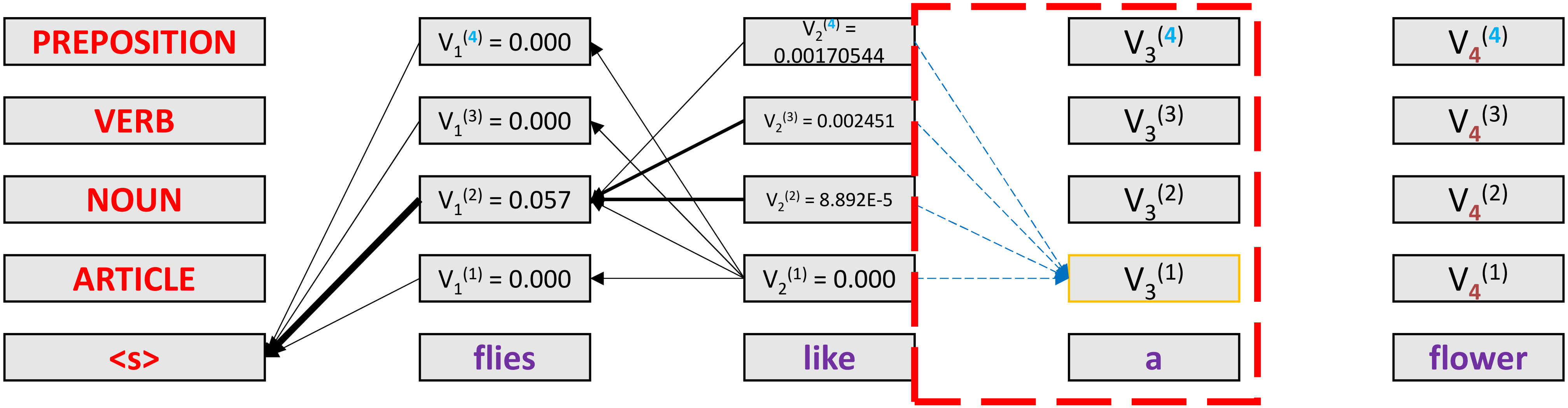
bestpath $\leftarrow$ the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Algorithm: Pseudocode

function VITERBI(*observations* of len *T*,*state-graph* of len *N*) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state *s* **from** 1 **to** *N* **do** ; initialization step
    $viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
    $backpointer[s,1] \leftarrow 0$
for each time step *t* **from** 2 **to** *T* **do** ; recursion step
    for each state *s* **from** 1 **to** *N* **do**
        $viterbi[s,t] \leftarrow \max_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
        $backpointer[s,t] \leftarrow \operatorname{argmax}_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
$bestpathprob \leftarrow \max_{s=1}^{N} viterbi[s,T]$; termination step
$bestpathpointer \leftarrow \operatorname{argmax}_{s=1}^{N} viterbi[s,T]$; termination step
bestpath ← the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Algorithm: Example



$V_3^{(1)}$ = viterbi[**ARTICLE**, a] = $\max_{s'}$(viterbi[state s', a] * $a_{s',1}$ * b_1(a)) = max(

$V_2^{(1)}$ * P(**ARTICLE** | **ARTICLE**) * P(a | **ARTICLE**) = 0.000 * 0.00 * 0.360 = 0.000
$V_2^{(2)}$ * P(**ARTICLE** | **NOUN**) * P(a | **ARTICLE**) = 8.892E-5 * 0.00 * 0.360 = 0.000
$V_2^{(3)}$ * P(**ARTICLE** | **VERB**) * P(a | **ARTICLE**) = 0.002451 * 0.65 * 0.360 = 0.000573
$V_2^{(4)}$ * P(**ARTICLE** | **PREPOSITION**) * P(a | **ARTICLE**) = 0.00170544 * 0.74 * 0.360 = 0.00045432921
)

Transition probability matrix

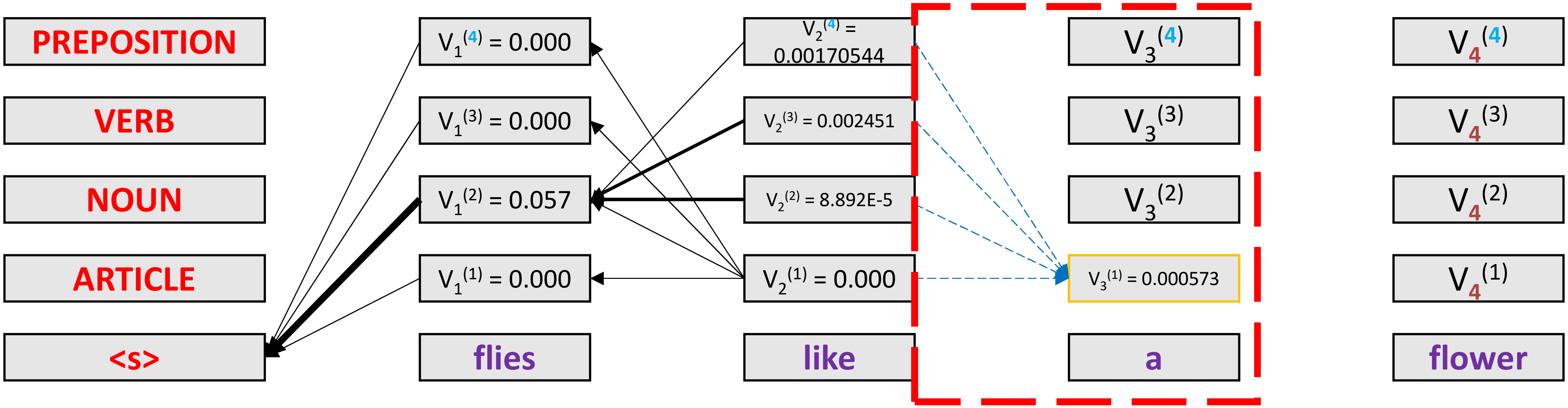
	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Example



$V_3^{(1)}$ = viterbi[ARTICLE, a] = $\max_{s'}$(viterbi[state s', a] * $a_{s',1}$ * b_1(a)) = 0.000573

Transition probability matrix

	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

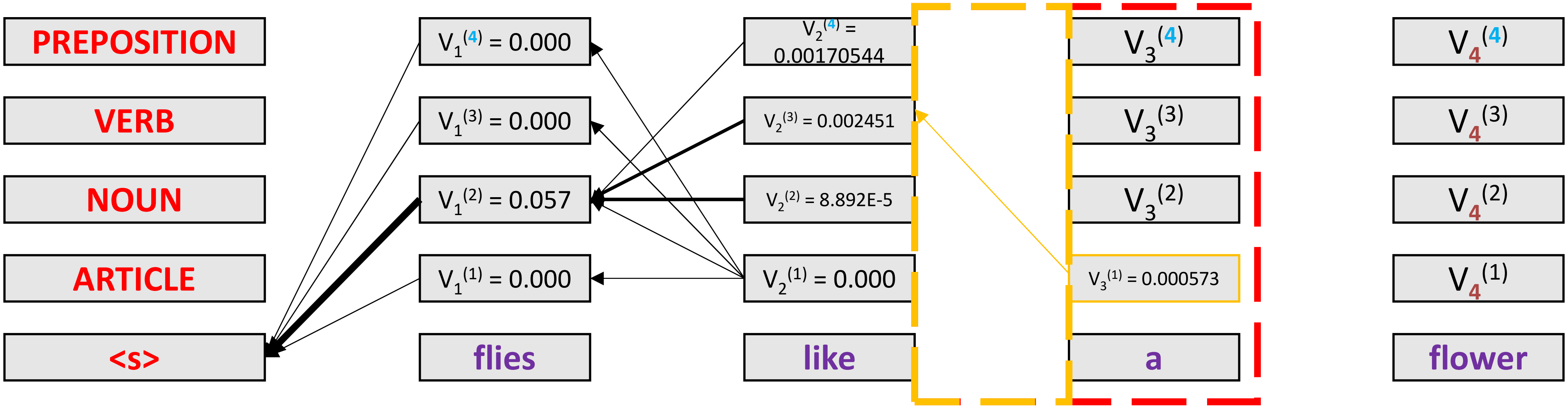
	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Pseudocode

function VITERBI(*observations* of len *T*,*state-graph* of len *N*) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state *s* **from** 1 **to** *N* **do** ; initialization step
    $viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
    $backpointer[s,1] \leftarrow 0$
for each time step *t* **from** 2 **to** *T* **do** ; recursion step
    for each state *s* **from** 1 **to** *N* **do**
        $viterbi[s,t] \leftarrow \max_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
        $backpointer[s,t] \leftarrow \operatorname{argmax}_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
$bestpathprob \leftarrow \max_{s=1}^{N} viterbi[s,T]$; termination step
$bestpathpointer \leftarrow \operatorname{argmax}_{s=1}^{N} viterbi[s,T]$; termination step
bestpath ← the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Algorithm: Example



$V_3^{(1)}$ = viterbi[**ARTICLE**, a] = $\max_{s'}$(viterbi[state s', a] * $a_{s',1}$ * b_1(a)) = 0.000573

Transition probability matrix					
	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

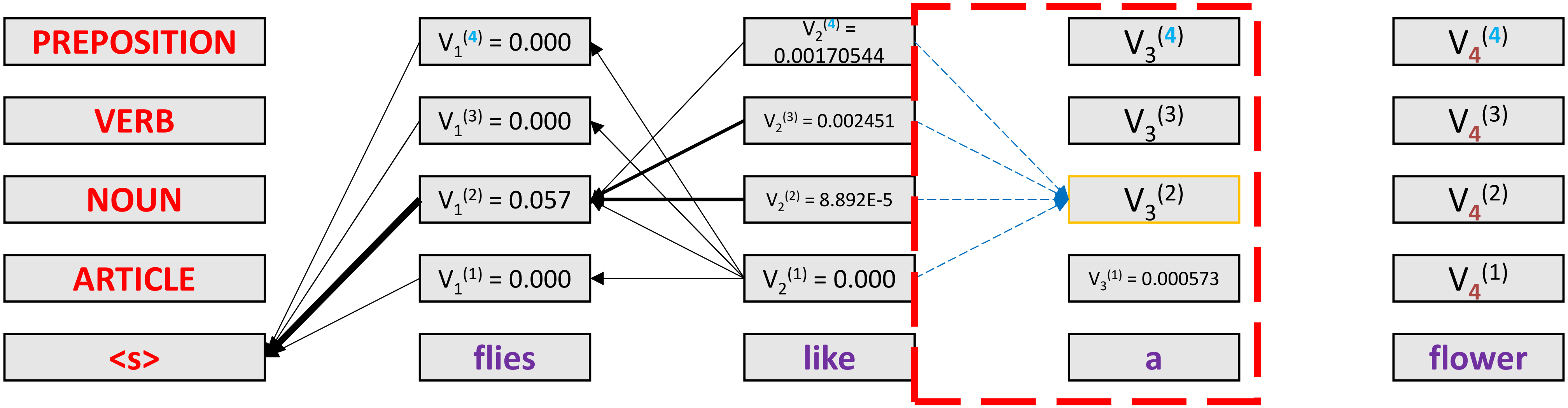
Emission probability matrix				
	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Pseudocode

function VITERBI(*observations* of len T,*state-graph* of len N) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state s **from** 1 **to** N **do** ; initialization step
$\quad viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
$\quad backpointer[s,1] \leftarrow 0$
for each time step t **from** 2 **to** T **do** ; recursion step
$\quad$ **for** each state s **from** 1 **to** N **do**
$$viterbi[s,t] \leftarrow \max_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$$
$$backpointer[s,t] \leftarrow \operatorname*{argmax}_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$$
$bestpathprob \leftarrow \max_{s=1}^{N} viterbi[s,T]$; termination step
$bestpathpointer \leftarrow \operatorname*{argmax}_{s=1}^{N} viterbi[s,T]$; termination step
bestpath ← the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Algorithm: Example



$V_3^{(2)}$ = viterbi[NOUN, a] = $\max_{s'}$(viterbi[state s', a] * $a_{s',2}$ * b_2(a)) = max(

$V_2^{(1)}$ * P(NOUN | ARTICLE) * P(a | NOUN) = 0.000 * 1.00 * 0.001 = 0.000

$V_2^{(2)}$ * P(NOUN | NOUN) * P(a | NOUN) = 8.892E-5 * 0.13 * 0.001 = 1.15596E-8

$V_2^{(3)}$ * P(NOUN | VERB) * P(a | NOUN) = 0.002451 * 0.35 * 0.001 = 8.5785E-7

$V_2^{(4)}$ * P(NOUN | PREPOSITION) * P(a | NOUN) = 0.00170544 * 0.26 * 0.001 = 4.434144E-7

Transition probability matrix

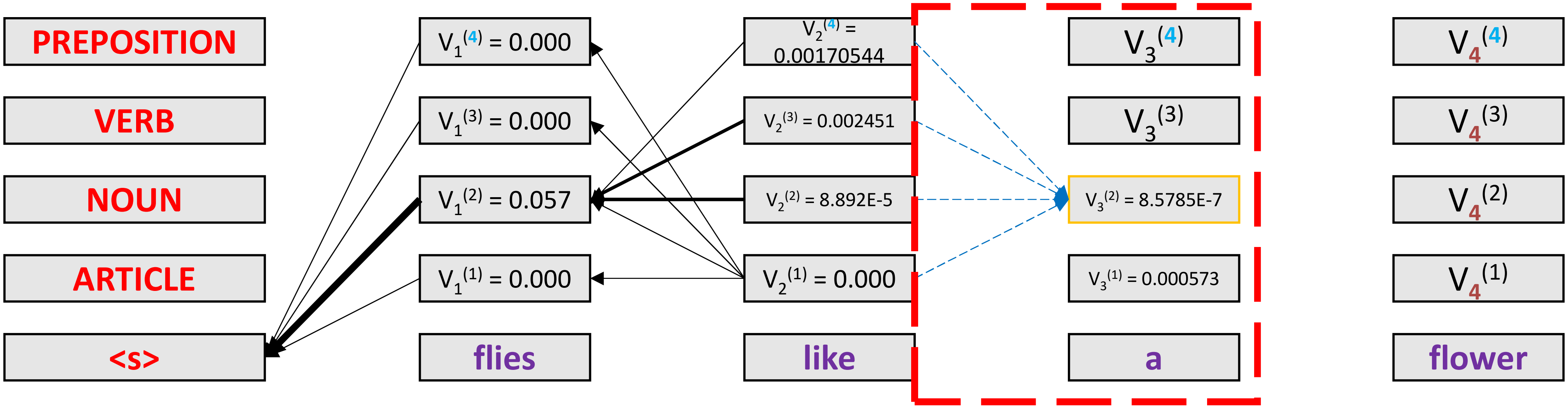
	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Example



$V_3^{(2)}$ = viterbi[NOUN, a] = $\max_{s'}$(viterbi[state s', a] * $a_{s',2}$ * b_2(a)) = 8.5785E-7

Transition probability matrix

	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

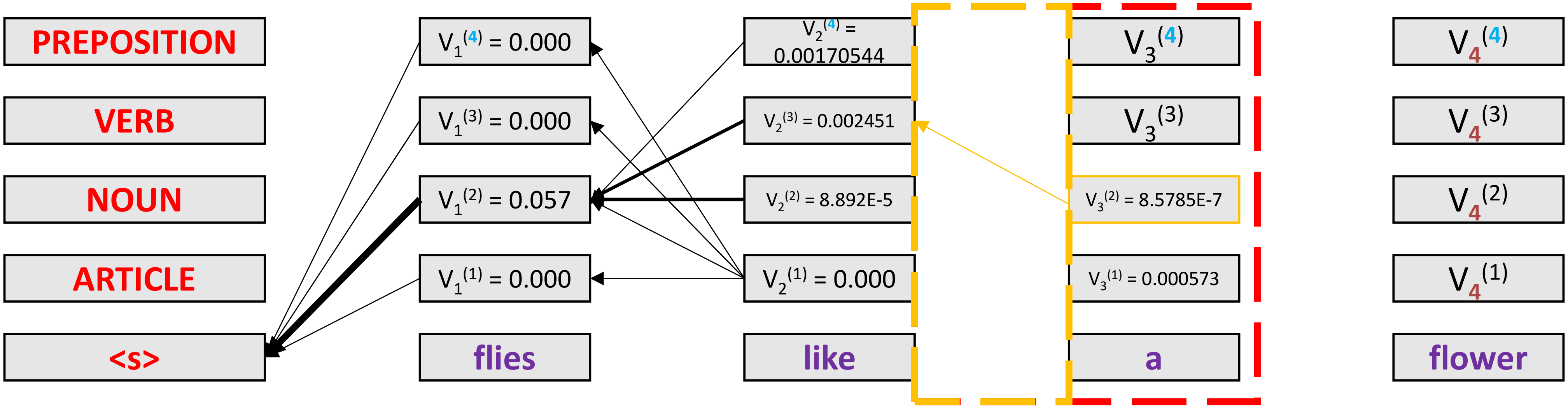
	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Pseudocode

function VITERBI(*observations* of len *T*,*state-graph* of len *N*) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state *s* **from** 1 **to** *N* **do** ; initialization step
    $viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
    $backpointer[s,1] \leftarrow 0$
for each time step *t* **from** 2 **to** *T* **do** ; recursion step
    for each state *s* **from** 1 **to** *N* **do**
        $viterbi[s,t] \leftarrow \max_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
        $backpointer[s,t] \leftarrow \operatorname{argmax}_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
$bestpathprob \leftarrow \max_{s=1}^{N} viterbi[s,T]$; termination step
$bestpathpointer \leftarrow \operatorname{argmax}_{s=1}^{N} viterbi[s,T]$; termination step
bestpath ← the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Algorithm: Example



$V_3^{(2)}$ = viterbi[NOUN, a] = $\max_{s'}$(viterbi[state s', a] * $a_{s',2}$ * b_2(a)) = 8.5785E-7

Transition probability matrix

	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

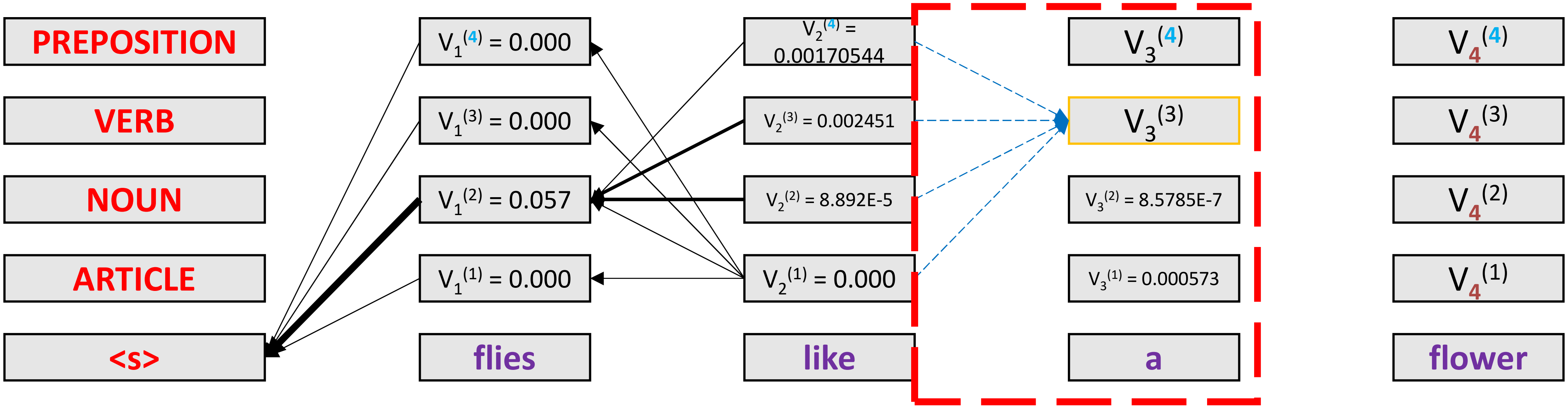
	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Pseudocode

function VITERBI(*observations* of len *T*,*state-graph* of len *N*) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state *s* **from** 1 **to** *N* **do** ; initialization step
  $viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
  $backpointer[s,1] \leftarrow 0$
for each time step *t* **from** 2 **to** *T* **do** ; recursion step
  for each state *s* **from** 1 **to** *N* **do**
    $viterbi[s,t] \leftarrow \max_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
    $backpointer[s,t] \leftarrow \operatorname{argmax}_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
$bestpathprob \leftarrow \max_{s=1}^{N} viterbi[s,T]$; termination step
$bestpathpointer \leftarrow \operatorname{argmax}_{s=1}^{N} viterbi[s,T]$; termination step
bestpath ← the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Algorithm: Example



$V_3^{(3)}$ = viterbi[**VERB**, **a**] = $\max_{s'}$(viterbi[state s', **a**] * $a_{s',3}$ * b_3(**a**)) = max(

$V_2^{(1)}$ * P(**VERB** | **ARTICLE**) * P(**a** | **VERB**) = 0.000 * 0.00 * 0.000 = 0.000
$V_2^{(2)}$ * P(**VERB** | **NOUN**) * P(**a** | **VERB**) = 8.892E-5 * 0.43 * 0.000 = 0.000
$V_2^{(3)}$ * P(**VERB** | **VERB**) * P(**a** | **VERB**) = 0.002451 * 0.00 * 0.000 = 0.000
$V_2^{(4)}$ * P(**VERB** | **PREPOSITION**) * P(**a** | **VERB**) = 0.00170544 * 0.00 * 0.000 = 0.000

)

Transition probability matrix

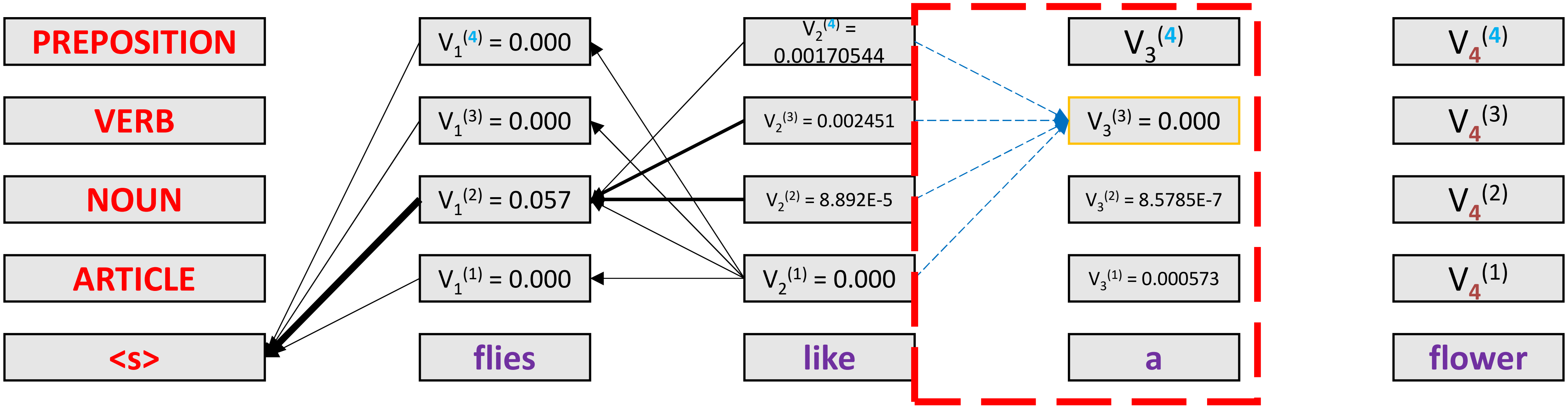
	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Example



$V_3^{(3)}$ = viterbi[VERB, a] = $\max_{s'}$(viterbi[state s', a] * $a_{s',3}$ * b_3(a)) = 0.000

Transition probability matrix					
	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

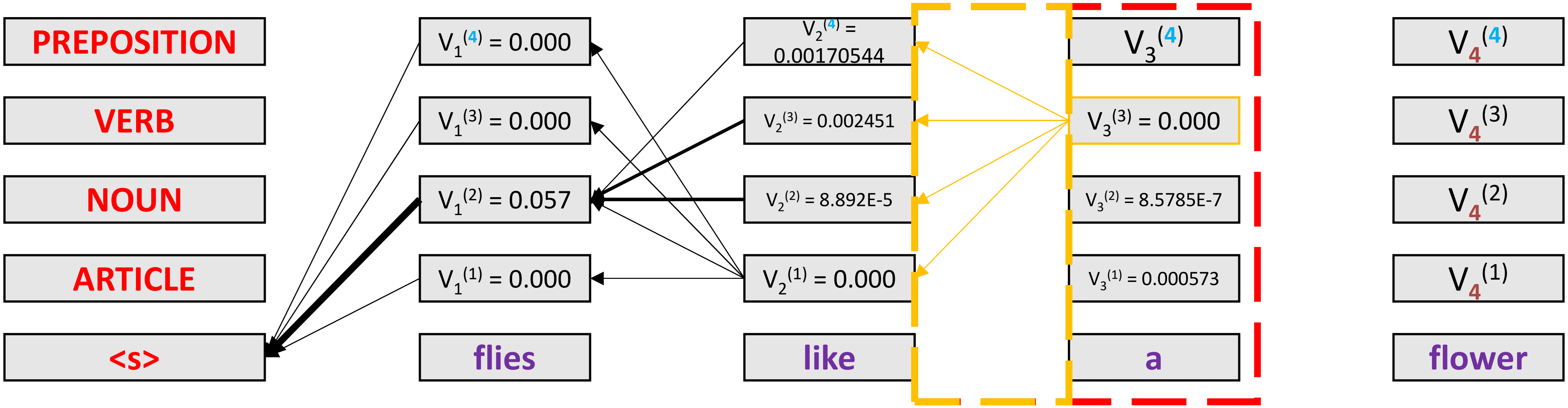
Emission probability matrix				
	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Pseudocode

function VITERBI(*observations* of len *T*,*state-graph* of len *N*) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state *s* **from** 1 **to** *N* **do** ; initialization step
    $viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
    $backpointer[s,1] \leftarrow 0$
for each time step *t* **from** 2 **to** *T* **do** ; recursion step
  for each state *s* **from** 1 **to** *N* **do**
    $viterbi[s,t] \leftarrow \max_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
    $backpointer[s,t] \leftarrow \operatorname{argmax}_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
$bestpathprob \leftarrow \max_{s=1}^{N} viterbi[s,T]$; termination step
$bestpathpointer \leftarrow \operatorname{argmax}_{s=1}^{N} viterbi[s,T]$; termination step
bestpath ← the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Algorithm: Example



$V_3^{(3)}$ = viterbi[VERB, a] = $\max_{s'}$(viterbi[state s', a] * $a_{s',3}$ * b_3(a)) = 0.000

Transition probability matrix

	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

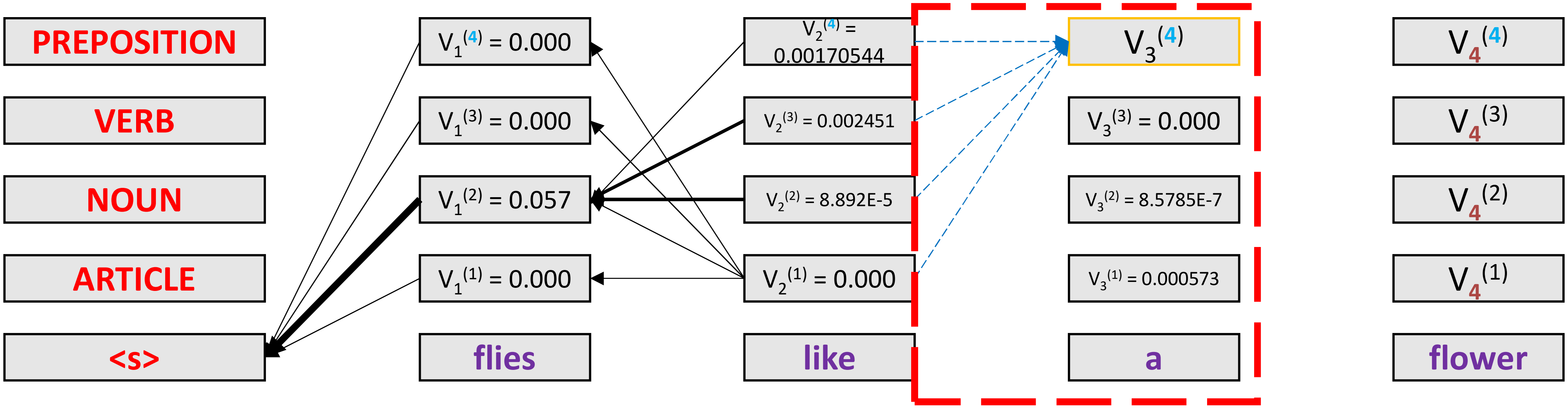
	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Pseudocode

function VITERBI(*observations* of len T,*state-graph* of len N) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state s **from** 1 **to** N **do** ; initialization step
$viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
$backpointer[s,1] \leftarrow 0$
for each time step t **from** 2 **to** T **do** ; recursion step
for each state s **from** 1 **to** N **do**
$viterbi[s,t] \leftarrow \max_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
$backpointer[s,t] \leftarrow \operatorname{argmax}_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
$bestpathprob \leftarrow \max_{s=1}^{N} viterbi[s,T]$; termination step
$bestpathpointer \leftarrow \operatorname{argmax}_{s=1}^{N} viterbi[s,T]$; termination step
bestpath $\leftarrow$ the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Algorithm: Example



$V_3^{(4)}$ = viterbi[PREPOSITION, a] = $\max_{s'}$(viterbi[state s', a] * $a_{s',4}$ * b_4(a)) = max(

$V_2^{(1)}$ * P(PREPOSITION | ARTICLE) * P(a | PREPOSITION) = 0.000 * 0.00 * 0.000 = 0.000
$V_2^{(2)}$ * P(PREPOSITION | NOUN) * P(a | PREPOSITION) = 8.892E-5 * 0.44 * 0.000 = 0.000
$V_2^{(3)}$ * P(PREPOSITION | VERB) * P(a | PREPOSITION) = 0.002451 * 0.00 * 0.000 = 0.000
$V_2^{(4)}$ * P(PREPOSITION | PREPOSITION) * P(a | PREPOSITION) = 0.00170544 * 0.00 * 0.000 = 0.000

)

Transition probability matrix

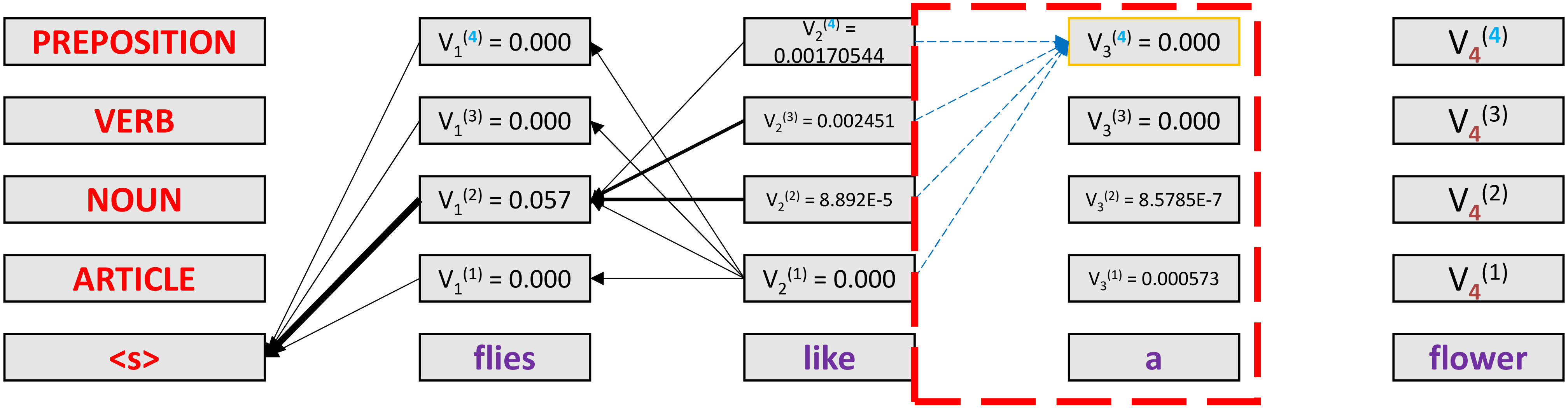
	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Example



$V_3^{(4)}$ = viterbi[PREPOSITION, a] = $\max_{s'}$(viterbi[state s', a] * $a_{s',4}$ * b_4(a)) = 0.000

Transition probability matrix	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

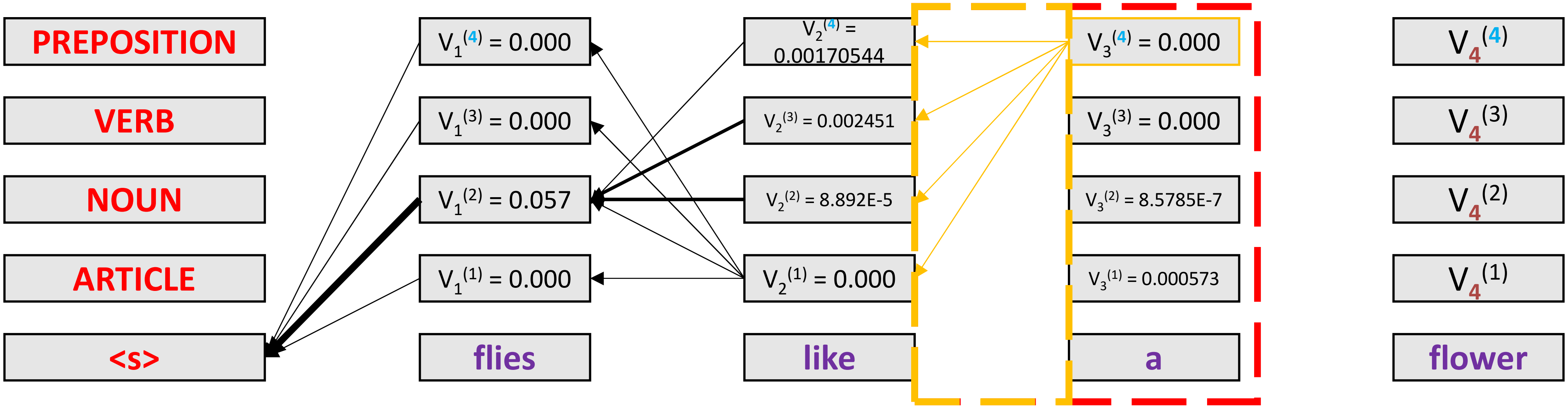
Emission probability matrix	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Pseudocode

function VITERBI(*observations* of len T,*state-graph* of len N) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state s **from** 1 **to** N **do** ; initialization step
    $viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
    $backpointer[s,1] \leftarrow 0$
for each time step t **from** 2 **to** T **do** ; recursion step
    for each state s **from** 1 **to** N **do**
        $viterbi[s,t] \leftarrow \max_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
        $backpointer[s,t] \leftarrow \operatorname{argmax}_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
$bestpathprob \leftarrow \max_{s=1}^{N} viterbi[s,T]$; termination step
$bestpathpointer \leftarrow \operatorname{argmax}_{s=1}^{N} viterbi[s,T]$; termination step
bestpath ← the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Algorithm: Example



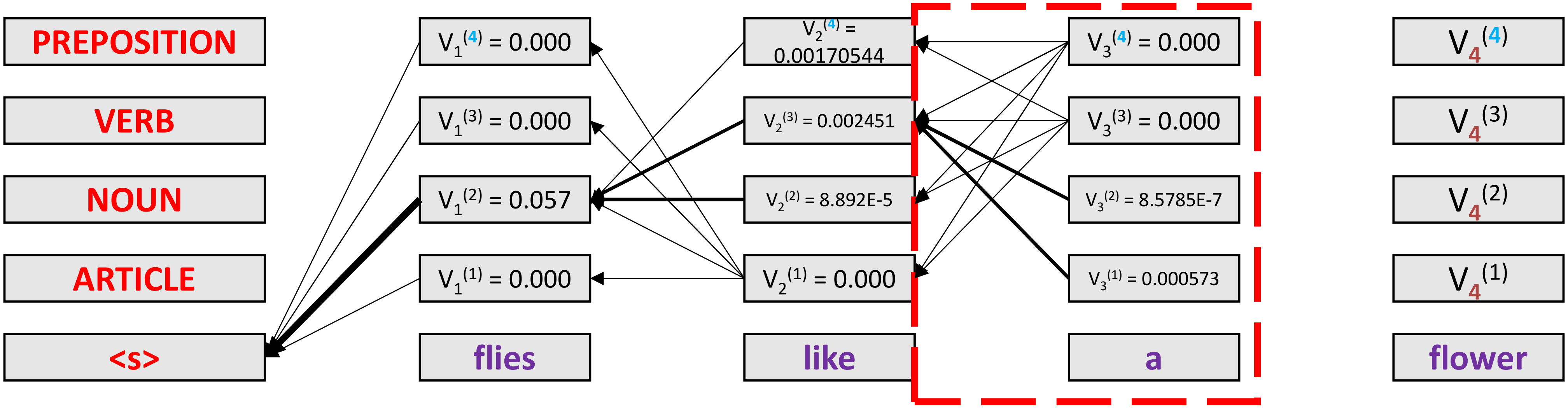
$V_3^{(4)}$ = viterbi[PREPOSITION, a] = $\max_{s'}$(viterbi[state s', a] * $a_{s',4}$ * b_4(a)) = 0.000

Transition probability matrix					
	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix				
	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Example



Transition probability matrix

	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Pseudocode

function VITERBI(*observations* of len T,*state-graph* of len N) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state s **from** 1 **to** N **do** ; initialization step
$viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
$backpointer[s,1] \leftarrow 0$
for each time step t **from** 2 **to** T **do** ; recursion step
for each state s **from** 1 **to** N **do**
$$viterbi[s,t] \leftarrow \max_{s'=1}^{N} \; viterbi[s',t-1] * a_{s',s} * b_s(o_t)$$
$$backpointer[s,t] \leftarrow \operatorname{argmax}_{s'=1}^{N} \; viterbi[s',t-1] * a_{s',s} * b_s(o_t)$$

t = 4
Fourth column

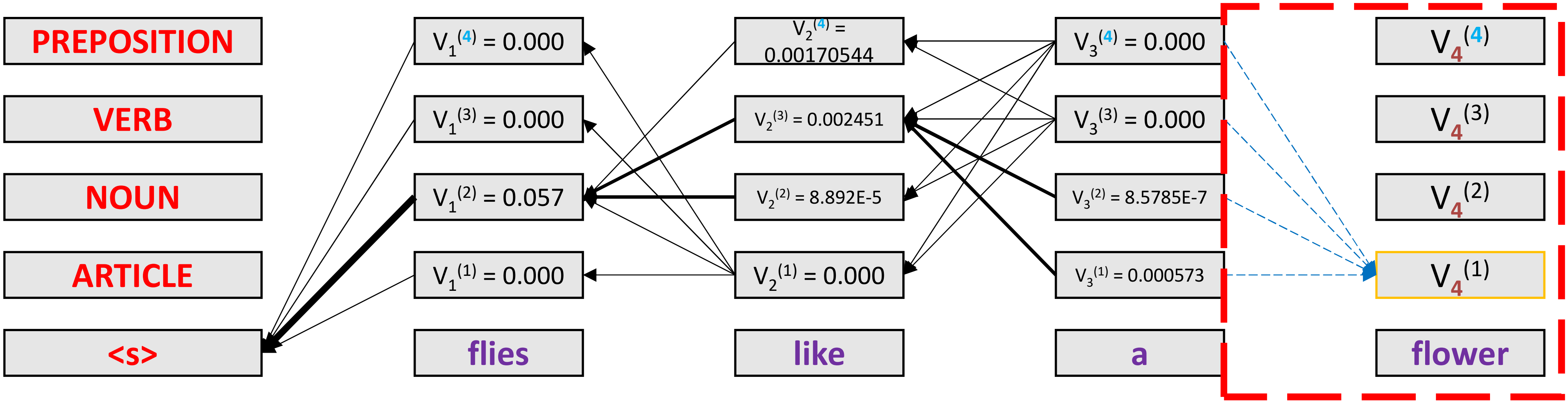
$bestpathprob \leftarrow \max_{s=1}^{N} \; viterbi[s,T]$; termination step
$bestpathpointer \leftarrow \operatorname{argmax}_{s=1}^{N} \; viterbi[s,T]$; termination step
bestpath ← the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Algorithm: Pseudocode

function VITERBI(*observations* of len T,*state-graph* of len N) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state s **from** 1 **to** N **do** ; initialization step
    $viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
    $backpointer[s,1] \leftarrow 0$
for each time step t **from** 2 **to** T **do** ; recursion step
    for each state s **from** 1 **to** N **do**
        $viterbi[s,t] \leftarrow \max_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
        $backpointer[s,t] \leftarrow \operatorname{argmax}_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
$bestpathprob \leftarrow \max_{s=1}^{N} viterbi[s,T]$; termination step
$bestpathpointer \leftarrow \operatorname{argmax}_{s=1}^{N} viterbi[s,T]$; termination step
bestpath ← the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Algorithm: Example



$V_4^{(1)}$ = viterbi[ARTICLE, flower] = $\max_{s'}$(viterbi[state s', flower] * $a_{s',1}$ * b_1(flower)) = max(

$V_3^{(1)}$ * P(ARTICLE | ARTICLE) * P(flower | ARTICLE) = 0.000573 * 0.00 * 0.000 = 0.000
$V_3^{(2)}$ * P(ARTICLE | NOUN) * P(flower | ARTICLE) = 8.5785E-7 * 0.00 * 0.000 = 0.000
$V_3^{(3)}$ * P(ARTICLE | VERB) * P(flower | ARTICLE) = 0.000 * 0.65 * 0.000 = 0.000
$V_3^{(4)}$ * P(ARTICLE | PREPOSITION) * P(flower | ARTICLE) = 0.000 * 0.74 * 0.000 = 0.000
)

Transition probability matrix

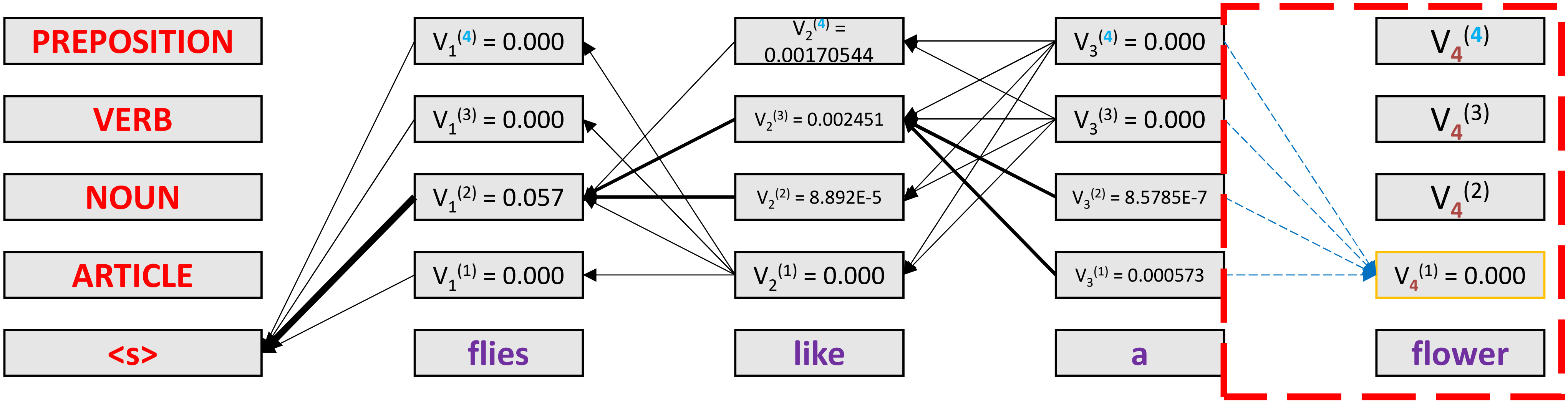
	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Example



$V_4^{(1)}$ = viterbi[ARTICLE, flower] = $\max_{s'}$(viterbi[state s', flower] * $a_{s',1}$ * b_1(flower)) = 0.000

Transition probability matrix

	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

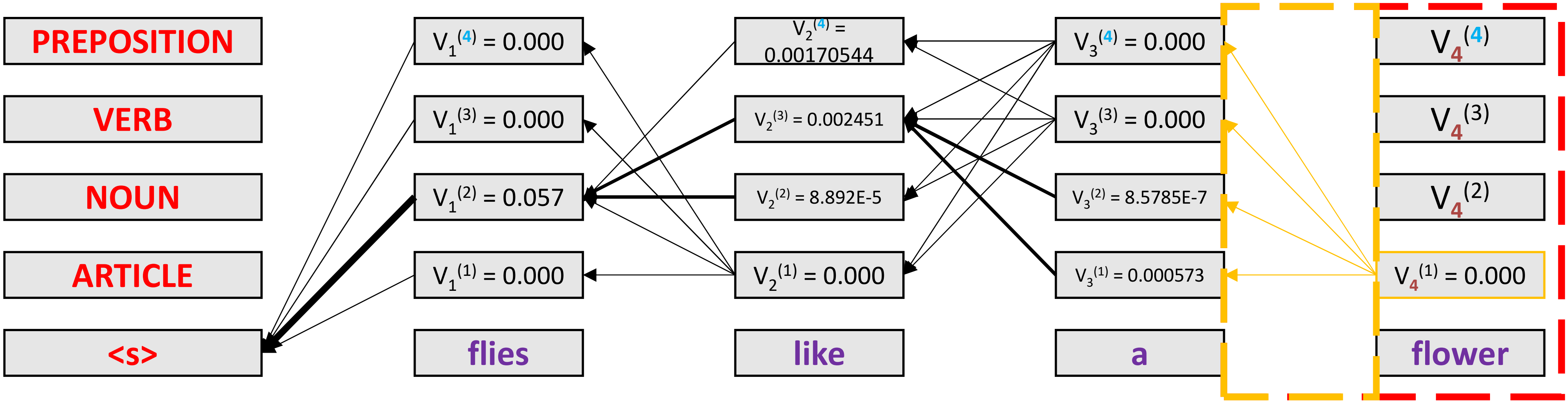
	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Pseudocode

function VITERBI(*observations* of len T,*state-graph* of len N) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state s **from** 1 **to** N **do** ; initialization step
    $viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
    $backpointer[s,1] \leftarrow 0$
for each time step t **from** 2 **to** T **do** ; recursion step
    for each state s **from** 1 **to** N **do**
        $viterbi[s,t] \leftarrow \max_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
        $backpointer[s,t] \leftarrow \operatorname{argmax}_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
$bestpathprob \leftarrow \max_{s=1}^{N} viterbi[s,T]$; termination step
$bestpathpointer \leftarrow \operatorname{argmax}_{s=1}^{N} viterbi[s,T]$; termination step
bestpath ← the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Algorithm: Example



$V_4^{(1)}$ = viterbi[ARTICLE, flower] = $\max_{s'}$(viterbi[state s', flower] * $a_{s',1}$ * b_1(flower)) = 0.000

Transition probability matrix

	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

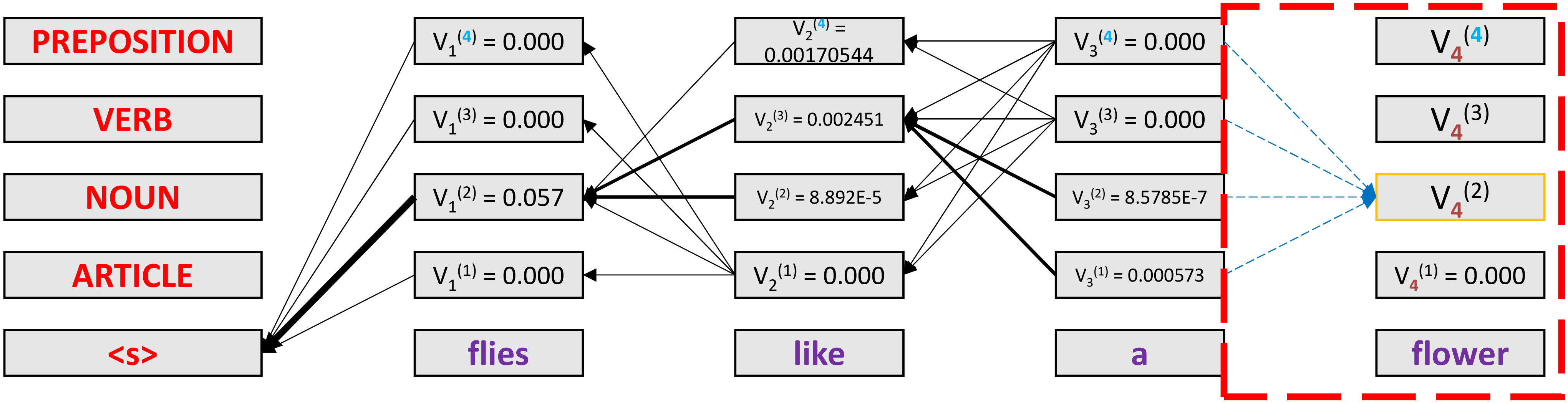
	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Pseudocode

function VITERBI(*observations* of len T,*state-graph* of len N) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state s **from** 1 **to** N **do** ; initialization step
$viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
$backpointer[s,1] \leftarrow 0$
for each time step t **from** 2 **to** T **do** ; recursion step
for each state s **from** 1 **to** N **do**
$viterbi[s,t] \leftarrow \max_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
$backpointer[s,t] \leftarrow \operatorname{argmax}_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
$bestpathprob \leftarrow \max_{s=1}^{N} viterbi[s,T]$; termination step
$bestpathpointer \leftarrow \operatorname{argmax}_{s=1}^{N} viterbi[s,T]$; termination step
bestpath $\leftarrow$ the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Algorithm: Example



$V_4^{(2)}$ = viterbi[NOUN, flower] = $\max_{s'}$(viterbi[state s', flower] * $a_{s',2}$ * b_2(flower)) = max(

$V_3^{(1)}$ * P(NOUN | ARTICLE) * P(flower | NOUN) = 0.000573 * 1.00 * 0.063 = <u>3.6099E-5</u>

$V_3^{(2)}$ * P(NOUN | NOUN) * P(flower | NOUN) = 8.5785E-7 * 0.13 * 0.063 = 7.0257915E-9

$V_3^{(3)}$ * P(NOUN | VERB) * P(flower | NOUN) = 0.000 * 0.35 * 0.063 = 0.000

$V_3^{(4)}$ * P(NOUN | PREPOSITION) * P(flower | NOUN) = 0.000 * 0.26 * 0.063 = 0.000

)

Transition probability matrix

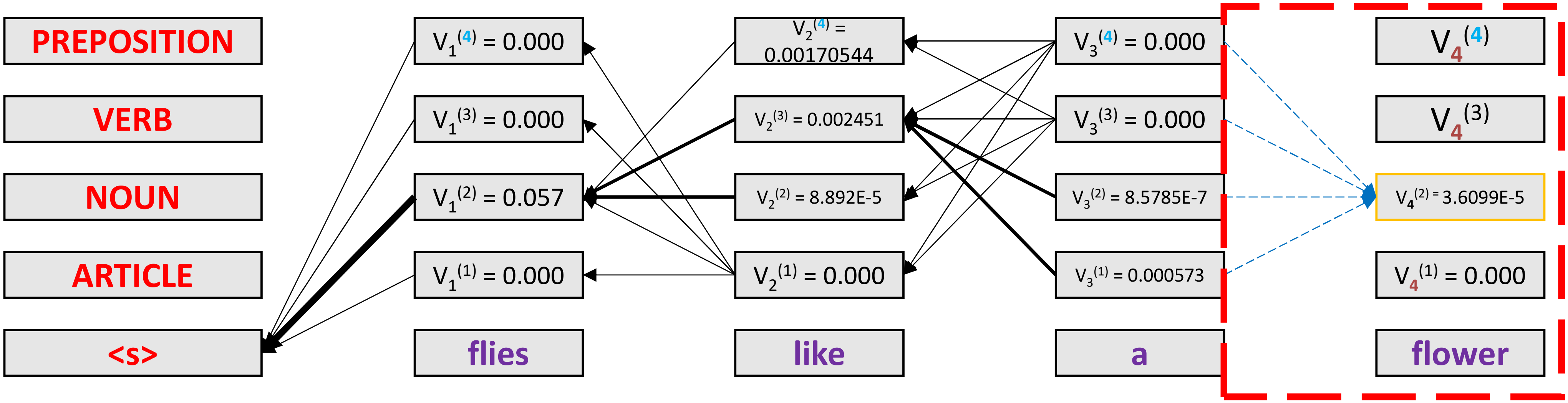
	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Example



$V_4^{(2)}$ = viterbi[NOUN, flower] = $\max_{s'}$(viterbi[state s', flower] * $a_{s',2}$ * b_2(flower)) = 3.6099E-5

Transition probability matrix

	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

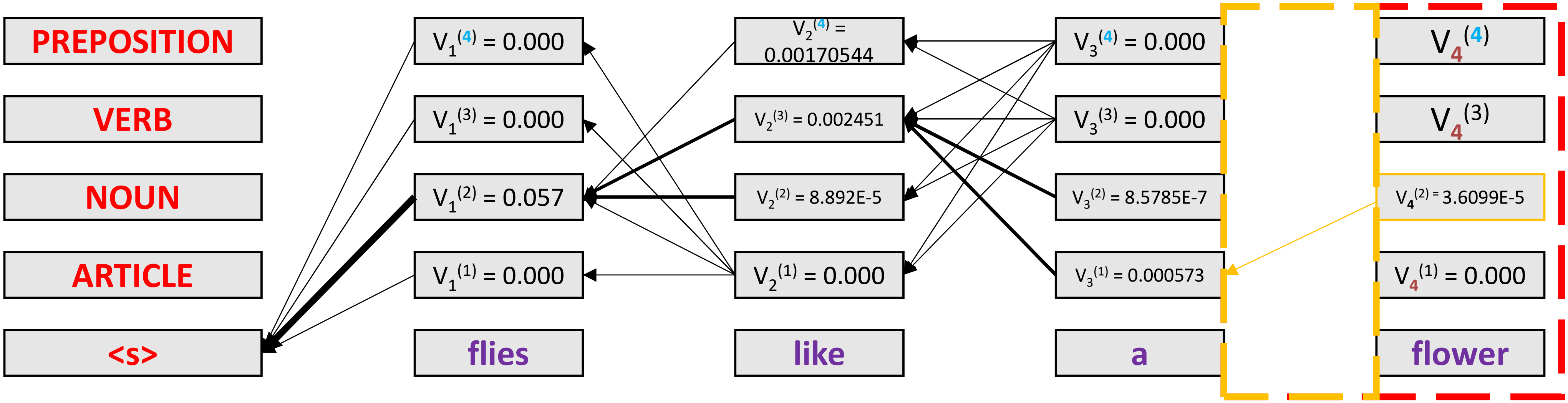
	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Pseudocode

function VITERBI(*observations* of len T,*state-graph* of len N) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state s **from** 1 **to** N **do** ; initialization step
$\quad viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
$\quad backpointer[s,1] \leftarrow 0$
for each time step t **from** 2 **to** T **do** ; recursion step
$\quad$ **for** each state s **from** 1 **to** N **do**
$$viterbi[s,t] \leftarrow \max_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$$
$$backpointer[s,t] \leftarrow \operatorname{argmax}_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$$
$bestpathprob \leftarrow \max_{s=1}^{N} viterbi[s,T]$; termination step
$bestpathpointer \leftarrow \operatorname{argmax}_{s=1}^{N} viterbi[s,T]$; termination step
$bestpath \leftarrow$ the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Algorithm: Example



$V_4^{(2)}$ = viterbi[NOUN, flower] = $\max_{s'}$(viterbi[state s', flower] * $a_{s',2}$ * b_2(flower)) = 3.6099E-5

Transition probability matrix

	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

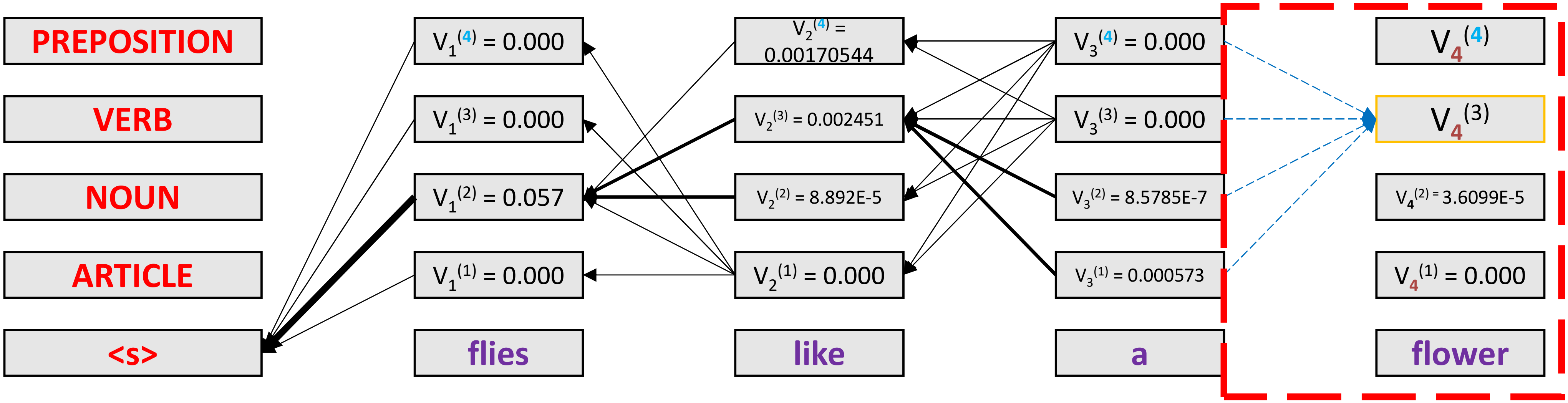
	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Pseudocode

function VITERBI(*observations* of len T,*state-graph* of len N) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state s **from** 1 **to** N **do** ; initialization step
$viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
$backpointer[s,1] \leftarrow 0$
for each time step t **from** 2 **to** T **do** ; recursion step
for each state s **from** 1 **to** N **do**
$\boxed{viterbi[s,t] \leftarrow \max_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)}$
$backpointer[s,t] \leftarrow \operatorname{argmax}_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
$bestpathprob \leftarrow \max_{s=1}^{N} viterbi[s,T]$; termination step
$bestpathpointer \leftarrow \operatorname{argmax}_{s=1}^{N} viterbi[s,T]$; termination step
bestpath ← the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Algorithm: Example



$V_4^{(3)}$ = viterbi[VERB, flower] = $\max_{s'}$(viterbi[state s', flower] * $a_{s',3}$ * b_3(flower)) = max(

$V_3^{(1)}$ * P(VERB | ARTICLE) * P(flower | VERB) = 0.000573 * 0.00 * 0.050 = 0.000

$V_3^{(2)}$ * P(VERB | NOUN) * P(flower | VERB) = 8.5785E-7 * 0.43 * 0.050 = <u>1.8443775E-8</u>

$V_3^{(3)}$ * P(VERB | VERB) * P(flower | VERB) = 0.000 * 0.00 * 0.050 = 0.000

$V_3^{(4)}$ * P(VERB | PREPOSITION) * P(flower | VERB) = 0.000 * 0.00 * 0.050 = 0.000

)

Transition probability matrix

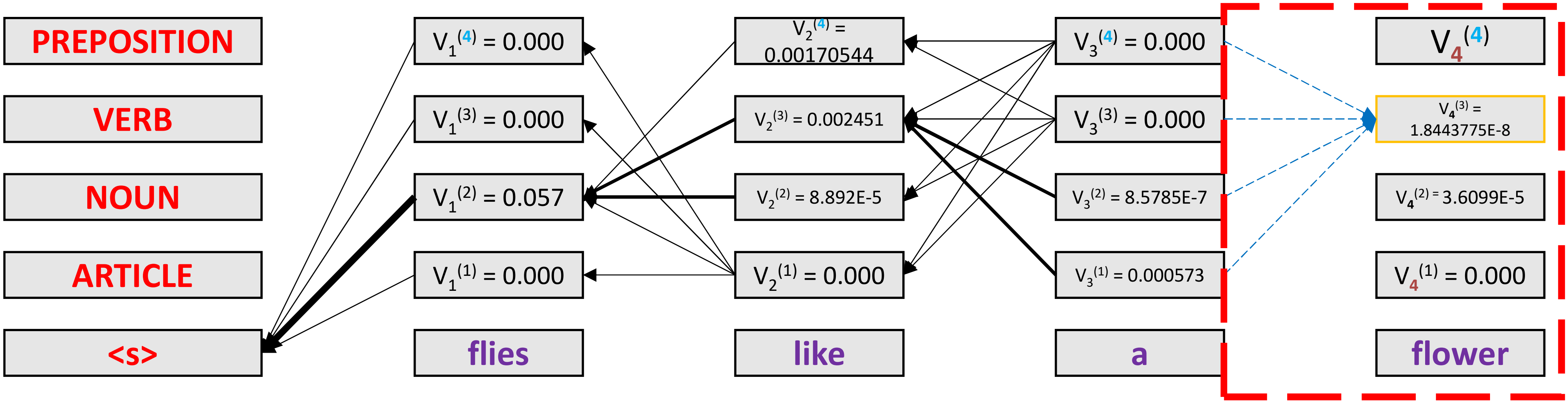
	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Example



$V_4^{(3)}$ = viterbi[VERB, flower] = $\max_{s'}$(viterbi[state s', flower] * $a_{s',3}$ * b_3(flower)) = 1.8443775E-8

Transition probability matrix

	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

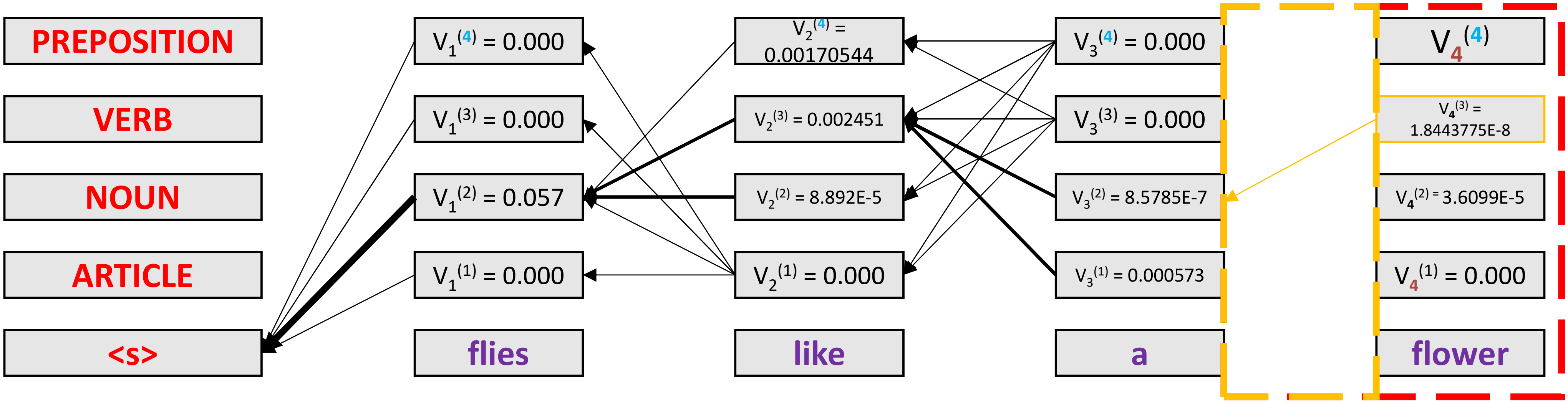
	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Pseudocode

function VITERBI(*observations* of len T,*state-graph* of len N) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state s **from** 1 **to** N **do** ; initialization step
  $viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
  $backpointer[s,1] \leftarrow 0$
for each time step t **from** 2 **to** T **do** ; recursion step
  for each state s **from** 1 **to** N **do**
   $viterbi[s,t] \leftarrow \max_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
   $backpointer[s,t] \leftarrow \operatorname{argmax}_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
$bestpathprob \leftarrow \max_{s=1}^{N} viterbi[s,T]$; termination step
$bestpathpointer \leftarrow \operatorname{argmax}_{s=1}^{N} viterbi[s,T]$; termination step
bestpath ← the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Algorithm: Example



$V_4^{(3)}$ = viterbi[VERB, flower] = $\max_{s'}$(viterbi[state s', flower] * $a_{s',3}$ * b_3(flower)) = 1.8443775E-8

Transition probability matrix

	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

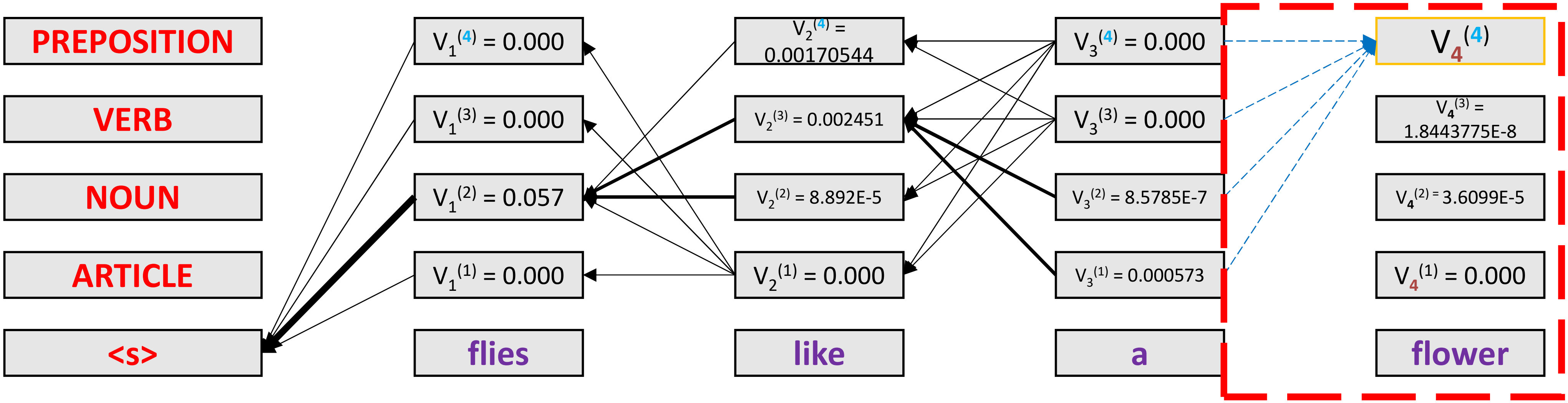
	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Pseudocode

function VITERBI(*observations* of len T,*state-graph* of len N) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state s **from** 1 **to** N **do** ; initialization step
 $viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
 $backpointer[s,1] \leftarrow 0$
for each time step t **from** 2 **to** T **do** ; recursion step
 for each state s **from** 1 **to** N **do**
 $viterbi[s,t] \leftarrow \max_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
 $backpointer[s,t] \leftarrow \operatorname{argmax}_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
$bestpathprob \leftarrow \max_{s=1}^{N} viterbi[s,T]$; termination step
$bestpathpointer \leftarrow \operatorname{argmax}_{s=1}^{N} viterbi[s,T]$; termination step
bestpath ← the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Algorithm: Example



$V_4^{(4)}$ = viterbi[PREPOSITION, flower] = $\max_{s'}$(viterbi[state s', flower] * $a_{s',4}$ * b_4(flower)) = max(

$V_3^{(1)}$ * P(PREPOSITION | ARTICLE) * P(flower | PREPOSITION) = 0.000573 * 0.00 * 0.050 = 0.000

$V_3^{(2)}$ * P(PREPOSITION | NOUN) * P(flower | PREPOSITION) = 8.5785E-7 * 0.44 * 0.050 = <u>1.8443775E-8</u>

$V_3^{(3)}$ * P(PREPOSITION | VERB) * P(flower | PREPOSITION) = 0.000 * 0.00 * 0.050 = 0.000

$V_3^{(4)}$ * P(PREPOSITION | PREPOSITION) * P(flower | PREPOSITION) = 0.000 * 0.00 * 0.050 = 0.000

)

Transition probability matrix

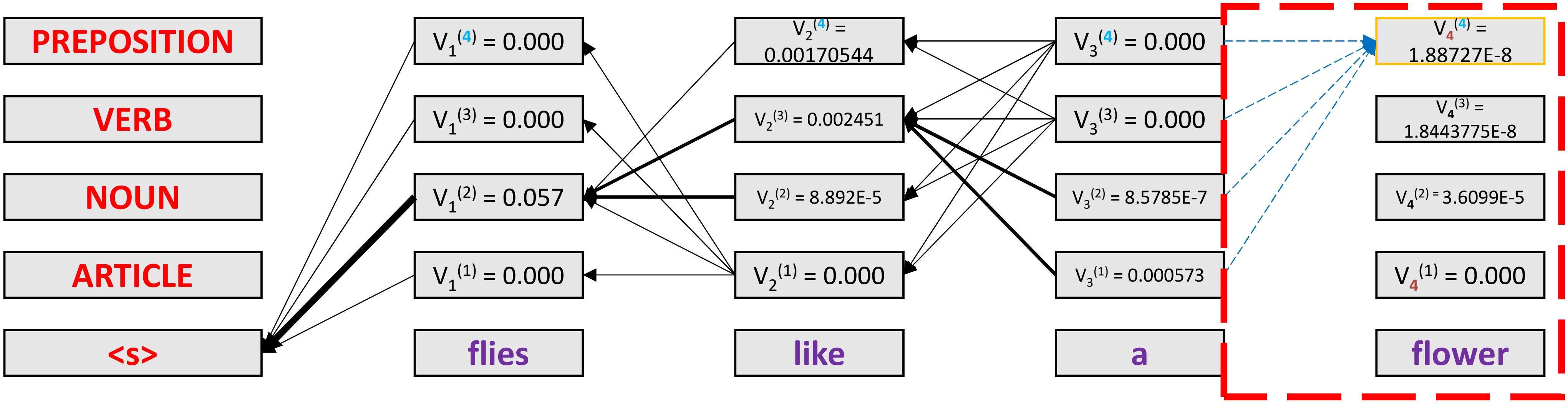
	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Example



$V_4^{(4)}$ = viterbi[**PREPOSITION**, **flower**] = $\max_{s'}$(viterbi[state s', **flower**] * $a_{s',4}$ * b_4(**flower**)) = 1.8443775E-8

Transition probability matrix

	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

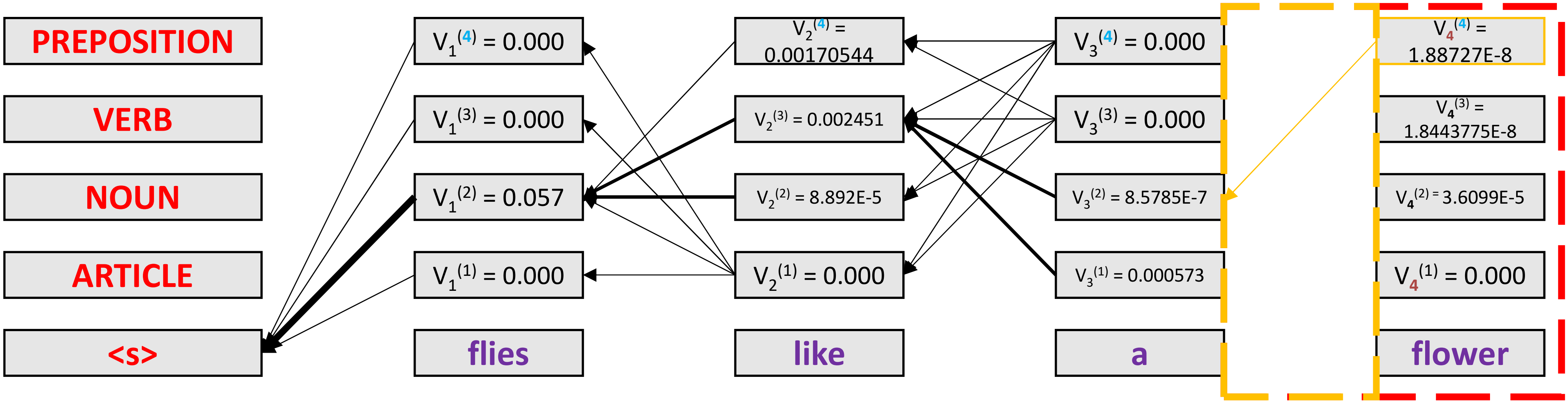
	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Pseudocode

function VITERBI(*observations* of len T,*state-graph* of len N) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state s **from** 1 **to** N **do** ; initialization step
$\quad viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
$\quad backpointer[\mathrm{s},1] \leftarrow 0$
for each time step t **from** 2 **to** T **do** ; recursion step
$\quad$ **for** each state s **from** 1 **to** N **do**
$$viterbi[\mathrm{s,t}] \leftarrow \max_{s'=1}^{N} \; viterbi[s',t-1] * a_{s',s} * b_s(o_t)$$
$$backpointer[\mathrm{s,t}] \leftarrow \operatorname*{argmax}_{s'=1}^{N} \; viterbi[s',t-1] * a_{s',s} * b_s(o_t)$$
$$bestpathprob \leftarrow \max_{s=1}^{N} \; viterbi[s,T] \qquad \text{; termination step}$$
$$bestpathpointer \leftarrow \operatorname*{argmax}_{s=1}^{N} \; viterbi[s,T] \qquad \text{; termination step}$$
bestpath ← the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Algorithm: Example



$V_4^{(4)}$ = viterbi[PREPOSITION, flower] = $\max_{s'}$(viterbi[state s', flower] * $a_{s',4}$ * b_4(flower)) = 1.8443775E-8

Transition probability matrix

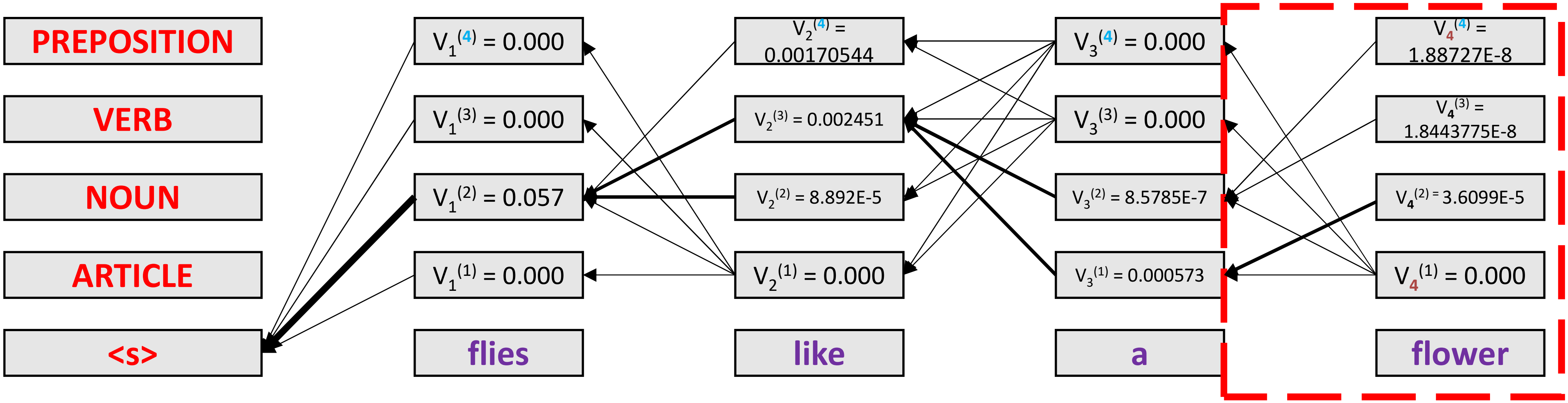
	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Example



Transition probability matrix

	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

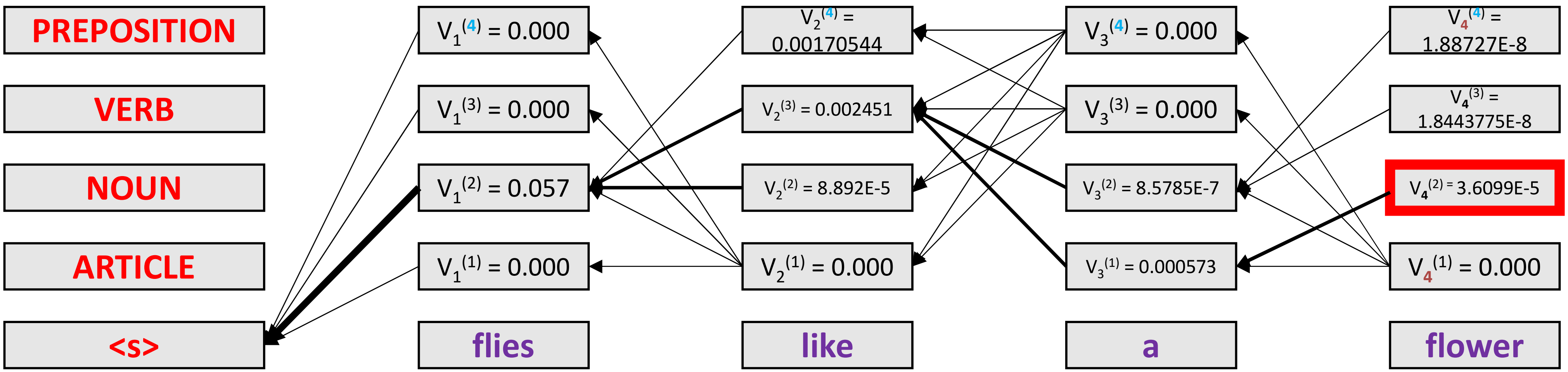
	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Pseudocode

function VITERBI(*observations* of len T,*state-graph* of len N) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state s **from** 1 **to** N **do** ; initialization step
$\quad viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
$\quad backpointer[s,1] \leftarrow 0$
for each time step t **from** 2 **to** T **do** ; recursion step
$\quad$ **for** each state s **from** 1 **to** N **do**
$$viterbi[s,t] \leftarrow \max_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$$
$$backpointer[s,t] \leftarrow \underset{s'=1}{\overset{N}{\operatorname{argmax}}} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$$
$$bestpathprob \leftarrow \max_{s=1}^{N} viterbi[s,T]$$; termination step
$$bestpathpointer \leftarrow \underset{s=1}{\overset{N}{\operatorname{argmax}}} viterbi[s,T]$$; termination step
bestpath $\leftarrow$ the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Algorithm: Example



Transition probability matrix

	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

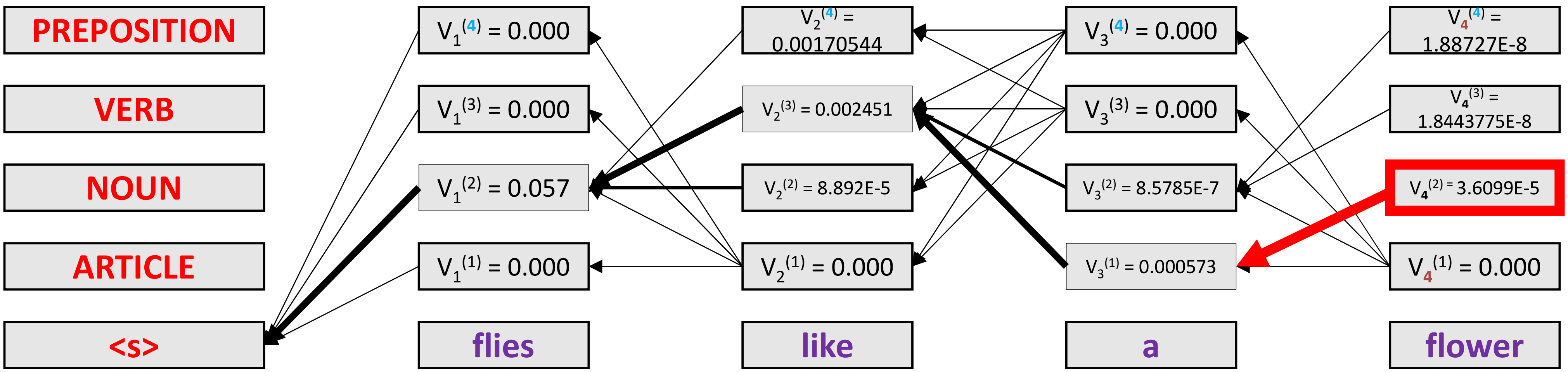
	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Pseudocode

function VITERBI(*observations* of len T,*state-graph* of len N) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state s **from** 1 **to** N **do** ; initialization step
$\quad viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
$\quad backpointer[s,1] \leftarrow 0$
for each time step t **from** 2 **to** T **do** ; recursion step
for each state s **from** 1 **to** N **do**
$$viterbi[s,t] \leftarrow \max_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$$
$$backpointer[s,t] \leftarrow \operatorname{argmax}_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$$
$bestpathprob \leftarrow \max_{s=1}^{N} viterbi[s,T]$; termination step
$\boxed{bestpathpointer \leftarrow \operatorname{argmax}_{s=1}^{N} viterbi[s,T] \quad \text{; termination step}}$
bestpath ← the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Algorithm: Example



Transition probability matrix

	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

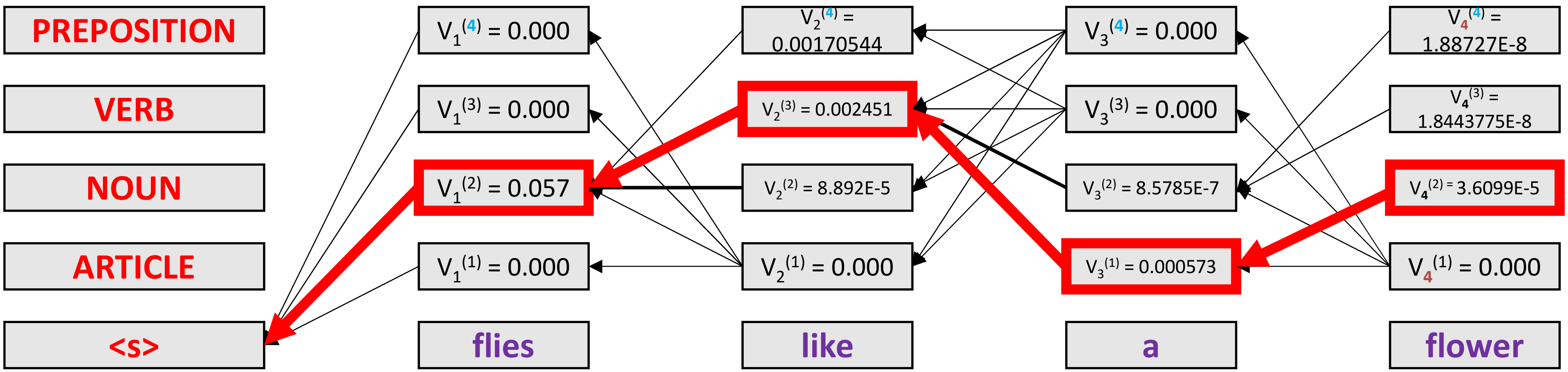
	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000

Viterbi Algorithm: Pseudocode

function VITERBI(*observations* of len T,*state-graph* of len N) **returns** *best-path*, *path-prob*

create a path probability matrix *viterbi[N,T]*
for each state s **from** 1 **to** N **do** ; initialization step
  $viterbi[s,1] \leftarrow \pi_s * b_s(o_1)$
  $backpointer[s,1] \leftarrow 0$
for each time step t **from** 2 **to** T **do** ; recursion step
  for each state s **from** 1 **to** N **do**
    $viterbi[s,t] \leftarrow \max_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
    $backpointer[s,t] \leftarrow \operatorname{argmax}_{s'=1}^{N} viterbi[s',t-1] * a_{s',s} * b_s(o_t)$
$bestpathprob \leftarrow \max_{s=1}^{N} viterbi[s,T]$; termination step
$bestpathpointer \leftarrow \operatorname{argmax}_{s=1}^{N} viterbi[s,T]$; termination step
bestpath ← the path starting at state *bestpathpointer*, that follows backpointer[] to states back in time
return *bestpath*, *bestpathprob*

Viterbi Algorithm: Example



SOLUTION:

POS tagged sequence: flies/NOUN, like/VERB, a/ARTICLE, flower/NOUN

P(NOUN, VERB, ARTICLE, NOUN | flies, like, a, flower) = $V_4^{(2)} = 3.6099E\text{-}5$

Transition probability matrix

	<s>	ARTICLE	NOUN	VERB	PREPOSITION
<s>	0.00	0.71	0.29	0.00	0.00
ARTICLE	0.00	0.00	1.00	0.00	0.00
NOUN	0.00	0.00	0.13	0.43	0.44
VERB	0.00	0.65	0.35	0.00	0.00
PREPOSITION	0.00	0.74	0.26	0.00	0.00

Hidden Markov Model

Emission probability matrix

	flies	like	a	flower
<s>	0.000	0.000	0.000	0.000
ARTICLE	0.000	0.000	0.360	0.000
NOUN	0.025	0.012	0.001	0.063
VERB	0.076	0.100	0.000	0.050
PREPOSITION	0.000	0.068	0.000	0.000