

2.2 词性标注及隐马尔科夫模型

陈 露

上海交通大学计算机学院

2025年秋季学期

- ▶ 什么是词性标注
- ▶ 隐马尔科夫模型 (HMM)
- ▶ HMM的几个重要算法
 - ▶ 给定文本推断词性：Viterbi算法
 - ▶ 计算给定文本出现的概率：前向算法
 - ▶ 计算给定文本出现的概率：后向算法
 - ▶ 有监督学习：最大似然参数估计
 - ▶ 无监督学习：前向-后向算法
- ▶ 词性标注的实战经验及常用工具推荐

什么是词性标注？

词性标注就是在给定句子中判定每个词的语法范畴，确定其词性并加以标注的过程

什么是词性标注？

英语



名词

动词

形容词

数词

代词

副词

介词

连词

冠词

感叹词

汉语



名词

动词

形容词

数词

量词

代词

副词

介词

连词

助词

叹词

拟声词

词性（词类）
Part-Of-Speech
(POS)



指词的语法分类，就是各位中学里学到的词类

词性标注
POS Tagging



就是用算法自动将句子中每个词的词性判断出来的过程

什么是词性标注？

▶ 词性与上下文相关

- ▶ 汉语是一种缺乏词形态变化的语言，词的类别不能像印欧语那样，直接从词的形态变化上来判别。
- ▶ 常用词兼类现象严重。《现代汉语八百词》收取的常用词中，兼类词所占的比例高达22.5%，而且发现越是常用的词，不同的用法越多。由于兼类使用程度高，兼类现象涉及汉语中大部分词类，因而造成在汉语文本中词类歧义排除的任务量大。
- ▶ 研究者主观原因造成的困难。语言学界在词性划分的目的、标准等问题上还存在分歧。目前还没有一个统一的被广泛认可汉语词类划分标准，词类划分的粒度和标记符号都不统一。词类划分标准和标记符号集的差异，以及分词规范的含混性，给中文信息处理带来了极大的困难。

词性标注

词性划分具有层次性

Tag	Description	Example	Tag	Description	Example
CC	Coordin. Conjunction	and, but, or	SYM	Symbol	+,%,&
CD	Cardinal number	one, two, three	TO	“to”	to
DT	Determiner	a, the	UH	Interjection	ah, oops
EX	Existential ‘there’	there	VB	Verb, base form	eat
FW	Foreign word	mea culpa	VBD	Verb, past tense	ate
IN	Preposition/sub-conj	of, in, by	VBG	Verb, gerund	eating
JJ	Adjective	yellow	VBN	Verb, past participle	eaten
JJR	Adj., comparative	bigger	VBP	Verb, non-3sg pres	eat
JJS	Adj., superlative	wildest	VBZ	Verb, 3sg pres	eats
LS	List item marker	1, 2, One	WDT	Wh-determiner	which, that
MD	Modal	can, should	WP	Wh-pronoun	what, who
NN	Noun, sing. or mass	llama	WPS	Possessive wh-	whose
NNS	Noun, plural	llamas	WRB	Wh-adverb	how, where
NNP	Proper noun, singular	IBM	\$	Dollar sign	\$
NNPS	Proper noun, plural	Carolinas	#	Pound sign	#
PDT	Predeterminer	all, both	“	Left quote	‘ or “
POS	Possessive ending	’s	”	Right quote	’ or ”
PRP	Personal pronoun	I, you, he	(Left parenthesis	[, (, {, <
PRP\$	Possessive pronoun	your, one’s)	Right parenthesis],), }, >
RB	Adverb	quickly, never	,	Comma	,
RBR	Adverb, comparative	faster	.	Sentence-final punc	. ! ?
RBS	Adverb, superlative	fastest	:	Mid-sentence punc	: ; ... - -
RP	Particle	up, off			

Figure 5.6 Penn Treebank part-of-speech tags (including punctuation).

Penn Treebank (英语) : 45个词性

代码	名称	代码	名称	代码	名称	代码	名称
Ag	形语素	g	语素	ns	地名	u	助词
a	形容词	h	前接成分	nt	机构团体	Vg	动语素
ad	副形词	i	成语	nz	其他专名	v	动词
an	名形词	j	简称略语	o	拟声词	vd	副动词
b	区别词	k	后接成分	p	介词	vn	名动词
c	连词	l	习用语	q	量词	w	标点符号
Dg	副语素	m	数词	r	代词	x	非语素字
d	副词	Ng	名语素	s	处所词	y	语气词
e	叹词	n	名词	Tg	时语素	z	状态词
f	方位词	nr	人名	t	时间词		

北大词性标记集 (汉语) : 39个词性

Tag	Description	Tag	Description
(opening parenthesis	RP	adverb or particle
)	closing parenthesis	TO	infinitive marker
*	negator	UH	interjection, exclamation
,	comma	VB	verb, base form
-	dash	VBD	verb, past tense
.	sentence terminator	VBG	verb, present participle, gerund
:	colon	VBN	verb, past participle
ABL	pre-qualifier	VBZ	verb, 3rd singular present
ABN	pre-quantifier	WDT	wh- determiner
ABX	pre-quantifier, double conjunction	WPS	possessive wh- pronoun
AP	post-determiner	WPO	objective wh- pronoun
AT	article	WPS	nominative wh- pronoun
BE/BED/BEDZ/BEG/BEM/BEN/BER/BEZ		WQL	how
CC	coordinating conjunction	WRB	wh- adverb
CD	cardinal numeral		
CS	subordinating conjunction	NPS	plural proper noun
DO/DOD/DOZ		NPSS	possessive plural proper noun
DT	singular determiner,	NR	adverbial noun
DTI	singular or plural determiner	NRS	possessive adverbial noun
DTS	plural determiner	OD	plural adverbial noun
DTX	determiner, double conjunction	PN	ordinal numeral
EX	existential there	PNS	nominal pronoun
HV/HVD/HVG/HVN/HVZ		PPS	possessive nominal pronoun
IN	preposition	PPSS	possessive personal pronoun
JJ	adjective	PPL	second possessive personal pronoun
JJR	comparative adjective	PPLS	singular reflexive personal pronoun
JJS	semantically superlative adj.	PPO	plural reflexive pronoun
JTT	morphologically superlative adj.	PPS	objective personal pronoun
MD	modal auxiliary	PPSS	3rd. sg. nominative pronoun
NN	(common) singular or mass noun	QL	other nominative pronoun
NN\$	possessive singular common noun	QLP	qualifier
NN%	plural common noun	RB	post-qualifier
NN\$%	possessive plural noun	RBR	adverb
NP	singular proper noun	RBT	comparative adverb
NP\$	possessive singular proper noun	RN	superlative adverb
			nominal adverb

Brown Corpus (英语) : 87个词性

如何建立模型来正确预测他们?

对该问题建模的流派

▶ 基于规则的词性标注方法

- ▶ 基于规则的词性标注方法是人们提出较早的一种词性标注方法，其基本思想是按兼类词搭配关系和上下文语境建造词类消歧规则。早期的词类标注规则一般由人工构建。
- ▶ 随着标注语料库规模的增大，可利用的资源也变得越来越多，这时候以人工提取规则的方法显然变得不现实，于是乎，人们提出了基于机器学习的规则自动提取方法。

▶ 基于统计模型的词性标注方法

- ▶ 统计方法将词性标注看作是一个序列标注问题。其基本思想是：给定带有各自标注的词的序列，我们可以确定下一个词最可能的词性。
- ▶ 现在已经有隐马尔可夫模型（HMM）或条件随机场（CRF）等统计模型了，这些模型可以使用有标记数据的大型语料库进行训练，而有标记的数据则是指其中每一个词都分配了正确的词性标注的文本。

▶ 基于统计方法与规则方法相结合的词性标注方法

- ▶ 理性主义方法与经验主义相结合的处理策略一直是自然语言处理领域的专家们不断研究和探索的问题，对于词性标注问题当然也不例外。
- ▶ 这类方法的主要特点在于对统计标注结果的筛选，只对那些被认为可疑的标注结果，才采用规则方法进行歧义消解，而不是对所有情况都既使用统计方法又使用规则方法。

▶ 基于深度学习的词性标注方法

- ▶ 可以当作序列标注的任务来做，目前深度学习解决序列标注任务常用方法包括LSTM+CRF、BiLSTM+CRF等。

对该问题建模的流派

▶ 基于规则的词性标注方法

- ▶ 基于规则的词性标注方法是人们提出较早的一种词性标注方法，其基本思想是按兼类词搭配关系和上下文语境建造词类消歧规则。早期的词类标注规则一般由人工构建。
- ▶ 随着标注语料库规模的增大，可利用的资源也变得越来越多，这时候以人工提取规则的方法显然变得不现实，于是乎，人们提出了基于机器学习的规则自动提出方法。

▶ 基于统计模型的词性标注方法

- ▶ 统计方法将词性标注看作是一个序列标注问题。其基本思想是：给定带有各自标注的词的序列，我们可以确定下一个词最可能的词性。
- ▶ 现在已经有隐马尔可夫模型（HMM）或条件随机场（CRF）等统计模型了，这些模型可以使用有标记数据的大型语料库进行训练，而有标记的数据则是指其中每一个词都分配了正确的词性标注的文本。

▶ 基于统计方法与规则方法相结合的词性标注方法

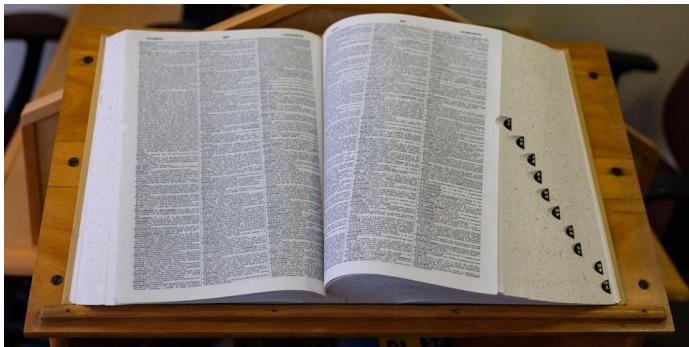
- ▶ 理性主义方法与经验主义相结合的处理策略一直是自然语言处理领域的专家们不断研究和探索的问题，对于词性标注问题当然也不例外。
- ▶ 这类方法的主要特点在于对统计标注结果的筛选，只对那些被认为可疑的标注结果，才采用规则方法进行歧义消解，而不是对所有情况都既使用统计方法又使用规则方法。

▶ 基于深度学习的词性标注方法

- ▶ 可以当作序列标注的任务来做，目前深度学习解决序列标注任务常用方法包括LSTM+CRF、BiLSTM+CRF等。

- ▶ 什么是词性标注
- ▶ 隐马尔科夫模型
- ▶ 三个场景

词性标注：从规则到HMM



Tag	Description	Example	Tag	Description	Example
CC	Coordin. Conjunction	<i>and, but, or</i>	SYM	Symbol	+,%,&
CD	Cardinal number	<i>one, two, three</i>	TO	“to”	<i>to</i>
DT	Determiner	<i>a, the</i>	UH	Interjection	<i>ah, oops</i>
EX	Existential ‘there’	<i>there</i>	VB	Verb, base form	<i>eat</i>
FW	Foreign word	<i>mea culpa</i>	VBD	Verb, past tense	<i>ate</i>
IN	Preposition/sub-conj	<i>of, in, by</i>	VBG	Verb, gerund	<i>eating</i>
JJ	Adjective	<i>yellow</i>	VBN	Verb, past participle	<i>eaten</i>
JJR	Adj., comparative	<i>bigger</i>	VBP	Verb, non-3sg pres	<i>eat</i>
JJS	Adj., superlative	<i>wildest</i>	VBZ	Verb, 3sg pres	<i>eats</i>
LS	List item marker	<i>1, 2, One</i>	WDT	Wh-determiner	<i>which, that</i>
MD	Modal	<i>can, should</i>	WP	Wh-pronoun	<i>what, who</i>
NN	Noun, sing. or mass	<i>llama</i>	WP\$	Possessive wh-	<i>whose</i>
NNS	Noun, plural	<i>llamas</i>	WRB	Wh-adverb	<i>how, where</i>
NNP	Proper noun, singular	<i>IBM</i>	\$	Dollar sign	\$
NNPS	Proper noun, plural	<i>Carolinas</i>	#	Pound sign	#
PDT	Predeterminer	<i>all, both</i>	“	Left quote	‘ or “
POS	Possessive ending	<i>'s</i>	”	Right quote	’ or ”
PRP	Personal pronoun	<i>I, you, he</i>	(Left parenthesis	[, (, {, <
PRPS	Possessive pronoun	<i>your, one's</i>)	Right parenthesis],), }, >
RB	Adverb	<i>quickly, never</i>	,	Comma	,
RBR	Adverb, comparative	<i>faster</i>	.	Sentence-final punc	. ! ?
RBS	Adverb, superlative	<i>fastest</i>	:	Mid-sentence punc	: ; ... --
RP	Particle	<i>up, off</i>			

Figure 5.6 Penn Treebank part-of-speech tags (including punctuation).

Steve Jobs , 42 years old , will join Apple ...
NNP NNPS , CD NNS JJ , MD VBP NN
NNP LS NNP VB NNP
NNS NN

词性标注：从规则到HMM

规则：

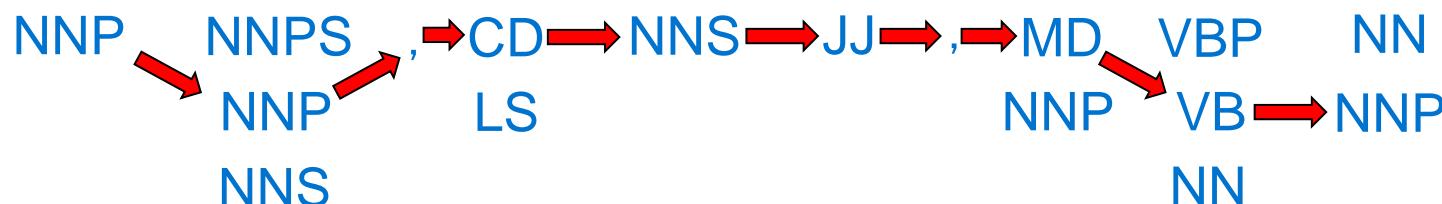
一阶 关联	人名	NNP -> NNP
	将来时	MD -> VB (will be)
	祈使	MD -> PRP (could you)
	动名词	VB -> NN may

二阶 关联	年龄短语	CD -> NNS/NN -> JJ

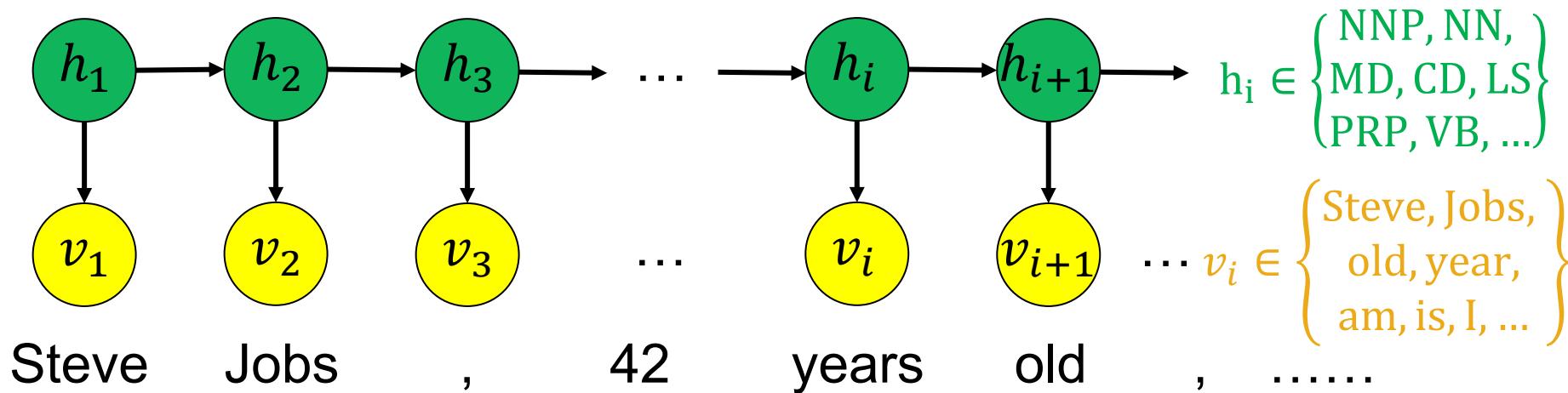
...

...

Steve Jobs , 42 years old , will join Apple ...



隐马尔科夫模型 (HMM)



已知句子单词，预测每个单词的词性，变成了这样一个任务：

$$\operatorname{argmax}_{h_1, h_2, \dots, h_i, \dots, h_N} P(h_1, h_2, \dots, h_i, \dots, h_N \mid v_1, v_2, \dots, v_i, \dots, v_N)$$

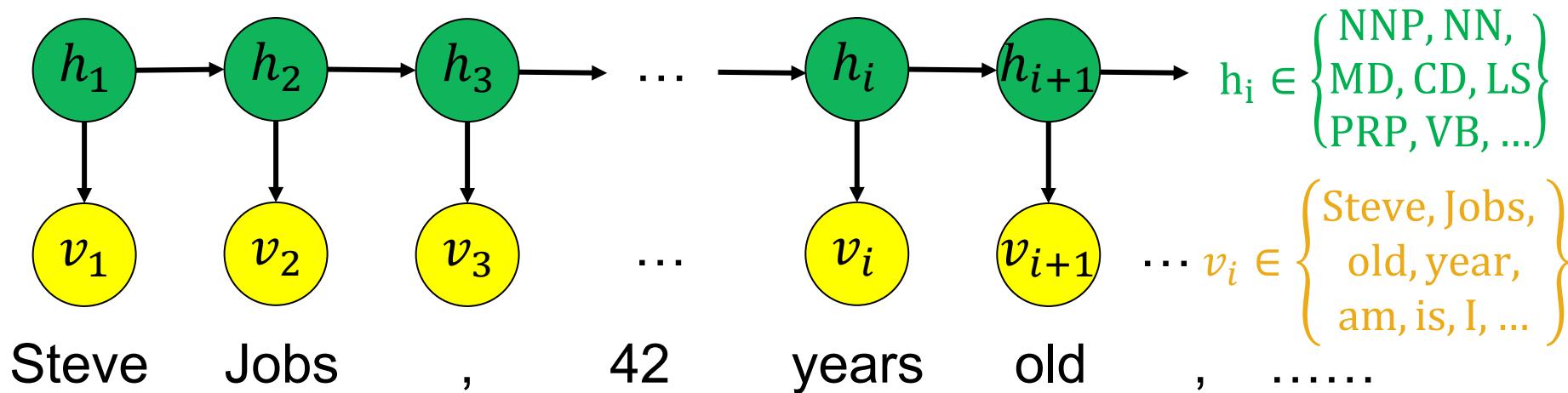
$$= \operatorname{argmax}_h P(h \mid v)$$

$$= \operatorname{argmax}_h \frac{P(v|h)P(h)}{P(v)}$$

$$= \operatorname{argmax}_h \boxed{P(v \mid h) \cdot P(h)}$$

$P(v, h)$

隐马尔科夫模型



已知句子单词，预测每个单词的词性，变成了这样一个任务：

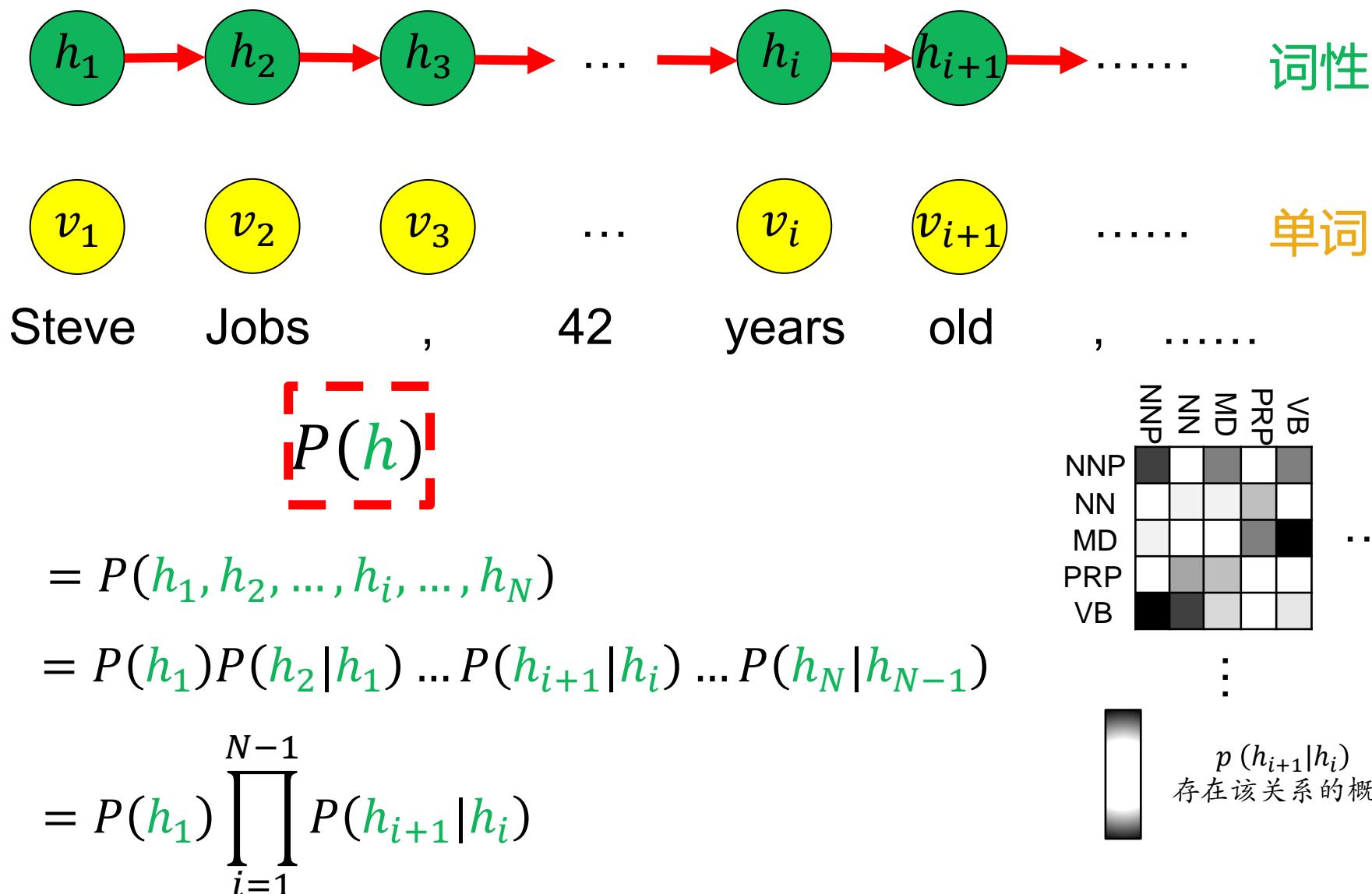
$$\operatorname{argmax}_{h_1, h_2, \dots, h_i, \dots, h_N} P(h_1, h_2, \dots, h_i, \dots, h_N \mid v_1, v_2, \dots, v_i, \dots, v_N)$$

$$= \operatorname{argmax}_h P(h \mid v)$$

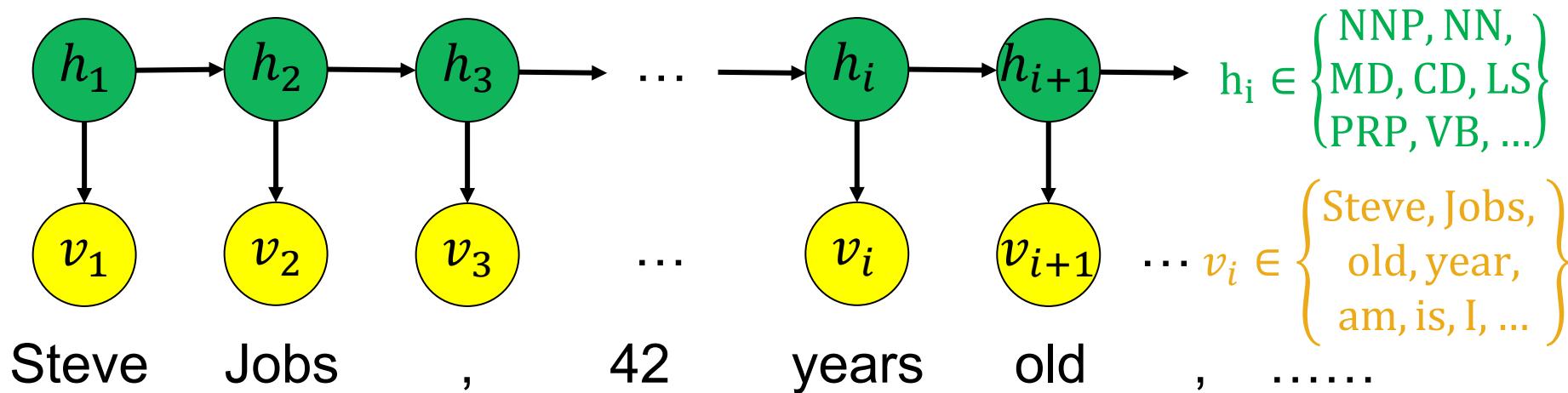
$$= \operatorname{argmax}_h \frac{P(v|h)P(h)}{P(v)}$$

$$= \operatorname{argmax}_h \boxed{\frac{P(v|h)}{P(v)}} \cdot \boxed{\frac{P(h)}{P(v)}}$$

隐马尔科夫模型



隐马尔科夫模型



已知句子单词，预测每个单词的词性，变成了这样一个任务：

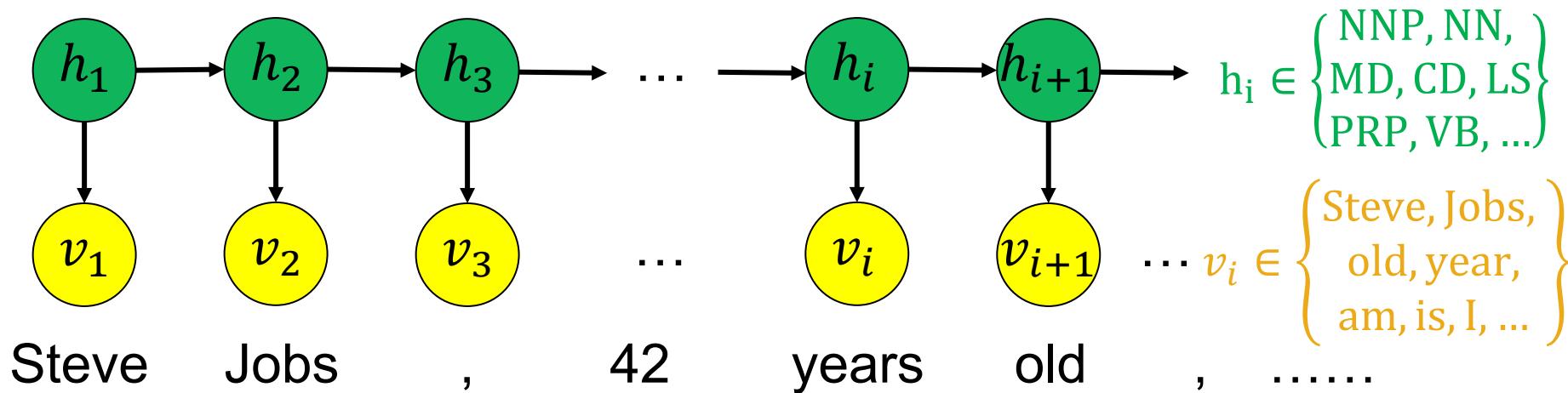
$$\operatorname{argmax}_{h_1, h_2, \dots, h_i, \dots, h_N} P(h_1, h_2, \dots, h_i, \dots, h_N \mid v_1, v_2, \dots, v_i, \dots, v_N)$$

$$= \operatorname{argmax}_h P(h \mid v)$$

$$= \operatorname{argmax}_h \frac{P(v|h)P(h)}{P(v)}$$

$$= \operatorname{argmax}_h \boxed{\frac{P(v|h)}{P(v)}} \cdot \boxed{\frac{P(h)}{P(v)}}$$

隐马尔科夫模型



已知句子单词，预测每个单词的词性，变成了这样一个任务：

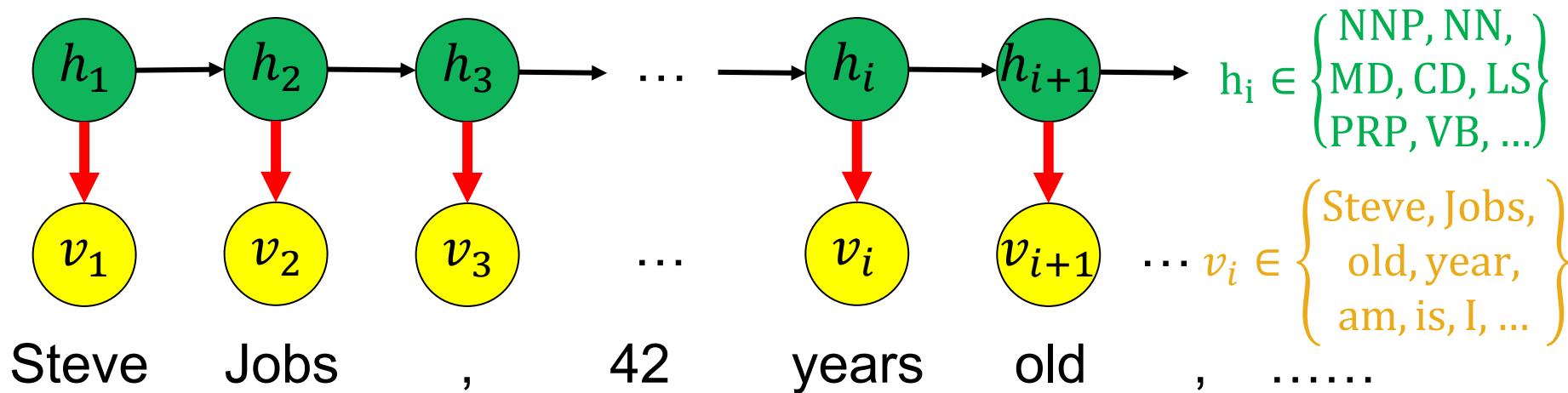
$$\operatorname{argmax}_{h_1, h_2, \dots, h_i, \dots, h_N} P(h_1, h_2, \dots, h_i, \dots, h_N \mid v_1, v_2, \dots, v_i, \dots, v_N)$$

$$= \operatorname{argmax}_h P(h \mid v)$$

$$= \operatorname{argmax}_h \frac{P(v|h)P(h)}{P(v)}$$

$$= \operatorname{argmax}_h \boxed{\frac{P(v|h)}{P(v)}} \cdot \boxed{\frac{P(h)}{P(v)}}$$

隐马尔科夫模型



已知句子单词，预测每个单词的词性，变成了这样一个任务：

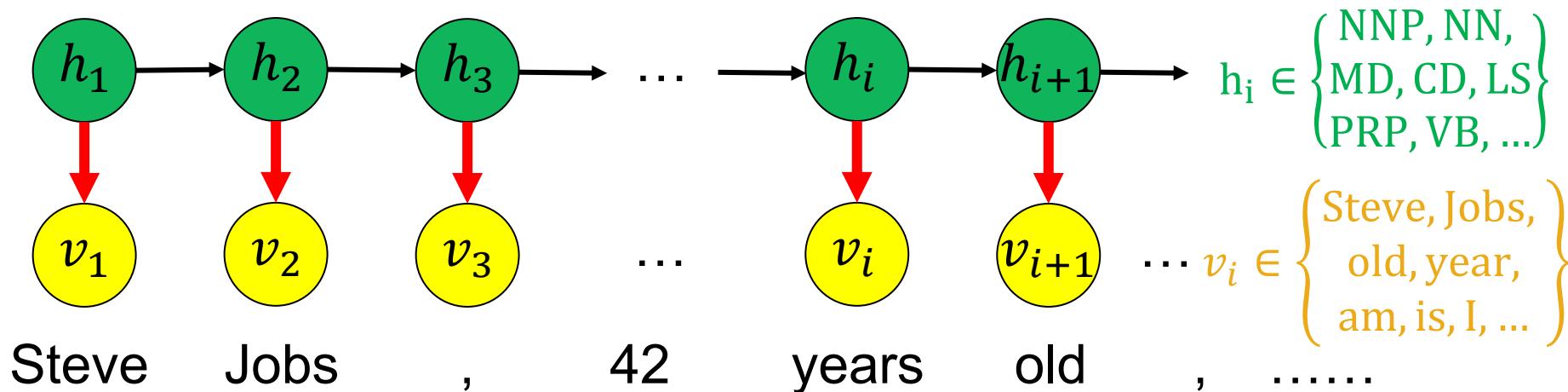
$$\operatorname{argmax}_{h_1, h_2, \dots, h_i, \dots, h_N} P(h_1, h_2, \dots, h_i, \dots, h_N \mid v_1, v_2, \dots, v_i, \dots, v_N)$$

$$= \operatorname{argmax}_h P(h \mid v)$$

$$= \operatorname{argmax}_h \frac{P(v|h)P(h)}{P(v)}$$

$$= \operatorname{argmax}_h \boxed{\frac{P(v|h)}{P(v)}} \cdot P(h)$$

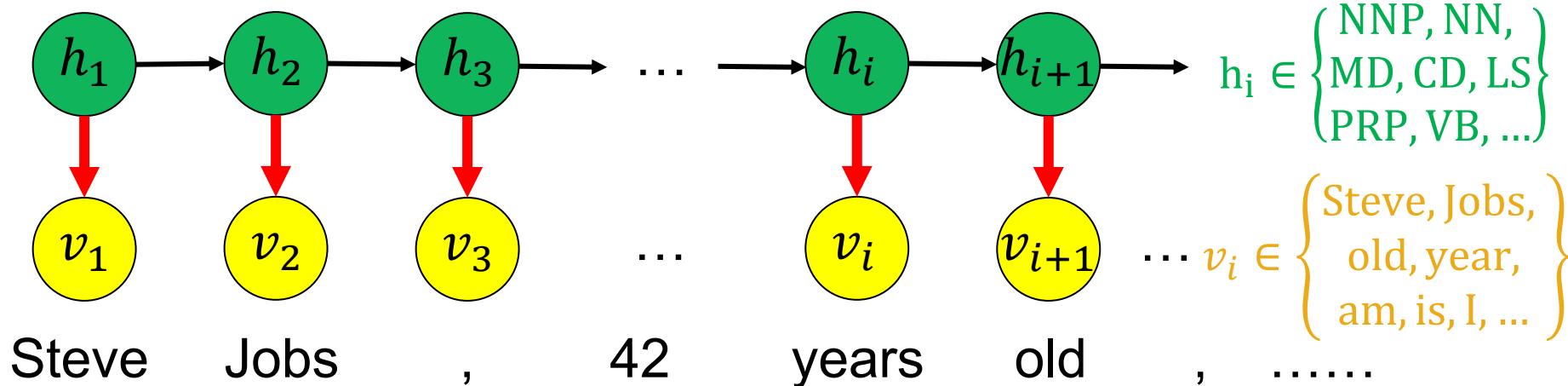
隐马尔科夫模型



am	is	are	a	an	the	old	help	sugar	price	help	sugar
NNP											
NN											
MD											
PRP											
VB											

$$P(v|h)$$

隐马尔科夫模型

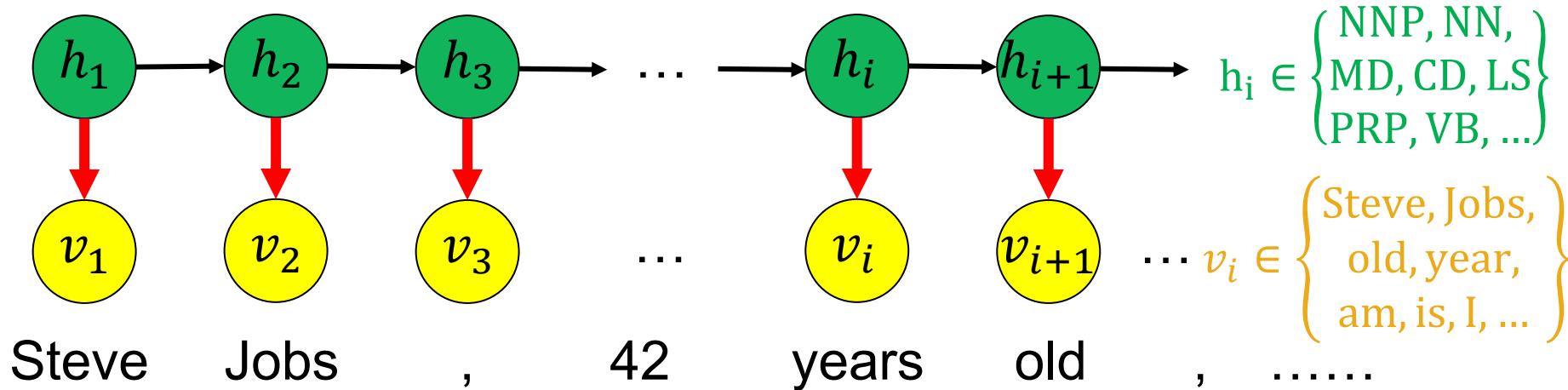


$$\begin{aligned}
 & P(v|h) \\
 & = P(v_1, v_2, \dots, v_N | h_1, h_2, \dots, h_N) \\
 & = P(v_1|h_1)P(v_2|h_2) \dots P(v_i|h_i) \dots P(v_N|h_N) \\
 & = \prod_{j=1}^N P(v_j|h_j)
 \end{aligned}$$

am	is	are	a	an	the	old	help	sugar
NNP								
NN								
MD								
PRP								
VB								

⋮

隐马尔科夫模型



$$\operatorname{argmax}_{\mathbf{h}} P(\mathbf{h} | \mathbf{v}) = \operatorname{argmax}_{\mathbf{h}} P(\mathbf{h}, \mathbf{v}) = \operatorname{argmax}_{\mathbf{h}} P(\mathbf{v} | \mathbf{h}) \cdot P(\mathbf{h})$$

$$= \operatorname{argmax}_{\mathbf{h}} \left(\prod_{j=1}^N P(v_j | h_j) \right) P(h_1) \prod_{i=1}^{N-1} P(h_{i+1} | h_i)$$

NNP							
NN							
MD							
PRP							
VB							

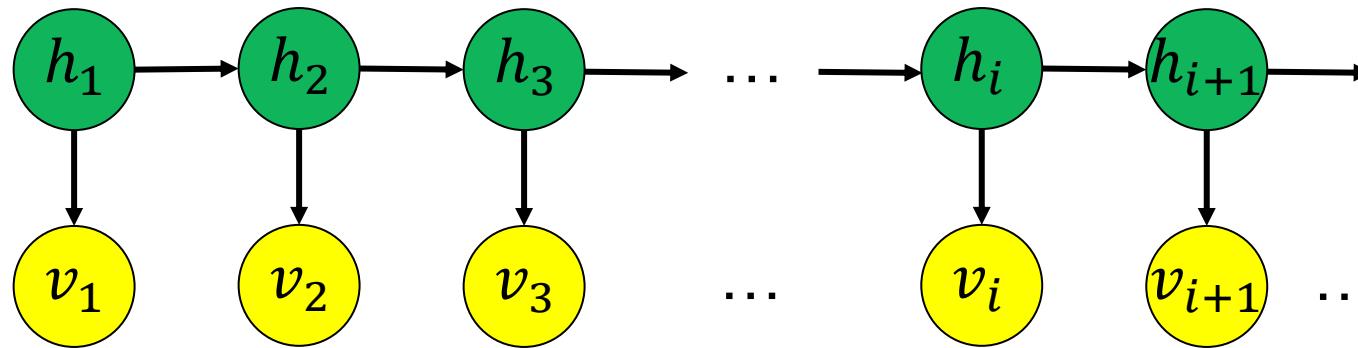
NNP							
NN							
MD							
PRP							
VB							

参数：

NNP							
NN							
MD							
PRP							
VB							

NNP							
NN							
MD							
PRP							
VB							

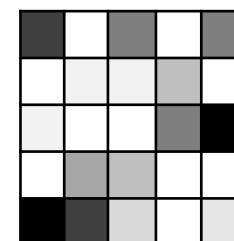
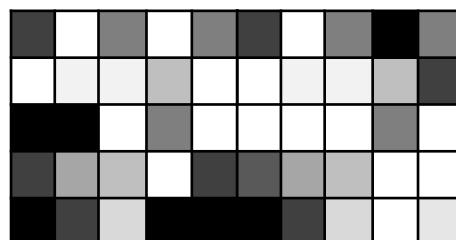
隐马尔科夫模型



隐马尔科夫模型的定义

$$\begin{aligned} P(\mathbf{h}, \mathbf{v}) &= P(\mathbf{v}|\mathbf{h}) \cdot P(\mathbf{h}) \\ &= \prod_{j=1}^N P(v_j|h_j) P(h_1) \prod_{i=1}^{N-1} P(h_{i+1}|h_i) \end{aligned}$$

依照左式对联合概率分布 $P(\mathbf{v}, \mathbf{h})$ 建模，即被称为 HMM 模型



\mathbf{v}, \mathbf{h} 都应是离散随机变量的离散序列

参数：

例子：计算联合概率

已知：

	jobs	years	steve	old	42	NNP	CD	SNS	JJ	
NNP	0.05	0.01	0.1	0.01	0.04	0.01	0.2	0.01	0.01	0.05
,	0.01	0.01	0.01	0.7	0.01	0.01	0.2	0.1		0.3
CD	0.01	0.01	0.01	0.01	0.01	0.9	0.15	0.1	0.4	0.2
NNS	0.2	0.1	0.01	0.01	0.01	0.01	0.2	0.01	0.01	0.2
JJ	0.01	0.01	0.01	0.01	0.4	0.01	0.1	0.1	0.4	0.3

求以下联合概率分布：

给定： Steve Jobs , 42 years old

对应POS为： NNP NNP , CD NNS JJ

$$\begin{aligned}
 P(h, v) &= P(\text{NNP}, \text{NNP}, \text{,}, \text{CD}, \text{NNS}, \text{JJ}, \text{Steve}, \text{Jobs}, \text{,}, \text{,}, 42, \text{years}, \text{old}) \\
 &= P(\text{NNP})P(\text{NNP}|\text{NNP})P(\text{,}|\text{NNP})P(\text{CD}|\text{,})P(\text{NNS}|\text{CD})P(\text{JJ}|\text{NNS}) \\
 &\quad P(\text{Steve}|\text{NNP})P(\text{Jobs}|\text{NNP})P(\text{,}|\text{,})P(42|\text{CD})P(\text{years}|\text{NNS})P(\text{old}|\text{JJ}) \\
 &= 0.05 \times 0.2 \times 0.2 \times 0.3 \times 0.4 \times 0.1 \\
 &\quad \times 0.1 \times 0.05 \times 0.7 \times 0.9 \times 0.1 \times 0.4 \\
 &= 0.00000003024
 \end{aligned}$$

例子：计算联合概率

已知：

	jobs	years	steve	old	42	NNP	CD	NNS	JJ	
NNP	0.05	0.01	0.1	0.01	0.04	0.01	0.2	0.01	0.01	0.05
,	0.01	0.01	0.01	0.7	0.01	0.01	0.2	0.1		0.3
CD	0.01	0.01	0.01	0.01	0.01	0.9	0.15	0.1	0.4	0.2
NNS	0.2	0.1	0.01	0.01	0.01	0.01	0.2	0.01	0.01	0.2
JJ	0.01	0.01	0.01	0.01	0.4	0.01	0.1	0.1	0.4	0.3

求以下联合概率分布：

给定： Steve Jobs , 42 years old

对应POS为： NNP NNS , CD NNS JJ

$$\begin{aligned}
 P(h, v) &= P(\text{NNP}, \text{NNS}, \text{,}, \text{CD}, \text{NNS}, \text{JJ}, \text{Steve}, \text{Jobs}, \text{,}, \text{,}, \text{42}, \text{years}, \text{old}) \\
 &= P(\text{NNP})P(\text{NNS}|\text{NNP})P(\text{,}|\text{NNS})P(\text{CD}|\text{,})P(\text{NNS}|\text{CD})P(\text{JJ}|\text{NNS}) \\
 &\quad P(\text{Steve}|\text{NNP})P(\text{Jobs}|\text{NNS})P(\text{,}|\text{,})P(\text{42}|\text{CD})P(\text{years}|\text{NNS})P(\text{old}|\text{JJ}) \\
 &= 0.05 \times 0.01 \times 0.2 \times 0.3 \times 0.4 \times 0.1 \\
 &\quad \times 0.1 \times 0.2 \times 0.7 \times 0.9 \times 0.1 \times 0.4 \\
 &= 0.0000000006048
 \end{aligned}$$

隐马尔科夫模型

已知：

	jobs	years	steve	,	old	42	NNP	CD	SNN	JJ	
NNP	0.05	0.01	0.1	0.01	0.04	0.01	0.2	0.2	0.01	0.01	0.05
,	0.01	0.01	0.01	0.7	0.01	0.01	0.2	0	0.3	0.1	0.3
CD	0.01	0.01	0.01	0.01	0.01	0.9	0.15	0.01	0.1	0.4	0.01
NNS	0.2	0.1	0.01	0.01	0.01	0.01	0.2	0.1	0.01	0.01	0.2
JJ	0.01	0.01	0.01	0.01	0.4	0.01	0.1	0.1	0.1	0.2	0.3

求以下联合概率分布：

给定： Steve Jobs , 42 years old

对应POS为： NNP NNP , CD NNS JJ



$$\log P(h_1, v) = -19.62$$

$$P(h_1, v) = 0.000000003024$$

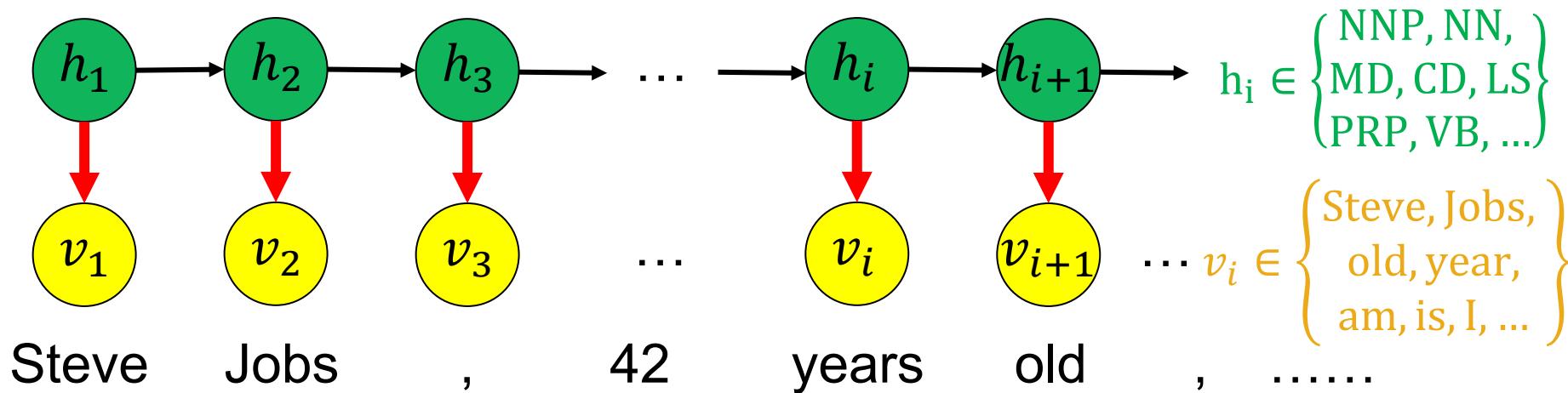
对应POS为： NNP NNS , CD NNS JJ



$$\log P(h_2, v) = -21.22$$

$$P(h_2, v) = 0.0000000006048$$

隐马尔科夫模型

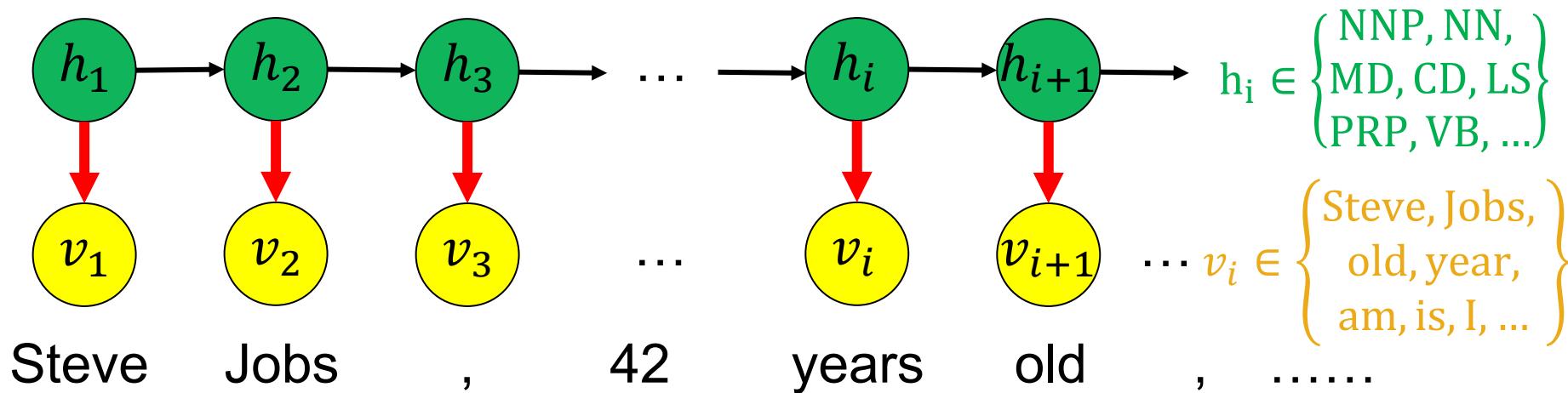


$$P(h_1, v)$$

$$P(h_2, v)$$

以上例子可知，给定一个句子及其POS序列我们可以算出它的联合分布概率

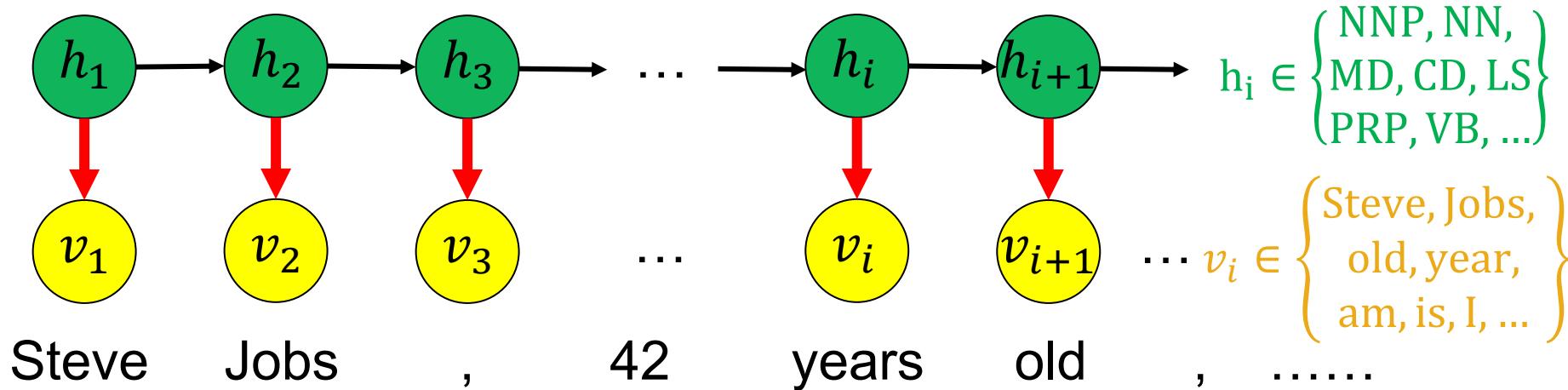
隐马尔科夫模型



$$\overline{\arg\max}_{\overline{h}} P(\overline{h} | \overline{v}) = \arg\max_h P(h, v)$$

以上例子可知，给定一个句子及其POS序列我们可以算出它的联合分布概率但是我们是要求 $\arg\max_{\overline{h}}(\cdot)$ ，总不能对所有可能的序列都去算一遍。所以我们怎么给定句子单词，找到最优的 h 序列呢？

隐马尔科夫模型：Viterbi 算法

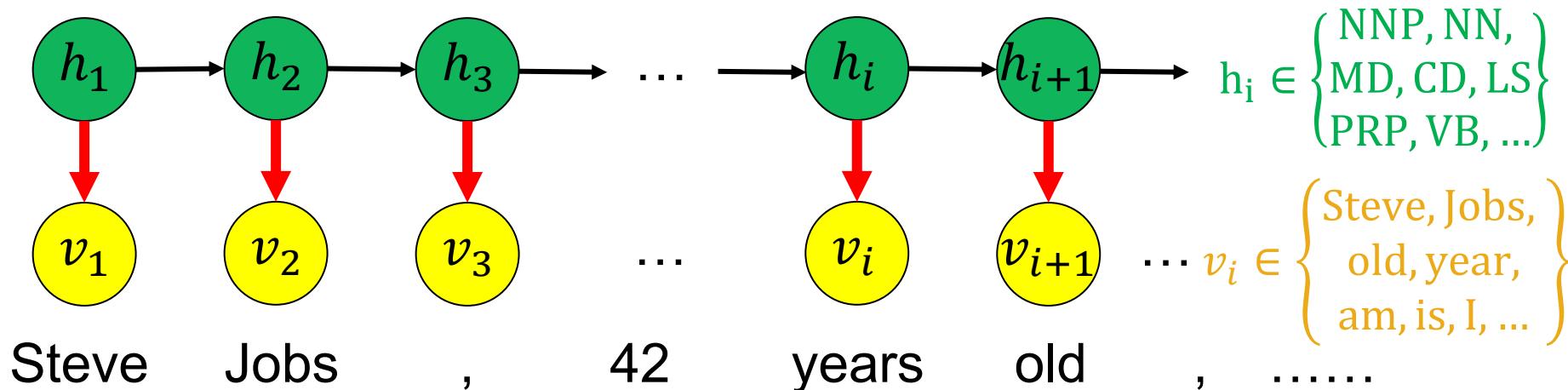


$$\underset{h}{\operatorname{argmax}} P(h | v) = \underset{h}{\operatorname{argmax}} P(h, v)$$

$$= \underset{h}{\operatorname{argmax}} (P(v|h) \cdot P(h))$$

$$= \underset{h}{\operatorname{argmax}} \left(\prod_{j=1}^N P(v_j | h_j) P(h_1) \prod_{i=1}^{N-1} P(h_{i+1} | h_i) \right)$$

隐马尔科夫模型：Viterbi算法



$$= \operatorname{argmax}_{\mathbf{h}} \left(\prod_{j=1}^N P(v_j | h_j) P(h_1) \prod_{i=1}^{N-1} P(h_{i+1} | h_i) \right)$$

$$= \operatorname{argmax}_{h_1, h_2, \dots, h_i, \dots, h_N} \left(\frac{P(v_1 | h_1) P(v_2 | h_2) \dots P(v_i | h_i) \dots P(v_N | h_N)}{P(h_1) P(h_2 | h_1) \dots P(h_{i+1} | h_i) \dots P(h_{N-1} | h_N)} \right)$$



隐马尔科夫模型：Viterbi算法

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	JJ	
NNP	0.05	0.01	0.1	0.01	0.04	0.01	0.2	0.01	0.01	0.05
,	0.01	0.01	0.01	0.7	0.01	0.01	0.2	0.1		0.3
CD	0.01	0.01	0.01	0.01	0.01	0.9	0.15	0.1	0.4	0.01
NNS	0.2	0.1	0.01	0.01	0.01	0.01	0.2	0.01	0.01	0.2
JJ	0.01	0.01	0.01	0.01	0.4	0.01	0.1	0.1	0.4	0.3

求以下句子的最佳POS标注：

给定： Steve Jobs , 42 years old

$$\begin{aligned}
 \text{NNP} & P(\text{NNP})P(\text{Steve}|\text{NNP}) \\
 , & P(,)P(\text{Steve},|\text{,}) \\
 \text{CD} & P(\text{CD})P(\text{Steve}|\text{CD}) \\
 \text{NNS} & P(\text{NNS})P(\text{Steve}|\text{NNS}) \\
 \text{JJ} & P(\text{JJ})P(\text{Steve}|\text{JJ})
 \end{aligned}$$

隐马尔科夫模型：Viterbi算法

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	JJ	
NNP	0.05	0.01	0.1	0.01	0.04	0.01	0.2	0.01	0.01	0.05
,	0.01	0.01	0.01	0.7	0.01	0.01	0.2	0.1		0.3
CD	0.01	0.01	0.01	0.01	0.01	0.9	0.15	0.1	0.4	0.01
NNS	0.2	0.1	0.01	0.01	0.01	0.01	0.2	0.01	0.01	0.2
JJ	0.01	0.01	0.01	0.01	0.4	0.01	0.1	0.1	0.2	0.3

求以下句子的最佳POS标注：

给定： Steve Jobs , 42 years old

NNP	0.005
,	0.003
CD	0.0001
NNS	0.002
JJ	0.003

隐马尔科夫模型：Viterbi算法

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	JJ	
NNP	0.05	0.01	0.1	0.01	0.04	0.01	0.2	0.01	0.01	0.05
,	0.01	0.01	0.01	0.7	0.01	0.01	0.2	0.1		0.3
CD	0.01	0.01	0.01	0.01	0.01	0.9	0.15	0.1	0.4	0.01
NNS	0.2	0.1	0.01	0.01	0.01	0.01	0.2	0.01	0.01	0.2
JJ	0.01	0.01	0.01	0.01	0.4	0.01	0.1	0.1	0.2	0.3

求以下句子的最佳POS标注：

给定： Steve Jobs , 42 years old

NNP 0.005 
, 0.003 
CD 0.0001 
NNS 0.002 
JJ 0.003 

$\max_{left} (left \cdot P(\text{NNP}|left_tag)P(\text{Jobs}|\text{NNP}))$

隐马尔科夫模型：Viterbi算法

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	?	
NNP	0.05	0.01	0.1	0.01	0.04	0.01	0.2	0.01	0.01	0.05
,	0.01	0.01	0.01	0.7	0.01	0.01	0.2	0.1		0.3
CD	0.01	0.01	0.01	0.01	0.01	0.9	0.15	0.1	0.4	0.01
NNS	0.2	0.1	0.01	0.01	0.01	0.01	0.2	0.01	0.01	0.2
JJ	0.01	0.01	0.01	0.01	0.4	0.01	0.1	0.1	0.2	0.3

求以下句子的最佳POS标注：

给定： Steve Jobs , 42 years old

NNP	0.005	→ 5e-4	
,	0.003		$\max_{left} (left \cdot P(, left_tag) P(Jobs ,))$
CD	0.0001		
NNS	0.002		
JJ	0.003		

隐马尔科夫模型：Viterbi算法

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	JJ	
NNP	0.05	0.01	0.1	0.01	0.04	0.01	0.2	0.01	0.01	0.05
,	0.01	0.01	0.01	0.7	0.01	0.01	0.2	0.1		0.3
CD	0.01	0.01	0.01	0.01	0.01	0.9	0.15	0.1	0.4	0.01
NNS	0.2	0.1	0.01	0.01	0.01	0.01	0.2	0.01	0.01	0.2
JJ	0.01	0.01	0.01	0.01	0.4	0.01	0.1	0.1	0.2	0.3

求以下句子的最佳POS标注：

给定： Steve Jobs , 42 years old

NNP	0.005	 5e-4
,	0.003	 1e-5
CD	0.0001	
NNS	0.002	
JJ	0.003	

隐马尔科夫模型：Viterbi算法

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	?	
NNP	0.05	0.01	0.1	0.01	0.04	0.01	0.2	0.01	0.01	0.05
,	0.01	0.01	0.01	0.7	0.01	0.01	0.2	0.1		0.3
CD	0.01	0.01	0.01	0.01	0.01	0.9	0.15	0.1	0.4	0.01
NNS	0.2	0.1	0.01	0.01	0.01	0.01	0.2	0.01	0.01	0.2
JJ	0.01	0.01	0.01	0.01	0.4	0.01	0.1	0.1	0.2	0.3

求以下句子的最佳POS标注：

给定： Steve Jobs , 42 years old

NNP	0.005	$\xrightarrow{\text{left}}$ 5e-4
,	0.003	$\xrightarrow{\text{left}}$ 1e-5
CD	0.0001	$\xrightarrow{\text{left}}$ $\max_{\text{left}}(\text{left} \cdot P(\text{CD} \text{left_tag})P(\text{Jobs} \text{CD}))$
NNS	0.002	
JJ	0.003	

隐马尔科夫模型：Viterbi算法

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	JJ	
NNP	0.05	0.01	0.1	0.01	0.04	0.01	0.2	0.01	0.01	0.05
,	0.01	0.01	0.01	0.7	0.01	0.01	0.2	0.1		0.3
CD	0.01	0.01	0.01	0.01	0.01	0.9	0.15	0.1	0.4	0.01
NNS	0.2	0.1	0.01	0.01	0.01	0.01	0.2	0.01	0.01	0.2
JJ	0.01	0.01	0.01	0.01	0.4	0.01	0.1	0.1	0.2	0.3

求以下句子的最佳POS标注：

给定： Steve Jobs , 42 years old

NNP	0.005	5e-4
,	0.003	1e-5
CD	0.0001	9e-5
NNS	0.002	
JJ	0.003	

隐马尔科夫模型：Viterbi算法

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	JJ	
NNP	0.05	0.01	0.1	0.01	0.04	0.01	0.2	0.01	0.01	0.05
,	0.01	0.01	0.01	0.7	0.01	0.01	0.2	0.1		0.3
CD	0.01	0.01	0.01	0.01	0.01	0.9	0.15	0.1	0.4	0.01
NNS	0.2	0.1	0.01	0.01	0.01	0.01	0.2	0.01	0.01	0.2
JJ	0.01	0.01	0.01	0.01	0.4	0.01	0.1	0.1	0.2	0.3

求以下句子的最佳POS标注：

给定： Steve Jobs , 42 years old

NNP	0.005	5e-4
,	0.003	1e-5
CD	0.0001	9e-5
NNS	0.002	1.2e-4
JJ	0.003	

隐马尔科夫模型：Viterbi算法

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	?	
NNP	0.05	0.01	0.1	0.01	0.04	0.01	0.2	0.01	0.01	0.05
,	0.01	0.01	0.01	0.7	0.01	0.01	0.2	0.1		0.3
CD	0.01	0.01	0.01	0.01	0.01	0.9	0.15	0.1	0.4	0.01
NNS	0.2	0.1	0.01	0.01	0.01	0.01	0.2	0.01	0.01	0.2
JJ	0.01	0.01	0.01	0.01	0.4	0.01	0.1	0.1	0.2	0.3

求以下句子的最佳POS标注：

给定： Steve Jobs , 42 years old

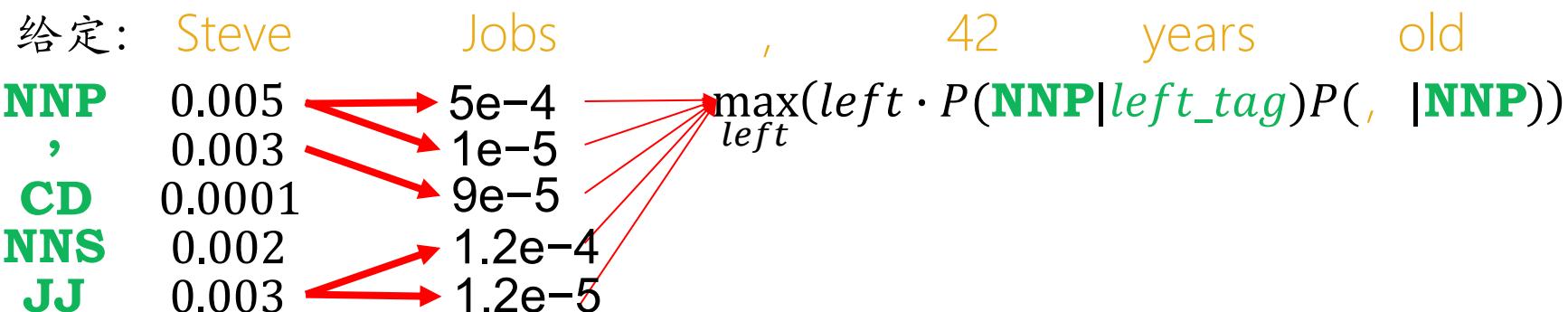
NNP	0.005	5e-4
,	0.003	1e-5
CD	0.0001	9e-5
NNS	0.002	1.2e-4
JJ	0.003	1.2e-5

隐马尔科夫模型：Viterbi算法

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	JJ	
NNP	0.05	0.01	0.1	0.01	0.04	0.01	0.2	0.01	0.01	0.05
,	0.01	0.01	0.01	0.7	0.01	0.01	0.2	0.1		0.3
CD	0.01	0.01	0.01	0.01	0.01	0.9	0.15	0.1	0.4	0.2
NNS	0.2	0.1	0.01	0.01	0.01	0.01	0.2	0.01	0.01	0.2
JJ	0.01	0.01	0.01	0.01	0.4	0.01	0.1	0.1	0.2	0.3

求以下句子的最佳POS标注：



隐马尔科夫模型：Viterbi算法

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	JJ	
NNP	0.05	0.01	0.1	0.01	0.04	0.01	0.2	0.01	0.01	0.05
,	0.01	0.01	0.01	0.7	0.01	0.01	0.2	0.1		0.3
CD	0.01	0.01	0.01	0.01	0.01	0.9	0.15	0.1	0.4	0.2
NNS	0.2	0.1	0.01	0.01	0.01	0.01	0.2	0.01	0.01	0.2
JJ	0.01	0.01	0.01	0.01	0.4	0.01	0.1	0.1	0.2	0.3

求以下句子的最佳POS标注：

给定： Steve Jobs , 42 years old

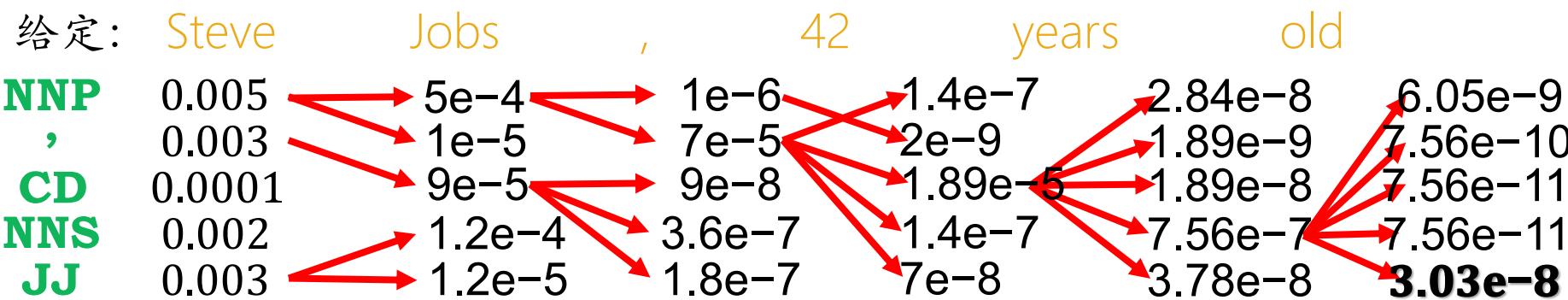
NNP	0.005	5e-4	1e-6
,	0.003	1e-5	7e-5
CD	0.0001	9e-5	9e-8
NNS	0.002	1.2e-4	3.6e-7
JJ	0.003	1.2e-5	1.8e-7

隐马尔科夫模型：Viterbi算法

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	JJ
NNP	0.05	0.01	0.1	0.01	0.04	0.01	0.2	0.01	0.05
,	0.01	0.01	0.01	0.7	0.01	0.01	0.2	0.1	0.3
CD	0.01	0.01	0.01	0.01	0.01	0.9	0.15	0.4	0.01
NNS	0.2	0.1	0.01	0.01	0.01	0.01	0.2	0.01	0.2
JJ	0.01	0.01	0.01	0.01	0.4	0.01	0.1	0.1	0.3

求以下句子的最佳POS标注：



隐马尔科夫模型：Viterbi算法

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	JJ
NNP	0.05	0.01	0.1	0.01	0.04	0.01	0.2	0.01	0.01
,	0.01	0.01	0.01	0.7	0.01	0.01	0.2	0.1	0.05
CD	0.01	0.01	0.01	0.01	0.01	0.9	0.15	0.01	0.01
NNS	0.2	0.1	0.01	0.01	0.01	0.01	0.2	0.01	0.01
JJ	0.01	0.01	0.01	0.01	0.4	0.01	0.1	0.1	0.3

求以下句子的最佳POS标注：

给定: Steve

Jobs

,

42

years

old

NNP

0.005



5e-4

1e-6

1.4e-7

6.05e-9

,

0.003



1e-5

7e-5

2e-9

7.56e-10

CD

0.0001



9e-5

9e-8

1.89e-5

7.56e-11

NNS

0.002



1.2e-4

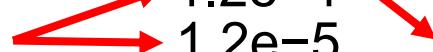
3.6e-7

1.4e-7

7.56e-11

JJ

0.003



1.2e-5

1.8e-7

7e-8

3.78e-8

隐马尔科夫模型：Viterbi算法

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	?	
NNP	0.05	0.01	0.1	0.01	0.04	0.01	0.2	0.01	0.01	0.05
,	0.01	0.01	0.01	0.7	0.01	0.01	0.2	0.3	0.2	0.3
CD	0.01	0.01	0.01	0.01	0.01	0.9	0.15	0.01	0.1	0.01
NNS	0.2	0.1	0.01	0.01	0.01	0.01	0.2	0.01	0.01	0.2
JJ	0.01	0.01	0.01	0.01	0.4	0.01	0.1	0.1	0.2	0.3

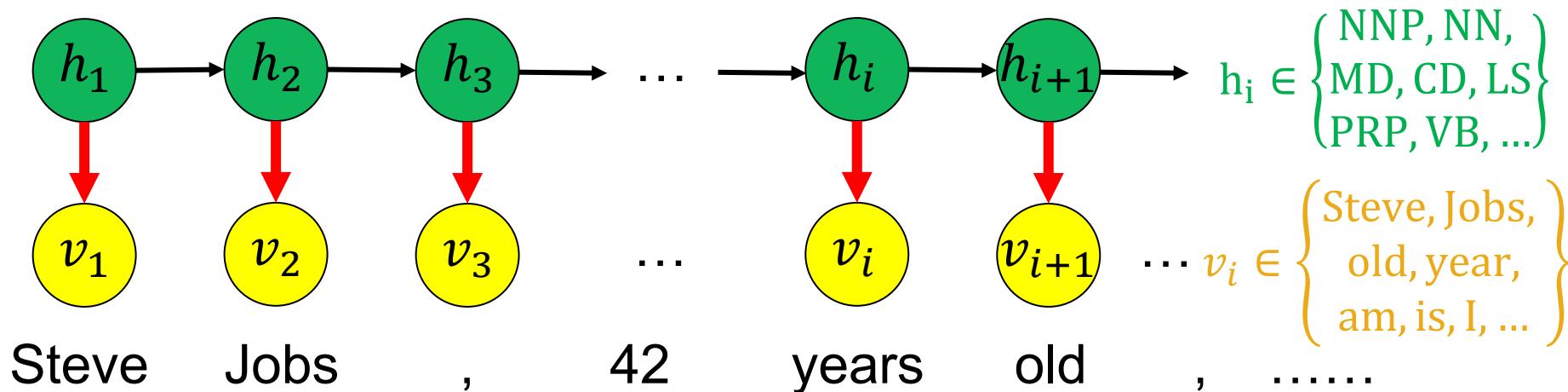
求以下句子的最佳POS标注：

给定:	Steve	Jobs	,	42	years	old
NNP	0.005	5e-4	1e-6	1.4e-7	2.84e-8	6.05e-9
,	0.003	1e-5	7e-5	2e-9	1.89e-9	7.56e-10
CD	0.0001	9e-5	9e-8	1.89e-5	1.89e-8	7.56e-11
NNS	0.002	1.2e-4	3.6e-7	1.4e-7	7.56e-7	7.56e-11
JJ	0.003	1.2e-5	1.8e-7	7e-8	3.78e-8	<u>3.03e-8</u>

NNP NNP CD NNS JJ



隐马尔科夫模型

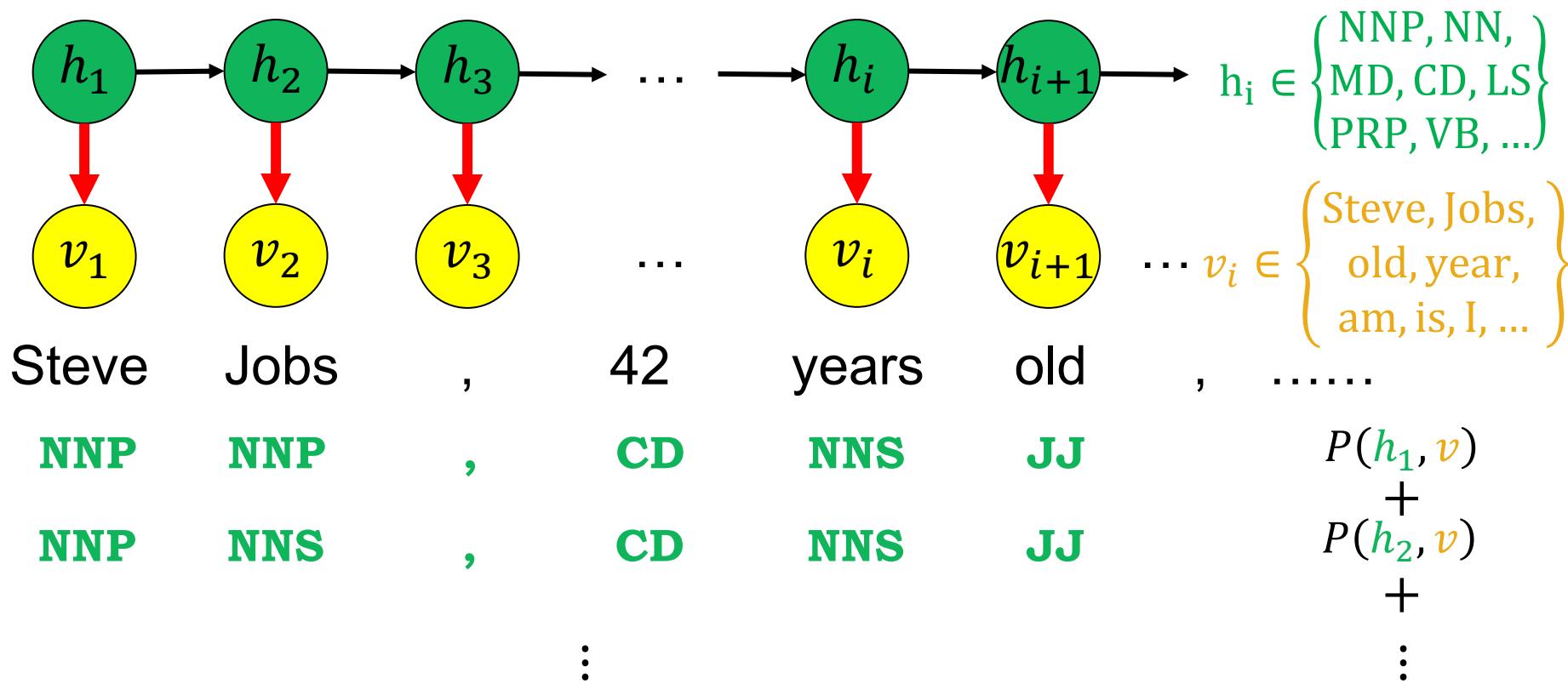


定义直接给出 • $P(\textcolor{teal}{h}, \textcolor{brown}{v})$

Viterbi 算法 • $\underset{h}{\operatorname{argmax}} P(\textcolor{teal}{h} | \textcolor{brown}{v}) = \underset{h}{\operatorname{argmax}} P(\textcolor{teal}{h}, \textcolor{brown}{v})$

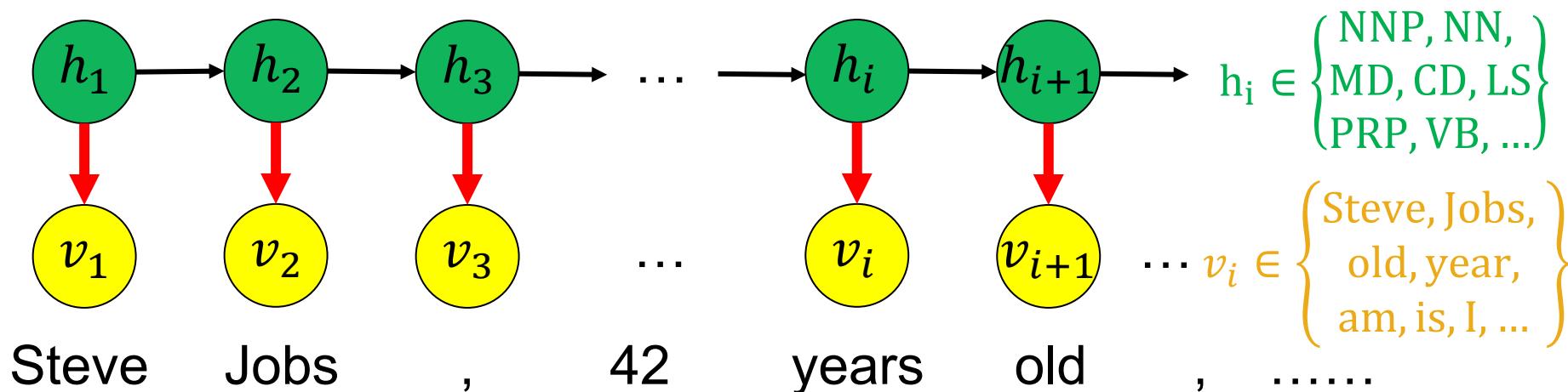
$$P(\textcolor{teal}{h} | \textcolor{brown}{v}) = \frac{P(\textcolor{brown}{v} | \textcolor{teal}{h}) P(\textcolor{teal}{h})}{P(\textcolor{brown}{v})}$$

隐马尔科夫模型



$$\bullet \quad P(v) = \sum_h P(h, v)$$

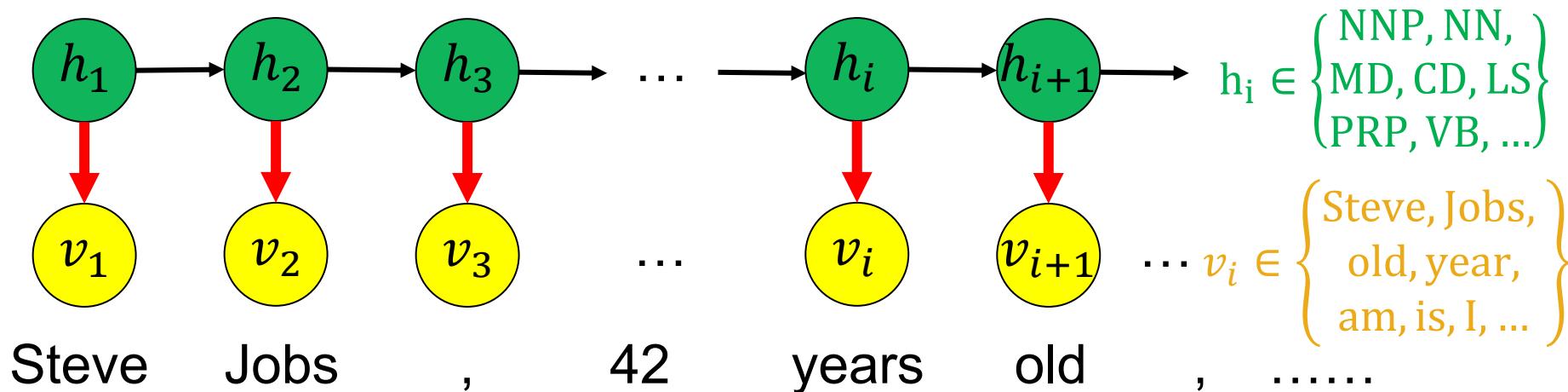
隐马尔科夫模型



$$\begin{aligned} P(v) &= \sum_h P(h, v) \\ &= \sum_h (P(v|h) \cdot P(h)) \\ &= \sum_h \left(\prod_{j=1}^N P(v_j|h_j) P(h_1) \prod_{i=1}^{N-1} P(h_{i+1}|h_i) \right) \end{aligned}$$

假设每个 h 有 S 个可能的状态，
序列长为 N ，则共有 S^N 个可能的 h 序列！

隐马尔科夫模型



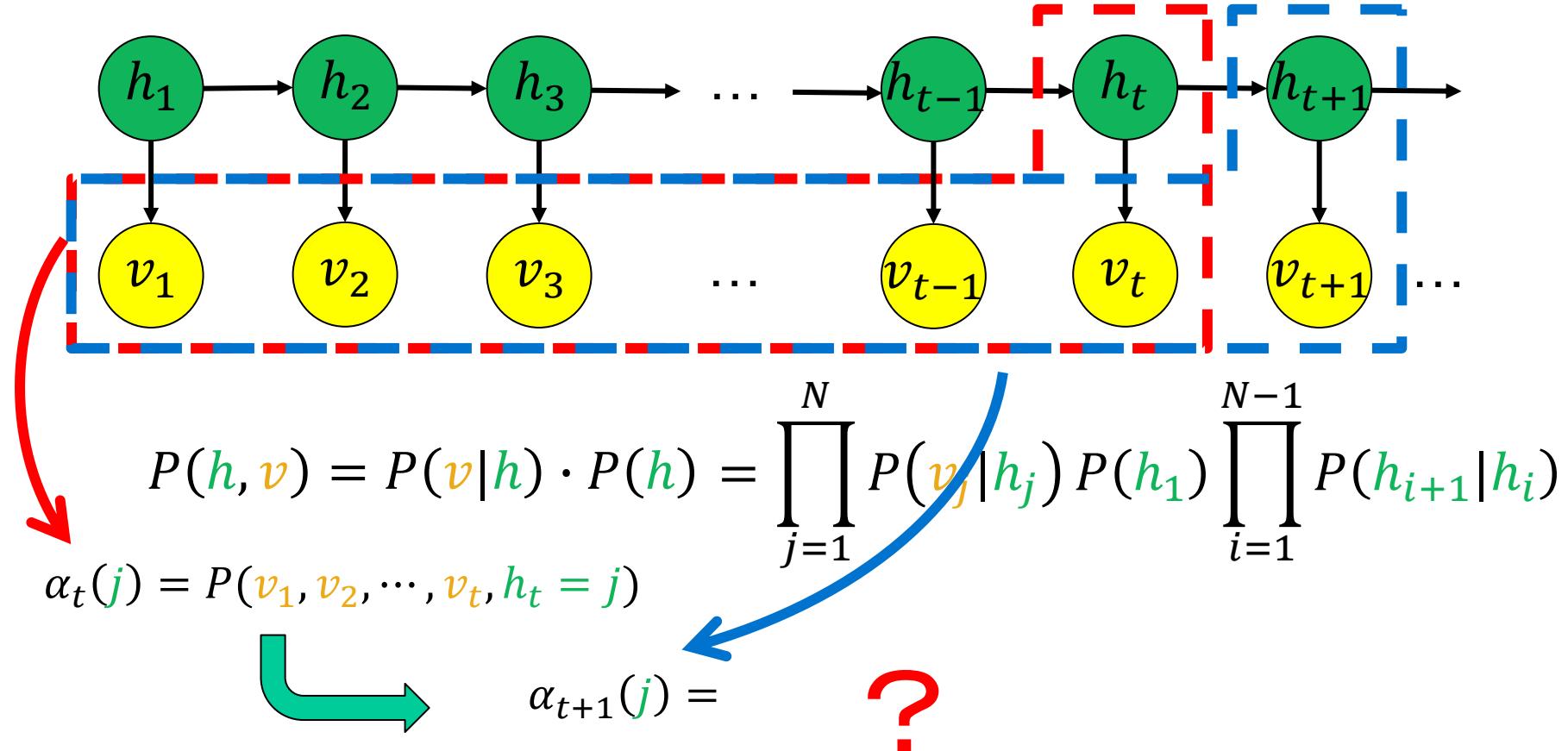
定义直接给出 • $P(\textcolor{teal}{h}, \textcolor{brown}{v})$

Viterbi 算法 • $\underset{h}{\operatorname{argmax}} P(\textcolor{teal}{h} | \textcolor{brown}{v}) = \underset{h}{\operatorname{argmax}} P(\textcolor{teal}{h}, \textcolor{brown}{v})$

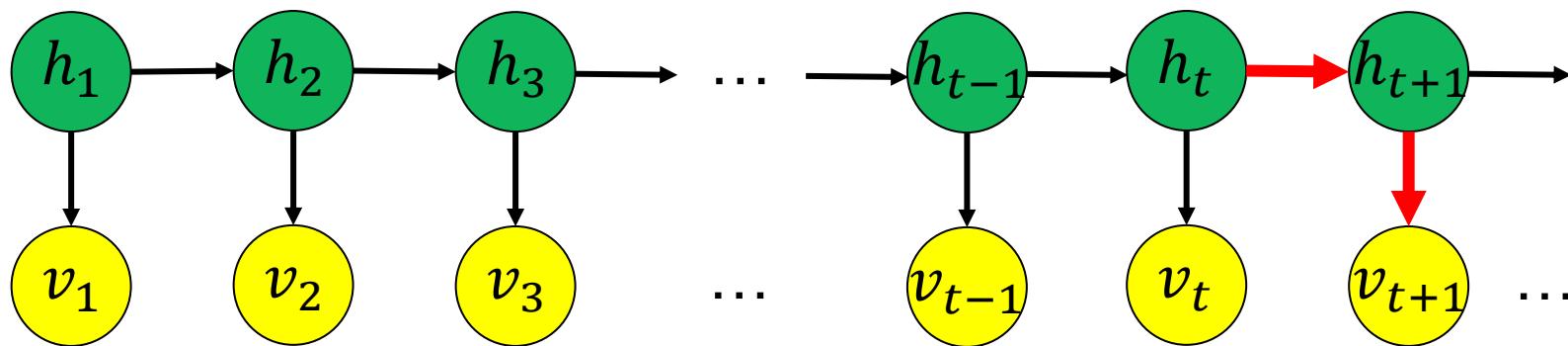
前向算法
后向算法

• $P(\textcolor{brown}{v})$

隐马尔科夫模型： α 与 β

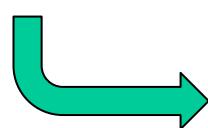


隐马尔科夫模型： α 与 β



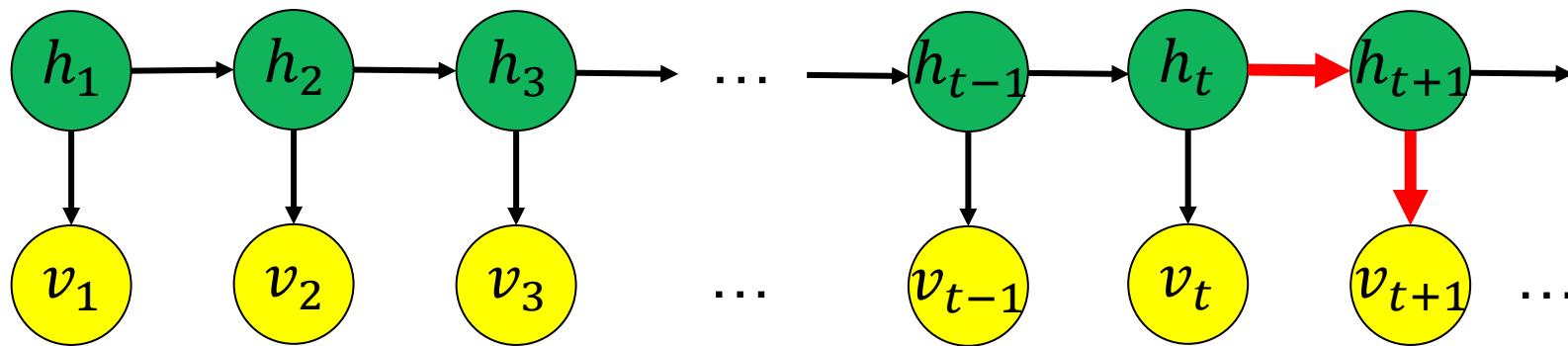
$$P(\mathbf{h}, \mathbf{v}) = P(\mathbf{v}|\mathbf{h}) \cdot P(\mathbf{h}) = \prod_{j=1}^N P(v_j|h_j) P(h_1) \prod_{i=1}^{N-1} P(h_{i+1}|h_i)$$

$$\alpha_t(j) = P(v_1, v_2, \dots, v_t, h_t = j)$$



$$\alpha_{t+1}(j) = \sum_{i=1}^S \alpha_t(i) P(h_{t+1} = j | h_t = i) P(v_{t+1} | h_{t+1} = j)$$

隐马尔科夫模型： α 与 β



$$\alpha_t(j) = P(v_1, v_2, \dots, v_t, h_t = j)$$

$\alpha_{t+1}(j) = \sum_{i=1}^S \alpha_t(i) P(h_{t+1} = j | h_t = i) P(v_{t+1} | h_{t+1} = j)$

$$\alpha_{t+1}(j) = P(v_1, v_2, \dots, v_t, v_{t+1}, h_{t+1} = j)$$

$$= \sum_{i=1}^S P(v_1, v_2, \dots, v_t, h_t = i, v_{t+1}, h_{t+1} = j)$$

$$= \sum_{\substack{i=1 \\ i \in S}}^S P(v_1, v_2, \dots, v_t, h_t = i) P(v_{t+1}, h_{t+1} = j | v_1, v_2, \dots, v_t, h_t = i)$$

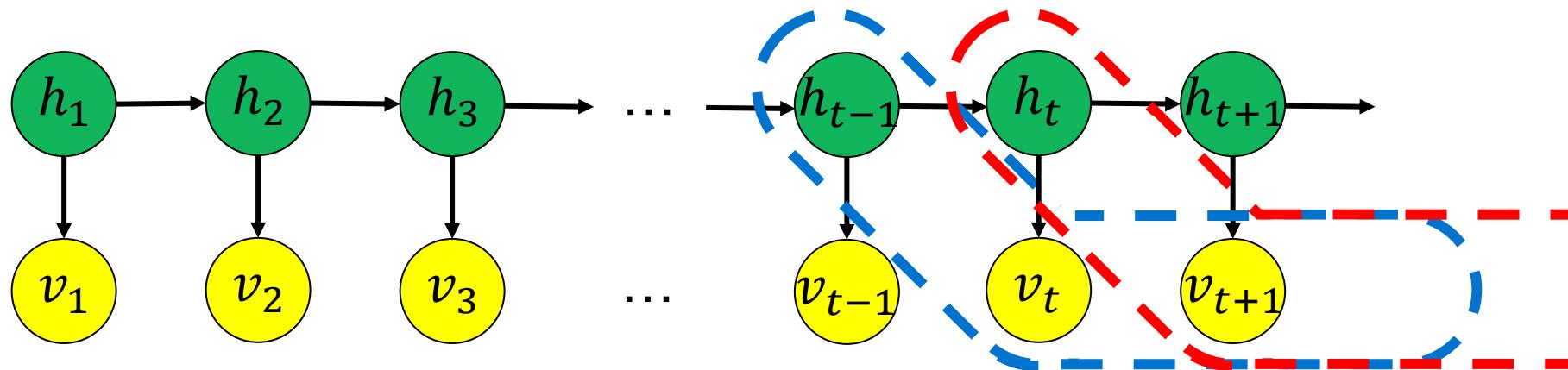
$$= \sum_{\substack{i=1 \\ i \in S}} \alpha_t(i) P(v_{t+1}, h_{t+1} = j | h_t = i)$$

$$= \sum_{\substack{i=1 \\ i \in S}} \alpha_t(i) P(h_{t+1} = j | h_t = i) P(v_{t+1} | h_{t+1} = j, h_t = i)$$

$$P(x) = \sum_y P(x, y)$$

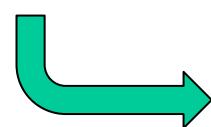
$$P(x, y) = P(x)P(y|x)$$

隐马尔科夫模型： α 与 β



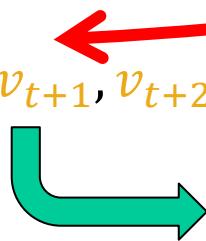
$$P(\mathbf{h}, \mathbf{v}) = P(\mathbf{v}|\mathbf{h}) \cdot P(\mathbf{h}) = \prod_{j=1}^N P(v_j|h_j) P(h_1) \prod_{i=1}^{N-1} P(h_{i+1}|h_i)$$

$$\alpha_t(j) = P(v_1, v_2, \dots, v_t, h_t = j)$$



$$\alpha_{t+1}(j) = \sum_{i=1}^S \alpha_t(i) P(h_{t+1} = j | h_t = i) P(v_{t+1} | h_{t+1} = j)$$

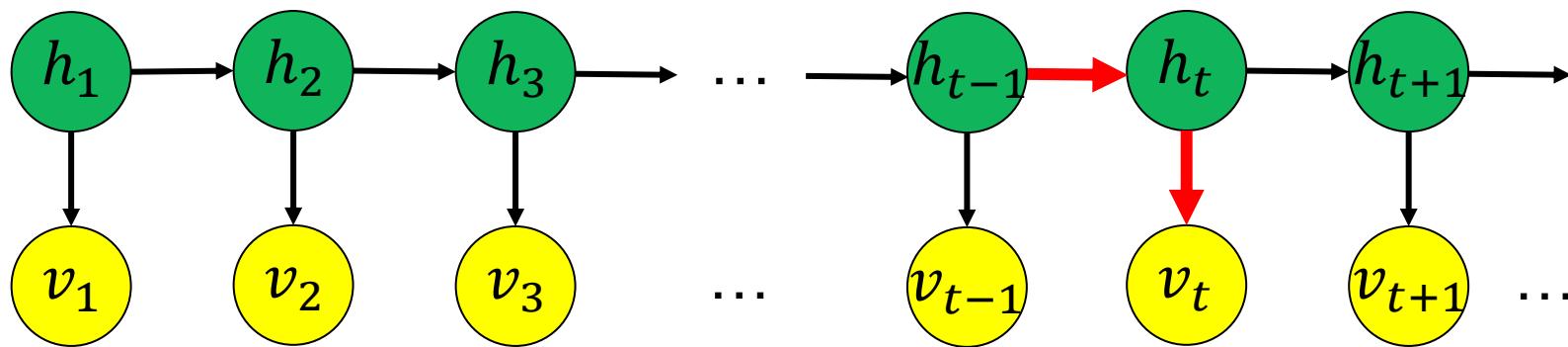
$$\beta_t(j) = P(v_{t+1}, v_{t+2}, \dots, v_N | h_t = j)$$



$$\beta_{t-1}(j) = ?$$

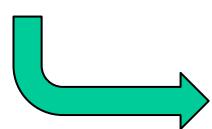
?

隐马尔科夫模型： α 与 β



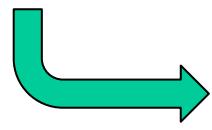
$$P(\textcolor{green}{h}, \textcolor{brown}{v}) = P(\textcolor{brown}{v}|\textcolor{green}{h}) \cdot P(\textcolor{green}{h}) = \prod_{j=1}^N P(\textcolor{brown}{v}_j | \textcolor{green}{h}_j) P(\textcolor{green}{h}_1) \prod_{i=1}^{N-1} P(\textcolor{green}{h}_{i+1} | \textcolor{green}{h}_i)$$

$$\alpha_t(j) = P(v_1, v_2, \dots, v_t, h_t = j)$$



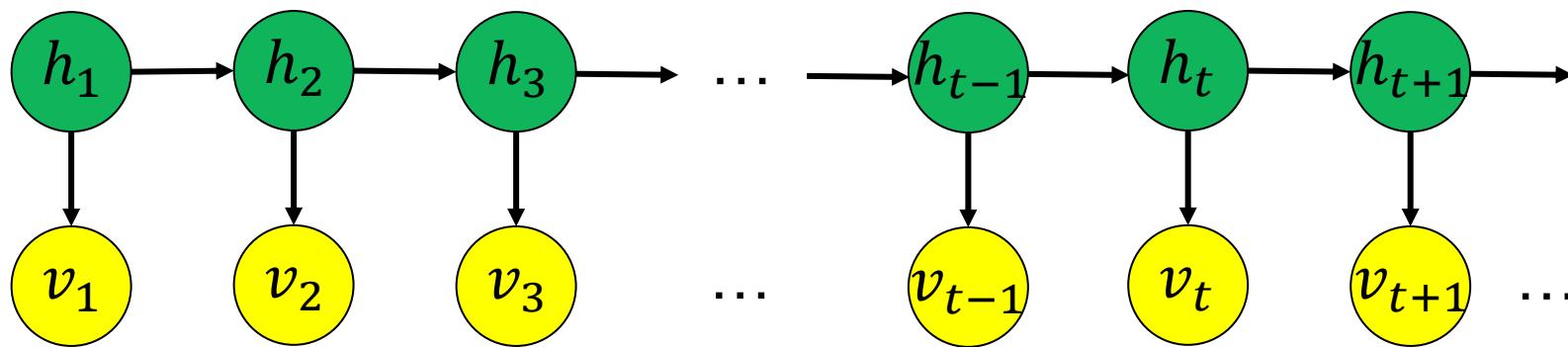
$$\alpha_{t+1}(j) = \sum_{i=1}^S \alpha_t(i) P(h_{t+1} = j | h_t = i) P(v_{t+1} | h_{t+1} = j)$$

$$\beta_t(j) = P(v_{t+1}, v_{t+2}, \dots, v_N | h_t = j)$$



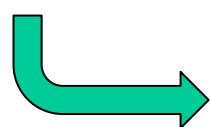
$$\beta_{t-1}(j) = \sum_{i=1}^S \beta_t(i) P(h_t = i | h_{t-1} = j) P(v_t | h_t = i)$$

隐马尔科夫模型： α 与 β



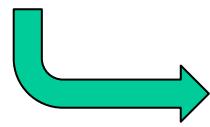
$$P(\textcolor{teal}{h}, \textcolor{brown}{v}) = P(\textcolor{brown}{v}|\textcolor{teal}{h}) \cdot P(\textcolor{teal}{h}) = \prod_{j=1}^N P(\textcolor{brown}{v}_j | \textcolor{teal}{h}_j) P(\textcolor{teal}{h}_1) \prod_{i=1}^{N-1} P(\textcolor{teal}{h}_{i+1} | \textcolor{teal}{h}_i)$$

$$\alpha_t(j) = P(v_1, v_2, \dots, v_t, h_t = j)$$



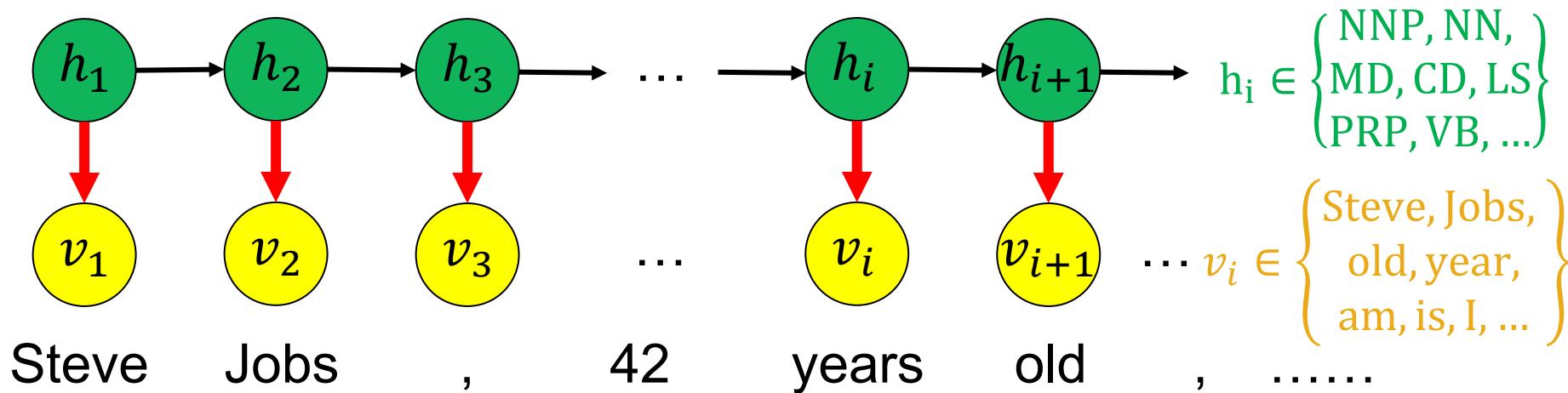
$$\alpha_{t+1}(j) = \sum_{i=1}^s \alpha_t(i) P(h_{t+1} = j | h_t = i) P(v_{t+1} | h_{t+1} = j)$$

$$\beta_t(j) = P(v_{t+1}, v_{t+2}, \dots, v_N | h_t = j)$$



$$\beta_{t-1}(j) = \sum_{i=1}^s \beta_t(i) P(h_t = i | h_{t-1} = j) P(v_t | h_t = i)$$

隐马尔科夫模型



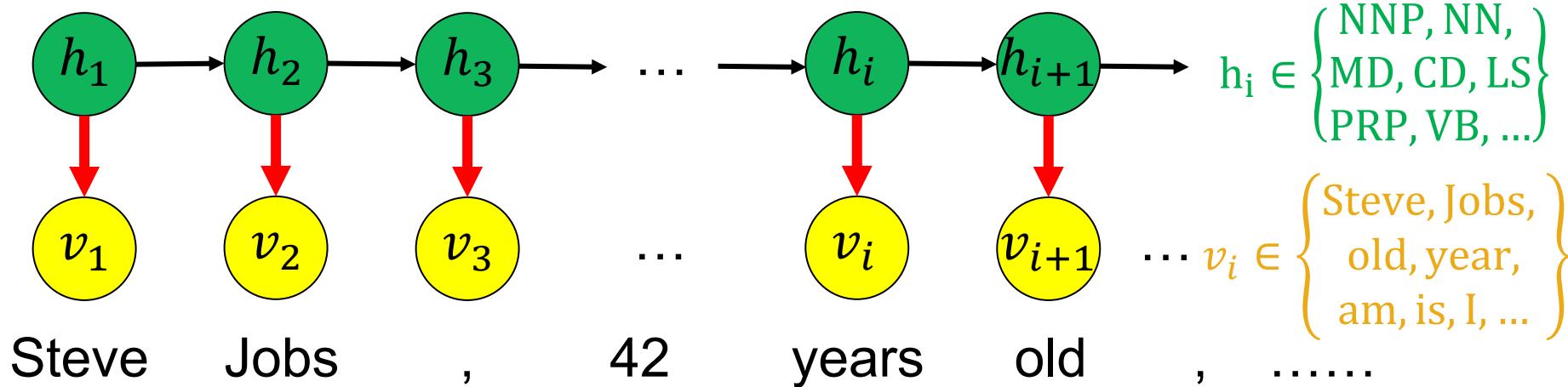
定义直接给出 • $P(\textcolor{teal}{h}, \textcolor{brown}{v})$

Viterbi 算法 • $\underset{h}{\operatorname{argmax}} P(\textcolor{teal}{h} | \textcolor{brown}{v}) = \underset{h}{\operatorname{argmax}} P(\textcolor{teal}{h}, \textcolor{brown}{v})$

前向算法
后向算法

• $P(\textcolor{brown}{v})$

隐马尔科夫模型：前向算法



$$P(v) = P(v_1, v_2, \dots, v_N)$$

$$= \sum_{h_N} P(v_1, v_2, \dots, v_N, h_N)$$

$$= \sum_{h_N} \alpha_N(j)$$

$$= \sum_{h_N} \sum_{h_{N-1}} \alpha_{N-1}(i) P(h_N | h_{N-1}) P(v_N | h_N)$$

$$= \sum_{h_N} \sum_{h_{N-1}} P(h_N | h_{N-1}) P(v_N | h_N) \sum_{h_{N-2}} \alpha_{N-2}(i) P(h_{N-1} | h_{N-2}) P(v_{N-1} | h_{N-1})$$

$$= \sum_{h_N} \sum_{h_{N-1}} \left\{ P(h_N | h_{N-1}) P(v_N | h_N) \cdots \sum_{h_{N-2}} \cdots \sum_{h_2} [\alpha_1(i) P(h_2 | h_1) P(v_2 | h_2)] \right\}$$

$$\alpha_{t+1}(j) = \sum_{i=1}^s \alpha_t(i) P(h_{t+1} = j | h_t = i) P(v_{t+1} | h_{t+1} = j)$$



隐马尔科夫模型：前向算法

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	JJ		
NNP	0.3	0.03	0.4	0.03	0.2	0.04	0.45	0.05	0.02	0.03	0.06
CD	0.01	0.02	0.02	0.9	0.02	0.03	0.3	0.4	0.2	0.09	0.35
NNS	0.02	0.02	0.02	0.02	0.02	0.9	0.15	0.1	0.5	0.24	0.01
JJ	0.5	0.3	0.05	0.05	0.05	0.05	0.45	0.07	0.03	0.2	0.23
	0.02	0.02	0.02	0.02	0.9	0.02	0.1	0.1	0.3	0.4	0.35

使用前向算法求以下句子出现的概率：

给定： Steve Jobs , 42 years old

隐马尔科夫模型：前向算法

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	?		
NNP	0.3	0.03	0.4	0.03	0.2	0.04	0.45	0.05	0.02	0.03	0.06
，	0.01	0.02	0.02	0.9	0.02	0.03	0.3	0.4	0.2	0.09	0.35
CD	0.02	0.02	0.02	0.02	0.02	0.9	0.15	0.1	0.5	0.24	0.01
NNS	0.5	0.3	0.05	0.05	0.05	0.05	0.45	0.07	0.03	0.2	0.23
JJ	0.02	0.02	0.02	0.02	0.9	0.02	0.1	0.1	0.3	0.4	0.35

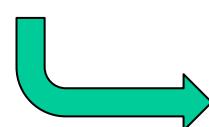
使用前向算法求以下句子出现的概率：

给定： Steve Jobs , 42 years old

S →

NNP
，
CD
NNS
JJ

$$\alpha_t(j) = P(v_1, v_2, \dots, v_t, h_t = j)$$



$$\alpha_{t+1}(j) = \sum_{i=1}^S \alpha_t(i) P(h_{t+1} = j | h_t = i) P(v_{t+1} | h_{t+1} = j)$$

隐马尔科夫模型：前向算法

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	?		
NNP	0.3	0.03	0.4	0.03	0.2	0.04	0.45	0.05	0.02	0.03	0.06
,	0.01	0.02	0.02	0.9	0.02	0.03	0.3	0.4	0.2	0.09	0.35
CD	0.02	0.02	0.02	0.02	0.02	0.9	0.15	0.1	0.5	0.24	0.01
NNS	0.5	0.3	0.05	0.05	0.05	0.05	0.45	0.07	0.03	0.2	0.23
JJ	0.02	0.02	0.02	0.02	0.9	0.02	0.1	0.1	0.3	0.4	0.35

使用前向算法求以下句子出现的概率：

给定： Steve Jobs , 42 years old

$$\text{NNP } P(\text{NNP})P(\text{Steve}|\text{NNP})$$

$$, \quad P(,)P(\text{Steve}|,)$$

$$\text{CD } P(\text{CD})P(\text{Steve}|\text{CD})$$

$$\text{NNS } P(\text{NNS})P(\text{Steve}|\text{NNS})$$

$$\text{JJ } P(\text{JJ})P(\text{Steve}|\text{JJ})$$

隐马尔科夫模型：前向算法

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	?		
NNP	0.3	0.03	0.4	0.03	0.2	0.04	0.45	0.05	0.02	0.03	0.06
,	0.01	0.02	0.02	0.9	0.02	0.03	0.3	0.4	0.2	0.09	0.35
CD	0.02	0.02	0.02	0.02	0.02	0.9	0.15	0.1	0.5	0.24	0.01
NNS	0.5	0.3	0.05	0.05	0.05	0.05	0.45	0.07	0.03	0.2	0.23
JJ	0.02	0.02	0.02	0.02	0.9	0.02	0.1	0.1	0.3	0.4	0.35

使用前向算法求以下句子出现的概率：

给定： Steve Jobs , 42 years old

$$\sum_{left} left \cdot P(\text{NNP}|left_tag)P(\text{Jobs}|\text{NNP})$$

NNP 0.024
, 0.007
CD 0.0002
NNS 0.0115
JJ 0.007

The diagram shows five red arrows originating from the words 'NNP', 'Jobs', 'old' in the input sentence and pointing to the values 0.024, 0.007, and 0.007 respectively in the NNP row of the transition matrix.

隐马尔科夫模型：前向算法

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	?		
NNP	0.3	0.03	0.4	0.03	0.2	0.04	0.45	0.05	0.02	0.03	0.06
,	0.01	0.02	0.02	0.9	0.02	0.03	0.3	0.4	0.2	0.09	0.35
CD	0.02	0.02	0.02	0.02	0.02	0.9	0.15	0.1	0.5	0.24	0.01
NNS	0.5	0.3	0.05	0.05	0.05	0.05	0.45	0.07	0.03	0.2	0.23
JJ	0.02	0.02	0.02	0.02	0.9	0.02	0.1	0.1	0.3	0.4	0.35

求以下句子的最佳POS标注：

给定： Steve Jobs , 42 years old

NNP 0.024
, 0.007
CD 0.0002
NNS 0.0115
JJ 0.007

$\max_{left} (left \cdot P(\text{NNP}|left_tag)P(\text{Jobs}|\text{NNP}))$

(Viterbi算法)

隐马尔科夫模型：前向算法

已知：

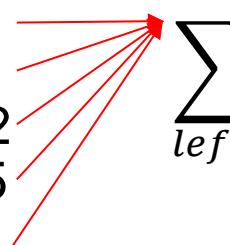
	jobs	years	steve	old	42	NNP	CD	SNN	?		
NNP	0.3	0.03	0.4	0.03	0.2	0.04	0.45	0.05	0.02	0.03	0.06
,	0.01	0.02	0.02	0.9	0.02	0.03	0.3	0.4	0.2	0.09	0.35
CD	0.02	0.02	0.02	0.02	0.02	0.9	0.15	0.1	0.5	0.24	0.01
NNS	0.5	0.3	0.05	0.05	0.05	0.05	0.45	0.07	0.03	0.2	0.23
JJ	0.02	0.02	0.02	0.02	0.9	0.02	0.1	0.1	0.3	0.4	0.35

使用前向算法求以下句子出现的概率：

给定： Steve Jobs , 42 years old

$$\sum_{left} left \cdot P(\text{NNP}|left_tag)P(\text{Jobs}|\text{NNP})$$

NNP 0.024
, 0.007
CD 0.0002
NNS 0.0115
JJ 0.007



隐马尔科夫模型：前向算法

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	?		
NNP	0.3	0.03	0.4	0.03	0.2	0.04	0.45	0.05	0.02	0.03	0.06
,	0.01	0.02	0.02	0.9	0.02	0.03	0.3	0.4	0.2	0.09	0.35
CD	0.02	0.02	0.02	0.02	0.02	0.9	0.15	0.1	0.5	0.24	0.01
NNS	0.5	0.3	0.05	0.05	0.05	0.05	0.45	0.07	0.03	0.2	0.23
JJ	0.02	0.02	0.02	0.02	0.9	0.02	0.1	0.1	0.3	0.4	0.35

使用前向算法求以下句子出现的概率：

给定： Steve Jobs , 42 years old

NNP 0.024
, 0.007
CD 0.0002
NNS 0.0115
JJ 0.007

5.64e-3

隐马尔科夫模型：前向算法

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	?		
NNP	0.3	0.03	0.4	0.03	0.2	0.04	0.45	0.05	0.02	0.03	0.06
,	0.01	0.02	0.02	0.9	0.02	0.03	0.3	0.4	0.2	0.09	0.35
CD	0.02	0.02	0.02	0.02	0.02	0.9	0.15	0.1	0.5	0.24	0.01
NNS	0.5	0.3	0.05	0.05	0.05	0.05	0.45	0.07	0.03	0.2	0.23
JJ	0.02	0.02	0.02	0.02	0.9	0.02	0.1	0.1	0.3	0.4	0.35

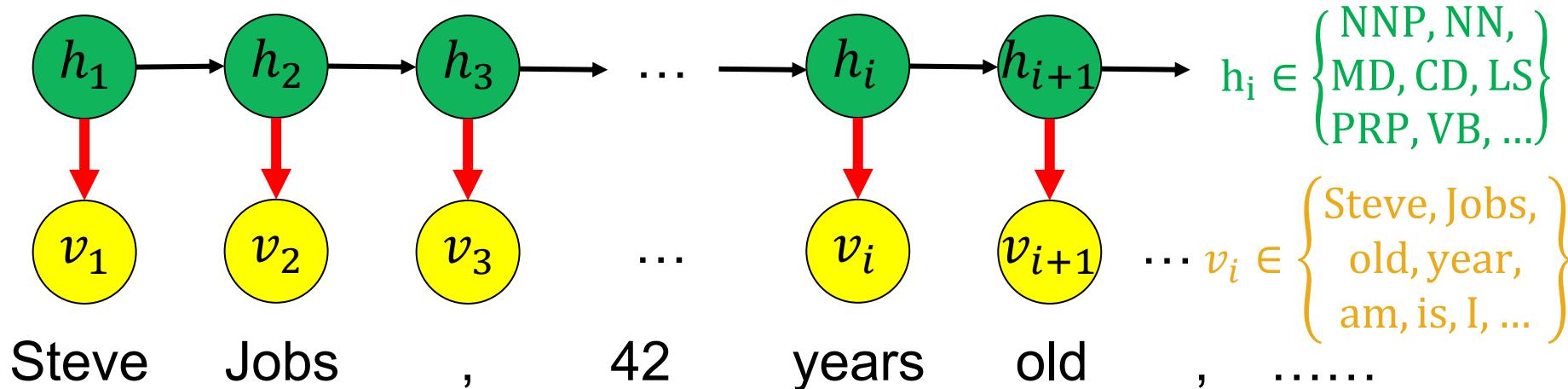
使用前向算法求以下句子出现的概率：

	Steve	Jobs	,	42	years	old	
NNP	0.024	5.64e-3	1.08e-4	3.59e-5	5.47e-6	1.45e-5	
,	0.007	1.44e-4	2.80e-3	2.46e-6	6.80e-7	8.27e-7	
CD	0.0002	1.10e-4	1.04e-5	1.01e-3	2.14e-6	2.40e-7	
NNS	0.0115	2.21e-3	1.51e-5	2.86e-5	1.53e-4	3.71e-7	
JJ	0.007	1.30e-4	1.41e-5	5.32e-6	5.05e-6	3.01e-5	

$$P(\text{Steve}, \text{Jobs}, , , 42, \text{years}, \text{old}) = 4.60e - 5$$

Σ

隐马尔科夫模型：后向算法



$$P(v) = P(v_1, v_2, \dots, v_N)$$

$$= \sum_{h_1} P(v_1, v_2, \dots, v_N, h_1)$$

$$= \sum_{h_1} P(v_1, h_1) P(v_2, \dots, v_N | v_1, h_1)$$

$$= \sum_{h_1} P(h_1) P(v_1 | h_1) \beta_1(j)$$

$$= \sum_{h_1} P(h_1) P(v_1 | h_1) \sum_{h_2} \beta_2(i) P(h_2 | h_1) P(v_2 | h_2)$$

$$= \sum_{h_1} P(h_1) P(v_1 | h_1) \sum_{h_2} P(h_2 | h_1) P(v_2 | h_2) \sum_{h_3} \dots \sum_{h_{N-1}} \beta_{N-1}(i) P(h_{N-1} | h_{N-2}) P(v_{N-1} | h_{N-1})$$

$$= \sum_{h_1} P(h_1) P(v_1 | h_1) \sum_{h_2} P(h_2 | h_1) P(v_2 | h_2) \sum_{h_3} \dots \sum_{h_N} \beta_N(i) P(h_N | h_{N-1}) P(v_N | h_N)$$

$$\beta_{t-1}(j) = \sum_{i=1}^S \beta_t(i) P(h_t = i | h_{t-1} = j) P(v_t | h_t = i)$$



隐马尔科夫模型：后向算法

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	?		
NNP	0.3	0.03	0.4	0.03	0.2	0.04	0.45	0.05	0.02	0.03	0.06
,	0.01	0.02	0.02	0.9	0.02	0.03	0.3	0.4	0.2	0.09	0.35
CD	0.02	0.02	0.02	0.02	0.02	0.9	0.15	0.1	0.5	0.24	0.01
NNS	0.5	0.3	0.05	0.05	0.05	0.05	0.45	0.07	0.03	0.2	0.23
JJ	0.02	0.02	0.02	0.02	0.9	0.02	0.1	0.1	0.3	0.4	0.35

使用后向算法求以下句子出现的概率：

给定： Steve Jobs , 42 years old

S →

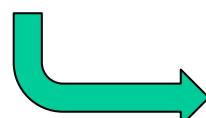
NNP
,

CD

NNS

JJ

$$\beta_t(j) = P(v_{t+1}, v_{t+2}, \dots, v_N | h_t = j)$$



$$\beta_{t-1}(j) = \sum_{i=1}^S \beta_t(i) P(h_t = i | h_{t-1} = j) P(v_t | h_t = i)$$

隐马尔科夫模型：后向算法

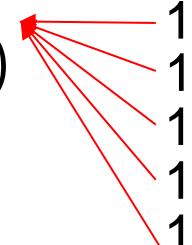
已知：

	jobs	years	steve	old	42	NNP	CD	SNN	?		
NNP	0.3	0.03	0.4	0.03	0.2	0.04	0.45	0.05	0.02	0.03	0.06
,	0.01	0.02	0.02	0.9	0.02	0.03	0.3	0.4	0.2	0.09	0.35
CD	0.02	0.02	0.02	0.02	0.02	0.9	0.15	0.1	0.5	0.24	0.01
NNS	0.5	0.3	0.05	0.05	0.05	0.05	0.45	0.07	0.03	0.2	0.23
JJ	0.02	0.02	0.02	0.02	0.9	0.02	0.1	0.1	0.3	0.4	0.35

使用后向算法求以下句子出现的概率：

给定： Steve Jobs , 42 years old

NNP
,
CD
NNS
JJ

$$\sum_{right} (right \cdot P(right_tag | NNP) P(old | right_tag))$$


隐马尔科夫模型：后向算法

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	?		
NNP	0.3	0.03	0.4	0.03	0.2	0.04	0.45	0.05	0.02	0.03	0.06
,	0.01	0.02	0.02	0.9	0.02	0.03	0.3	0.4	0.2	0.09	0.35
CD	0.02	0.02	0.02	0.02	0.02	0.9	0.15	0.1	0.5	0.24	0.01
NNS	0.5	0.3	0.05	0.05	0.05	0.05	0.45	0.07	0.03	0.2	0.23
JJ	0.02	0.02	0.02	0.02	0.9	0.02	0.1	0.1	0.3	0.4	0.35

使用后向算法求以下句子出现的概率：

给定： Steve Jobs , 42 years old

NNP
,
CD
NNS
JJ

0.128

$$\sum_{right} (right \cdot P(right_tag | ,)P(old | right_tag))$$

1
1
1
1
1
1

隐马尔科夫模型：后向算法

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	?		
NNP	0.3	0.03	0.4	0.03	0.2	0.04	0.45	0.05	0.02	0.03	0.06
,	0.01	0.02	0.02	0.9	0.02	0.03	0.3	0.4	0.2	0.09	0.35
CD	0.02	0.02	0.02	0.02	0.02	0.9	0.15	0.1	0.5	0.24	0.01
NNS	0.5	0.3	0.05	0.05	0.05	0.05	0.45	0.07	0.03	0.2	0.23
JJ	0.02	0.02	0.02	0.02	0.9	0.02	0.1	0.1	0.3	0.4	0.35

使用后向算法求以下句子出现的概率：

给定： Steve Jobs , 42 years old

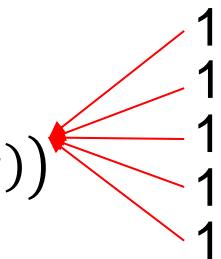
NNP
,

0.128

CD
NNS
JJ

0.1592

$$\sum_{right} (right \cdot P(right_tag|CD)P(old|right_tag))$$



隐马尔科夫模型：后向算法

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	?		
NNP	0.3	0.03	0.4	0.03	0.2	0.04	0.45	0.05	0.02	0.03	0.06
,	0.01	0.02	0.02	0.9	0.02	0.03	0.3	0.4	0.2	0.09	0.35
CD	0.02	0.02	0.02	0.02	0.02	0.9	0.15	0.1	0.5	0.24	0.01
NNS	0.5	0.3	0.05	0.05	0.05	0.05	0.45	0.07	0.03	0.2	0.23
JJ	0.02	0.02	0.02	0.02	0.9	0.02	0.1	0.1	0.3	0.4	0.35

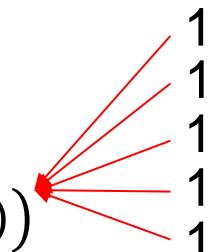
使用后向算法求以下句子出现的概率：

给定： Steve Jobs , 42 years old

NNP
,
CD
NNS
JJ

0.128
0.1592
0.2732

$$\sum_{right} (right \cdot P(right_tag | NNS) P(old | right_tag))$$



隐马尔科夫模型：后向算法

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	?		
NNP	0.3	0.03	0.4	0.03	0.2	0.04	0.45	0.05	0.02	0.03	0.06
,	0.01	0.02	0.02	0.9	0.02	0.03	0.3	0.4	0.2	0.09	0.35
CD	0.02	0.02	0.02	0.02	0.02	0.9	0.15	0.1	0.5	0.24	0.01
NNS	0.5	0.3	0.05	0.05	0.05	0.05	0.45	0.07	0.03	0.2	0.23
JJ	0.02	0.02	0.02	0.02	0.9	0.02	0.1	0.1	0.3	0.4	0.35

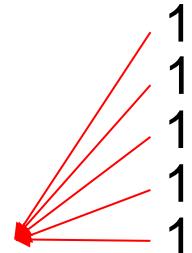
使用后向算法求以下句子出现的概率：

给定： Steve Jobs , 42 years old

NNP
,
CD
NNS
JJ

0.128
0.1592
0.2732
0.2779

$$\sum_{right} (right \cdot P(right_tag|JJ)P(old|right_tag))$$



隐马尔科夫模型：后向算法

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	?		
NNP	0.3	0.03	0.4	0.03	0.2	0.04	0.45	0.05	0.02	0.03	0.06
,	0.01	0.02	0.02	0.9	0.02	0.03	0.3	0.4	0.2	0.09	0.35
CD	0.02	0.02	0.02	0.02	0.02	0.9	0.15	0.1	0.5	0.24	0.01
NNS	0.5	0.3	0.05	0.05	0.05	0.05	0.45	0.07	0.03	0.2	0.23
JJ	0.02	0.02	0.02	0.02	0.9	0.02	0.1	0.1	0.3	0.4	0.35

使用后向算法求以下句子出现的概率：

给定： Steve Jobs , 42 years old

NNP
,
CD
NNS
JJ

0.128
0.1592
0.2732
0.2779
0.399

1
1
1
1
1

隐马尔科夫模型：后向算法

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	?				
，	NNP	0.3	0.03	0.4	0.03	0.2	0.04	0.45	0.45	0.05	0.02	0.03	0.06
		0.01	0.02	0.02	0.9	0.02	0.03	0.3	0.01	0.4	0.2	0.09	0.35
	CD	0.02	0.02	0.02	0.02	0.02	0.9	0.15	0.01	0.1	0.5	0.24	0.01
	NNS	0.5	0.3	0.05	0.05	0.05	0.05	0.45	0.25	0.07	0.03	0.2	0.23
	JJ	0.02	0.02	0.02	0.02	0.9	0.02	0.1	0.1	0.1	0.3	0.4	0.35

使用后向算法求以下句子出现的概率：

给定： Steve Jobs , 42 years old

NNP		0.128	1
,		0.1592	1
CD		0.2732	1
NNS		0.2779	1
JJ		0.399	1

隐马尔科夫模型：后向算法

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	?		
NNP	0.3	0.03	0.4	0.03	0.2	0.04	0.45	0.05	0.02	0.03	0.06
,	0.01	0.02	0.02	0.9	0.02	0.03	0.3	0.4	0.2	0.09	0.35
CD	0.02	0.02	0.02	0.02	0.02	0.9	0.15	0.1	0.5	0.24	0.01
NNS	0.5	0.3	0.05	0.05	0.05	0.05	0.45	0.07	0.03	0.2	0.23
JJ	0.02	0.02	0.02	0.02	0.9	0.02	0.1	0.1	0.3	0.4	0.35

使用后向算法求以下句子出现的概率：

给定： Steve Jobs , 42 years old

$$\begin{aligned} \text{NNP} & \quad 5.34e-3 \\ , & \quad 2.08e-2 \end{aligned}$$

$$\begin{aligned} \text{CD} & \quad 0.128 \\ \text{NNS} & \quad 0.1592 \\ \text{JJ} & \quad 0.2732 \end{aligned}$$

$$\sum_{right} (right \cdot P(right_tag|CD)P(years|right_tag))$$

$$\begin{aligned} & \quad 0.2779 \\ & \quad 0.399 \end{aligned}$$

隐马尔科夫模型：后向算法

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	?		
NNP	0.3	0.03	0.4	0.03	0.2	0.04	0.45	0.05	0.02	0.03	0.06
,	0.01	0.02	0.02	0.9	0.02	0.03	0.3	0.4	0.2	0.09	0.35
CD	0.02	0.02	0.02	0.02	0.02	0.9	0.15	0.1	0.5	0.24	0.01
NNS	0.5	0.3	0.05	0.05	0.05	0.05	0.45	0.07	0.03	0.2	0.23
JJ	0.02	0.02	0.02	0.02	0.9	0.02	0.1	0.1	0.3	0.4	0.35

使用后向算法求以下句子出现的概率：

给定： Steve Jobs , 42 years old

NNP	5.34e-3	0.128	1
,	2.08e-2	0.1592	1
CD	4.48e-2	0.2732	1
NNS	7.00e-3	0.2779	1
JJ	2.95e-2	0.399	1

隐马尔科夫模型：后向算法

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	?		
NNP	0.3	0.03	0.4	0.03	0.2	0.04	0.45	0.05	0.02	0.03	0.06
,	0.01	0.02	0.02	0.9	0.02	0.03	0.3	0.4	0.2	0.09	0.35
CD	0.02	0.02	0.02	0.02	0.02	0.9	0.15	0.1	0.5	0.24	0.01
NNS	0.5	0.3	0.05	0.05	0.05	0.05	0.45	0.07	0.03	0.2	0.23
JJ	0.02	0.02	0.02	0.02	0.9	0.02	0.1	0.1	0.3	0.4	0.35

使用后向算法求以下句子出现的概率：

给定： Steve Jobs , 42 years old

NNP	9.37e-4	6.65e-3	2.42e-3	5.34e-3	0.128	1
,	9.76e-4	2.44e-4	1.63e-2	2.08e-2	0.1592	1
CD	1.24e-3	2.68e-4	4.38e-3	4.48e-2	0.2732	1
NNS	9.60e-4	3.73e-3	3.20e-3	7.00e-3	0.2779	1
JJ	7.72e-4	1.57e-3	4.45e-3	2.95e-2	0.399	1

隐马尔科夫模型：后向算法

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	?		
NNP	0.3	0.03	0.4	0.03	0.2	0.04	0.45	0.05	0.02	0.03	0.06
,	0.01	0.02	0.02	0.9	0.02	0.03	0.3	0.4	0.2	0.09	0.35
CD	0.02	0.02	0.02	0.02	0.02	0.9	0.15	0.1	0.5	0.24	0.01
NNS	0.5	0.3	0.05	0.05	0.05	0.05	0.45	0.07	0.03	0.2	0.23
JJ	0.02	0.02	0.02	0.02	0.9	0.02	0.1	0.1	0.3	0.4	0.35

使用后向算法求以下句子出现的概率：

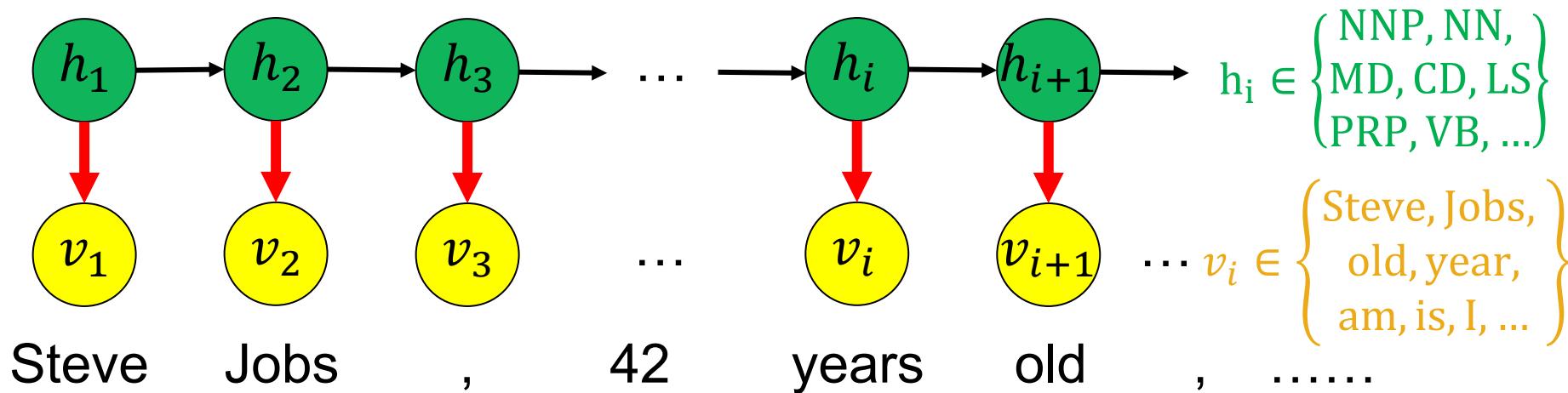
给定： Steve Jobs , 42 years old

NNP	9.37e-4	6.65e-3	2.42e-3	5.34e-3	0.128	1
,	9.76e-4	2.44e-4	1.63e-2	2.08e-2	0.1592	1
CD	1.24e-3	2.68e-4	4.38e-3	4.48e-2	0.2732	1
NNS	9.60e-4	3.73e-3	3.20e-3	7.00e-3	0.2779	1
JJ	7.72e-4	1.57e-3	4.45e-3	2.95e-2	0.399	1

$$\beta_1(j) = P(v_2, v_3, \dots, v_N | h_1 = j)$$

$$P(v_1, v_2, v_3, \dots, v_N) = \sum_j \beta_1(j) P(j) P(\text{Steve} | j) = 4.60e-5$$

隐马尔科夫模型



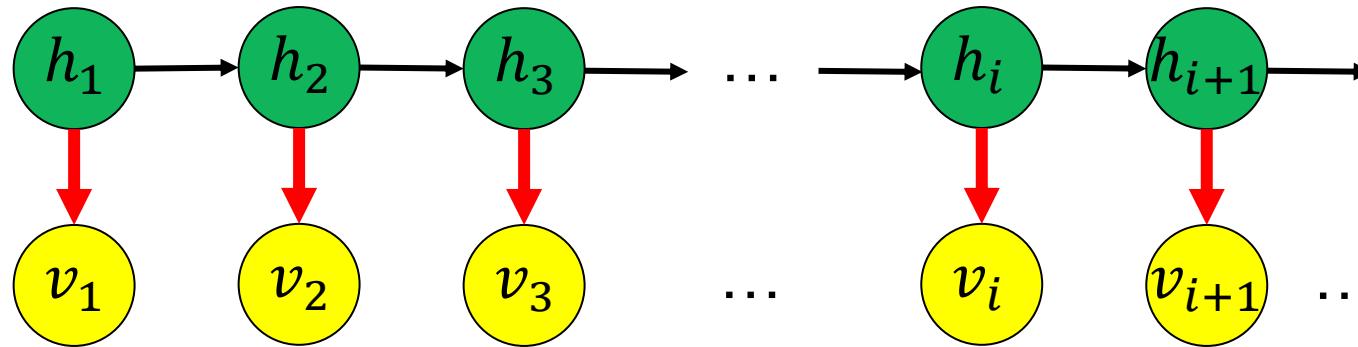
定义直接给出 • $P(\textcolor{teal}{h}, \textcolor{brown}{v})$

Viterbi 算法 • $\underset{h}{\operatorname{argmax}} P(\textcolor{teal}{h} | \textcolor{brown}{v}) = \underset{h}{\operatorname{argmax}} P(\textcolor{teal}{h}, \textcolor{brown}{v})$

前向算法
后向算法

• $P(\textcolor{brown}{v})$

隐马尔科夫模型

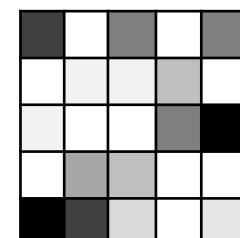
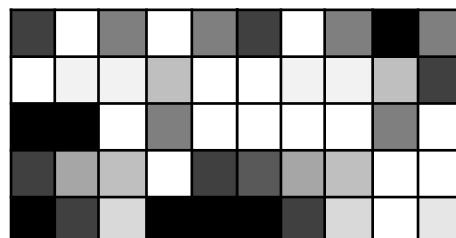


隐马尔科夫模型的定义

$$P(v, h) = P(v|h) \cdot P(h)$$

$$= \prod_{j=1}^N P(v_j|h_j) P(h_1) \prod_{i=1}^{N-1} P(h_{i+1}|h_i)$$

依照左式对联合概率分布 $P(v, h)$ 建模，即被称为 HMM 模型

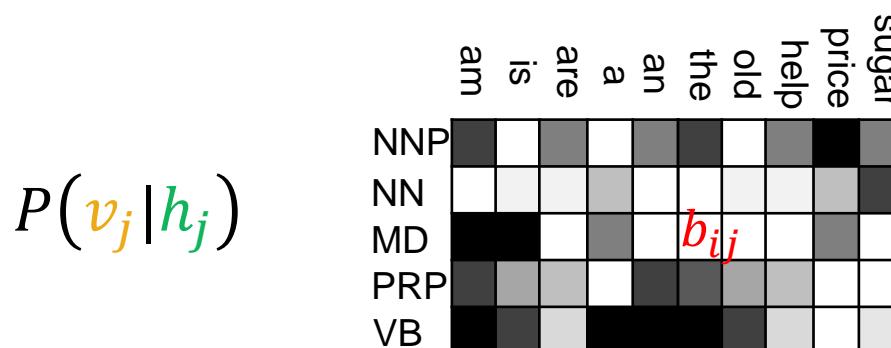
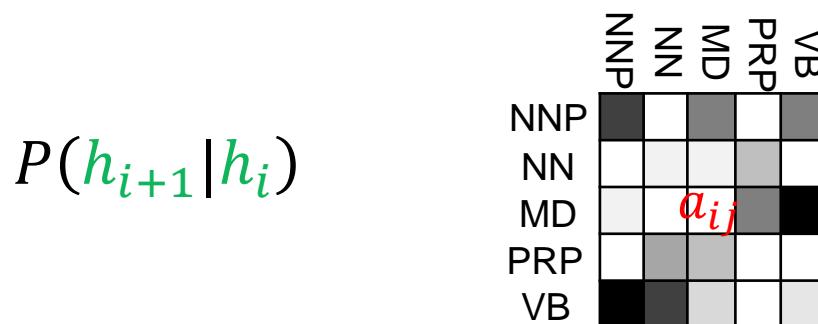


v, h 都应是离散随机变量的离散序列

参数：

隐马尔科夫模型：参数的有监督学习

$$P(v, h) = \prod_{j=1}^N P(v_j | h_j) P(h_1) \prod_{i=1}^{N-1} P(h_{i+1} | h_i)$$



给定POS的标注数据集，
这些参数即可通过数数得到：

$$\hat{\pi}_i = P(h_1 = i) = \frac{\#(h_1 = i)}{\#(\text{句子})}$$

$$\widehat{a_{ij}} = P(h_{t+1} = i | h_t = j) = \frac{\#\text{(tags} = (i,j)\text{)}}{\#\text{(tag} = i\text{)}}$$

$$\widehat{b_{ij}} = P(v_t = j | h_t = i) = \frac{\#\text{(tag} = i, word = j\text{)}}{\#\text{(tag} = i\text{)}}$$

隐马尔科夫模型：参数的有监督学习

$$P(v, h) = \prod_{j=1}^N P(v_j | h_j) P(h_1) \prod_{i=1}^{N-1} P(h_{i+1} | h_i)$$

给定POS的标注数据集，
这些参数即可通过数数得到：

$P(h_1)$

NNP	
NN	π_i
MD	
PRP	
VB	

$P(h_{i+1} | h_i)$

	NNP	NN	MD	PRP	VB
NNP					
NN					
MD			a_{ij}		
PRP					
VB					

$P(v_j | h_j)$

	am	is	are	a	an	the	old	help	price	sugar
NNP										
NN										
MD						b_{ij}				
PRP										
VB										

$\widehat{a}_{ij} = P(h_{t+1} = i | h_t = j)$
 $= \frac{\#(\text{tags} = (i, j))}{\#(\text{tag} = i)}$

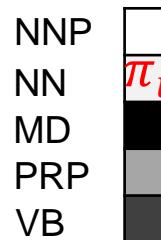
$\widehat{b}_{ij} = P(v_t = j | h_t = i)$
 $= \frac{\#(\text{tag} = i, \text{word} = j)}{\#(\text{tag} = i)}$

隐马尔科夫模型：参数的无监督学习

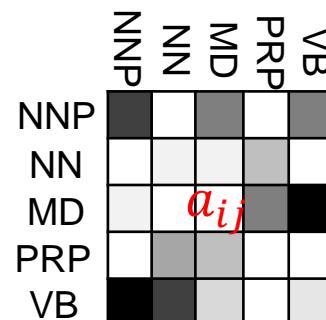
$$P(v, h) = \prod_{j=1}^N P(v_j | h_j) P(h_1) \prod_{i=1}^{N-1} P(h_{i+1} | h_i)$$

给定POS的无标注标注数据集，
这些参数如何计算？

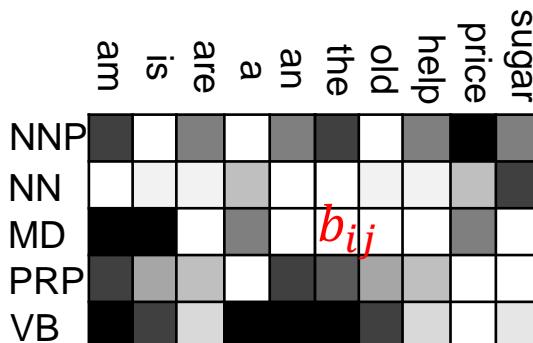
$$P(h_1)$$



$$P(h_{i+1} | h_i)$$



$$P(v_j | h_j)$$



$$\hat{\pi}_i = P(h_1 = i) = \frac{\#(h_1 = i)}{\#(\text{句子})}$$

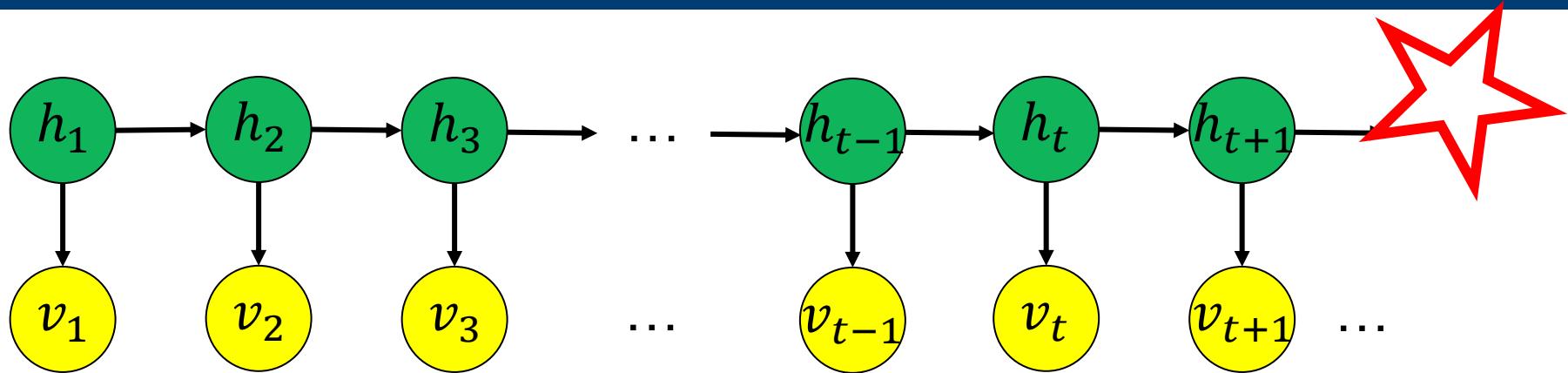
$$\widehat{a}_{ij} = P(h_{t+1} = i | h_t = j)$$

$$= \frac{\#(\text{tags} = (i, j))}{\#(\text{tag} = i)}$$

$$\widehat{b}_{ij} = P(v_t = j | h_t = i)$$

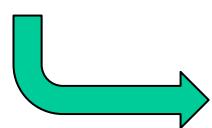
$$= \frac{\#(\text{tag} = i, \text{word} = j)}{\#(\text{tag} = i)}$$

隐马尔科夫模型： α 与 β



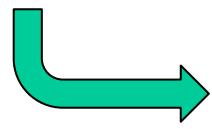
$$P(\mathbf{h}, \mathbf{v}) = P(\mathbf{v}|\mathbf{h}) \cdot P(\mathbf{h}) = \prod_{j=1}^N P(v_j|h_j) P(h_1) \prod_{i=1}^{N-1} P(h_{i+1}|h_i)$$

$$\alpha_t(j) = P(v_1, v_2, \dots, v_t, h_t = j)$$



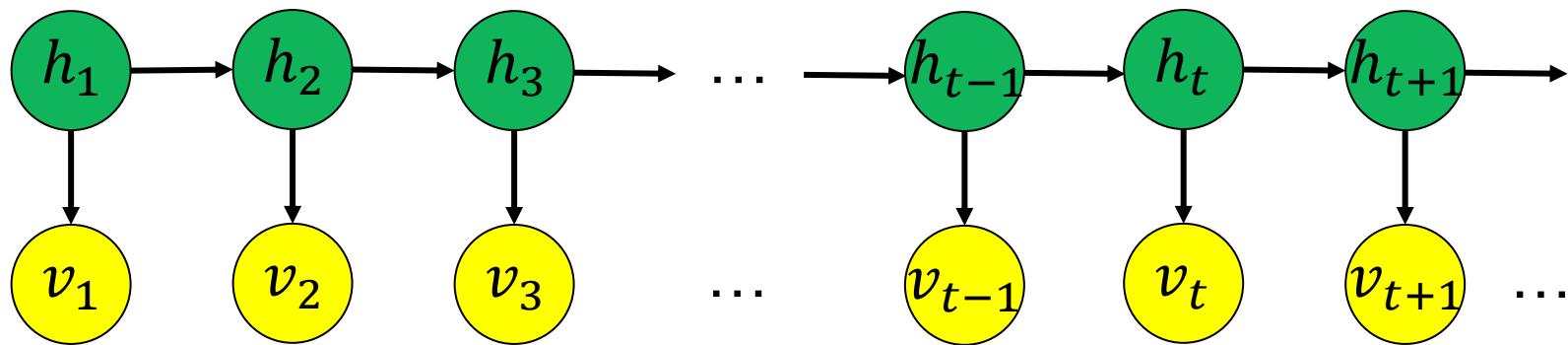
$$\alpha_{t+1}(j) = \sum_{i=1}^S \alpha_t(i) P(h_{t+1} = j | h_t = i) P(v_{t+1} | h_{t+1} = j)$$

$$\beta_t(j) = P(v_{t+1}, v_{t+2}, \dots, v_N | h_t = j)$$



$$\beta_{t-1}(j) = \sum_{i=1}^S \beta_t(i) P(h_t = i | h_{t-1} = j) P(v_t | h_t = i)$$

隐马尔科夫模型：参数的无监督学习



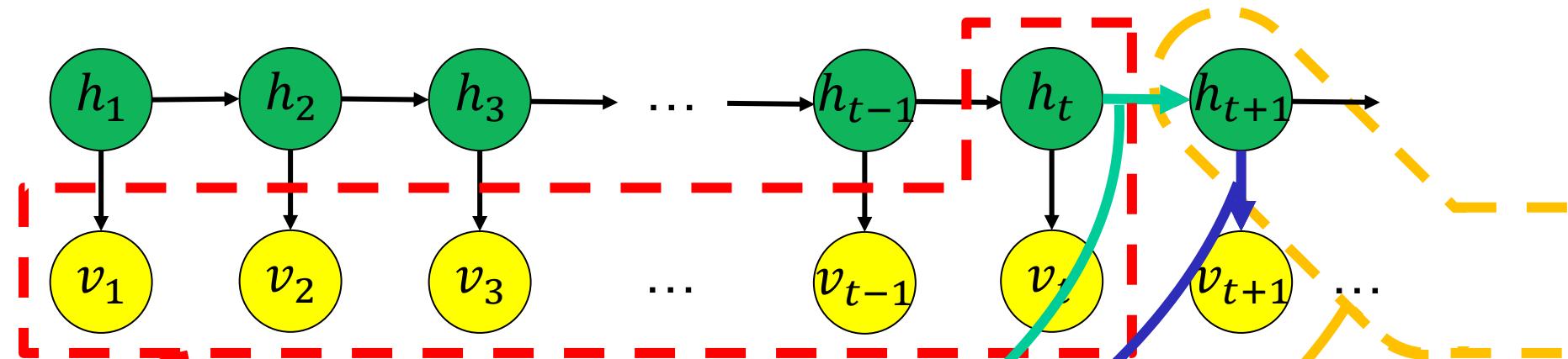
$$\xi_t(i, j) = P(h_t = i, h_{t+1} = j | v_1, v_2, \dots, v_N)$$

$$= \frac{P(h_t = i, h_{t+1} = j, v)}{P(v)}$$

$$\alpha_t(j) = P(v_1, v_2, \dots, v_t, h_t = j)$$

$$\beta_t(j) = P(v_{t+1}, v_{t+2}, \dots, v_N | h_t = j)$$

隐马尔科夫模型：参数的无监督学习



$$\xi_t(i, j) = P(h_t = i, h_{t+1} = j | v_1, v_2, \dots, v_N)$$

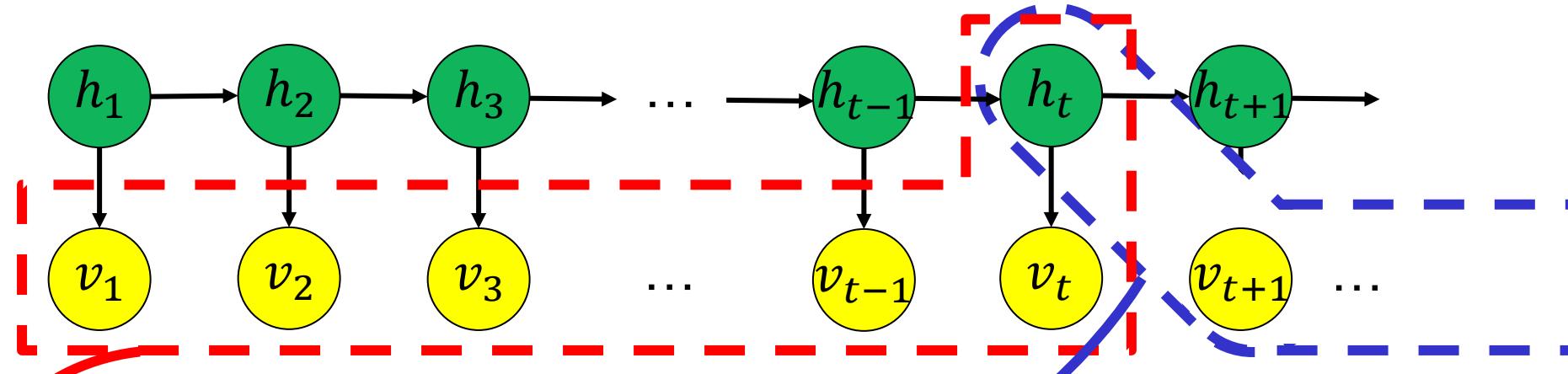
$$= \frac{P(h_t = i, h_{t+1} = j, v)}{P(v)}$$

$$= \frac{\alpha_t(i) a_{ij} b_{j, v_{t+1}} \beta_{t+1}(j)}{P(v)}$$

$$\alpha_t(j) = P(v_1, v_2, \dots, v_t, h_t = j)$$

$$\beta_t(j) = P(v_{t+1}, v_{t+2}, \dots, v_N | h_t = j)$$

隐马尔科夫模型：参数的无监督学习



$$\gamma_t(i) = P(h_t = i | v_1, v_2, \dots, v_N)$$

$$= \frac{P(h_t = i, v)}{P(v)}$$

$$= \frac{\alpha_t(i)\beta_t(i)}{P(v)}$$

$$\alpha_t(j) = P(v_1, v_2, \dots, v_t, h_t = j)$$

$$\beta_t(j) = P(v_{t+1}, v_{t+2}, \dots, v_N | h_t = j)$$

一个例子

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	?		
NNP	0.3	0.03	0.4	0.03	0.2	0.04	0.45	0.05	0.02	0.03	0.06
,	0.01	0.02	0.02	0.9	0.02	0.03	0.3	0.2	0.09		0.35
CD	0.02	0.02	0.02	0.02	0.02	0.9	0.15	0.1	0.5	0.24	0.01
NNS	0.5	0.3	0.05	0.05	0.05	0.05	0.45	0.07	0.03	0.2	0.23
JJ	0.02	0.02	0.02	0.02	0.9	0.02	0.1	0.1	0.3	0.4	0.35

给定： Steve Jobs , 42 years old

α

NNP	0.024	5.64e-3	1.08e-4	3.59e-5	5.47e-6	1.45e-5
,	0.007	1.44e-4	2.80e-3	2.46e-6	6.80e-7	8.27e-7
CD	0.0002	1.10e-4	1.04e-5	1.01e-3	2.14e-6	2.40e-7
NNS	0.0115	2.21e-3	1.51e-5	2.86e-5	1.53e-4	3.71e-7
JJ	0.007	1.30e-4	1.41e-5	5.32e-6	5.05e-6	3.01e-5

β

NNP	9.37e-4	6.65e-3	2.42e-3	5.34e-3	0.128	1
,	9.76e-4	2.44e-4	1.63e-2	2.08e-2	0.1592	1
CD	1.24e-3	2.68e-4	4.38e-3	4.48e-2	0.2732	1
NNS	9.60e-4	3.73e-3	3.20e-3	7.00e-3	0.2779	1
JJ	7.72e-4	1.57e-3	4.45e-3	2.95e-2	0.399	1

一个例子

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	?		
NNP	0.3	0.03	0.4	0.03	0.2	0.04	0.45	0.05	0.02	0.03	0.06
,	0.01	0.02	0.02	0.9	0.02	0.03	0.3	0.4	0.2	0.09	0.35
CD	0.02	0.02	0.02	0.02	0.02	0.9	0.15	0.1	0.5	0.24	0.01
NNS	0.5	0.3	0.05	0.05	0.05	0.05	0.45	0.07	0.03	0.2	0.23
JJ	0.02	0.02	0.02	0.02	0.9	0.02	0.1	0.1	0.3	0.4	0.35

给定： Steve Jobs , 42 years old

NNP	0.489	0.815	0.006	0.004	0.015	0.315
,	0.148	0.001	0.992	0.001	0.002	0.018
CD	0.005	0.001	0.001	0.983	0.013	0.005
NNS	0.24	0.179	0.001	0.004	0.924	0.008
JJ	0.117	0.004	0.001	0.003	0.044	0.654

γ_1

γ_2

γ_3

γ_4

γ_5

γ_6

一个例子

已知：

	jobs	years	steve	old	42	NNP	CD	SNN	Σ	
Noun	0.3	0.03	0.4	0.03	0.2	0.04	0.45	0.05	0.02	0.06
	0.01	0.02	0.02	0.9	0.02	0.03	0.3	0.2	0.09	0.35
	0.02	0.02	0.02	0.02	0.02	0.9	0.15	0.1	0.5	0.01
	0.5	0.3	0.05	0.05	0.05	0.05	0.45	0.07	0.03	0.23
	0.02	0.02	0.02	0.02	0.9	0.02	0.1	0.1	0.3	0.35

给定：Steve Jobs ,

42 years old

ξ_1 [[4.68e-01, 5.73e-04, 1.40e-04, 1.95e-02, 4.91e-04],
[9.11e-02, 3.71e-06, 3.26e-04, 5.67e-02, 4.30e-04],
[1.30e-03, 1.06e-07, 2.33e-06, 4.05e-03, 3.28e-05],
[2.24e-01, 1.52e-04, 9.38e-05, 1.40e-02, 1.57e-03],
[3.04e-02, 3.71e-05, 8.15e-05, 8.51e-02, 1.91e-03]]

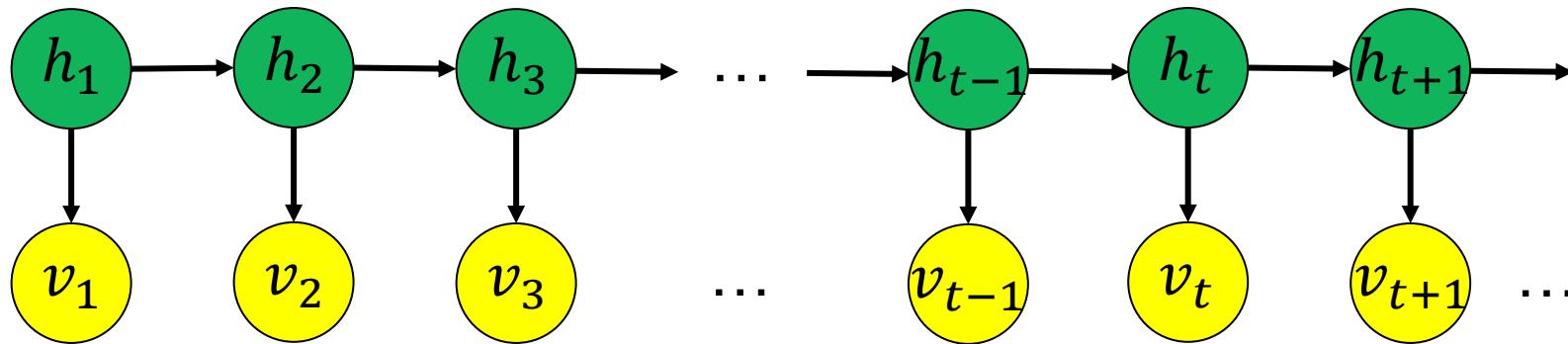
ξ_2 [[4.00e-03, 8.09e-01, 5.37e-04, 3.92e-04, 3.27e-04],
[6.82e-05, 4.59e-04, 1.10e-04, 1.00e-04, 2.51e-05],
[2.60e-05, 3.51e-04, 2.09e-05, 1.91e-04, 5.11e-05],
[1.57e-03, 1.76e-01, 2.95e-04, 2.31e-04, 8.55e-04],
[2.05e-05, 4.14e-03, 2.48e-05, 1.36e-04, 1.01e-04]]

ξ_3 [[2.26e-04, 6.59e-04, 4.73e-03, 1.64e-05, 4.15e-05],
[3.90e-03, 3.80e-04, 9.81e-01, 4.26e-03, 3.23e-03],
[7.24e-06, 1.41e-06, 9.11e-04, 3.96e-05, 3.20e-05],
[3.15e-05, 5.12e-05, 9.26e-04, 3.45e-06, 3.87e-05],
[6.55e-06, 1.91e-05, 1.24e-03, 3.22e-05, 7.23e-05]]

ξ_4 [[1.35e-03, 1.12e-03, 2.13e-04, 1.30e-03, 1.87e-04],
[6.16e-05, 1.70e-06, 1.17e-04, 8.91e-04, 3.84e-05],
[1.26e-02, 6.99e-04, 1.20e-02, 9.15e-01, 4.20e-02],
[1.07e-03, 4.95e-04, 2.38e-04, 1.55e-03, 9.92e-04],
[4.44e-05, 3.68e-05, 6.32e-05, 2.89e-03, 3.69e-04]]

ξ_5 [[1.07e-02, 1.07e-03, 1.19e-04, 1.19e-04, 3.21e-03],
[8.87e-04, 2.96e-06, 1.18e-04, 1.48e-04, 1.20e-03],
[1.40e-03, 9.30e-06, 9.30e-05, 1.16e-03, 1.00e-02],
[2.99e-01, 1.66e-02, 4.66e-03, 4.99e-03, 5.99e-01],
[2.20e-03, 2.20e-04, 2.20e-04, 1.65e-03, 3.95e-02]]

隐马尔科夫模型：参数的无监督学习



$$\widehat{a}_{ij} = P(h_{t+1} = i | h_t = j) = \frac{\#\text{(tags} = (i,j)\text{)}}{\#\text{(tag} = i\text{)}}$$

$$= \frac{\mathbb{E}_{s:i \rightarrow j}}{\mathbb{E}_{s:i}}$$

← 序列s中从状态*i*转移到*j*的次数的数学期望
← 序列s中状态*i*出现次数的数学期望

$$= \frac{\sum_{t=1}^N \xi_t(i, j)}{\sum_{t=1}^N \gamma_t(i)}$$

$$\xi_t(i, j) = P(h_t = i, h_{t+1} = j | v_1, v_2, \dots, v_N)$$

$$\gamma_t(i) = P(h_t = i | v_1, v_2, \dots, v_N)$$

一个例子

已知：

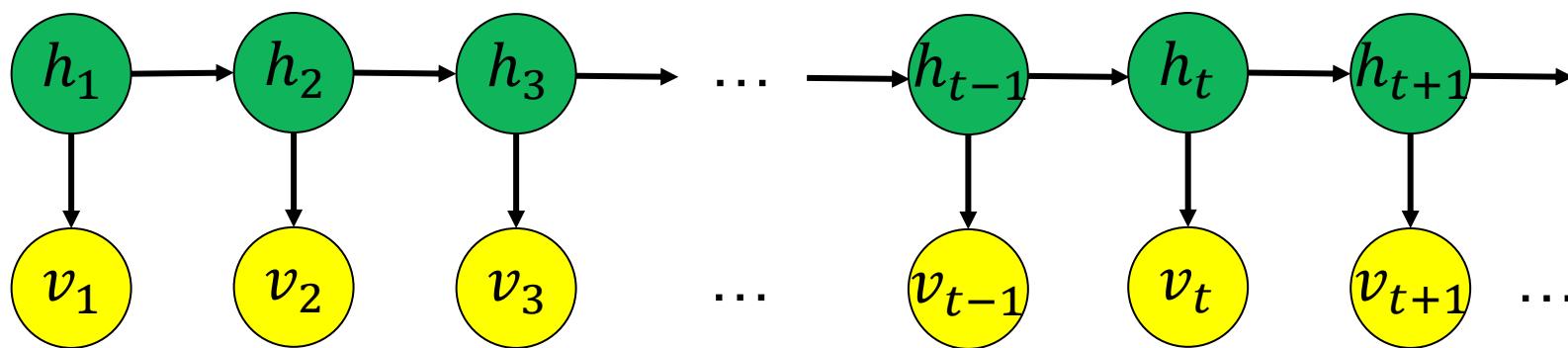
	jobs	years	steve	old	42	NNP	CD	SNN	JJ		
NNP	0.3	0.03	0.4	0.03	0.2	0.04	0.45	0.05	0.02	0.03	0.06
CD	0.01	0.02	0.02	0.9	0.02	0.03	0.3	0.4	0.2	0.09	0.35
NNS	0.02	0.02	0.02	0.02	0.02	0.9	0.15	0.1	0.5	0.24	0.01
JJ	0.5	0.3	0.05	0.05	0.05	0.05	0.45	0.07	0.03	0.2	0.23
	0.02	0.02	0.02	0.02	0.9	0.02	0.1	0.1	0.3	0.4	0.35

给定： Steve Jobs , 42 years old

\widehat{a}_{ij}

$[[2.95e-01, 4.94e-01, 3.49e-03, 1.29e-02, 2.59e-03],$
 $[8.25e-02, 7.29e-04, 8.45e-01, 5.35e-02, 4.23e-03],$
 $[1.52e-02, 1.05e-03, 1.29e-02, 9.13e-01, 5.18e-02],$
 $[3.88e-01, 1.43e-01, 4.58e-03, 1.53e-02, 4.44e-01],$
 $[3.96e-02, 5.41e-03, 1.97e-03, 1.09e-01, 5.09e-02]]$

隐马尔科夫模型：参数的无监督学习



$$\widehat{b}_{ij} = P(v_t = j | h_t = i) = \frac{\#(\text{tag} = i, \text{word} = j)}{\#(\text{tag} = i)}$$

$$= \frac{\mathbb{E}_{s:i \rightarrow j}}{\mathbb{E}_{s:i}}$$

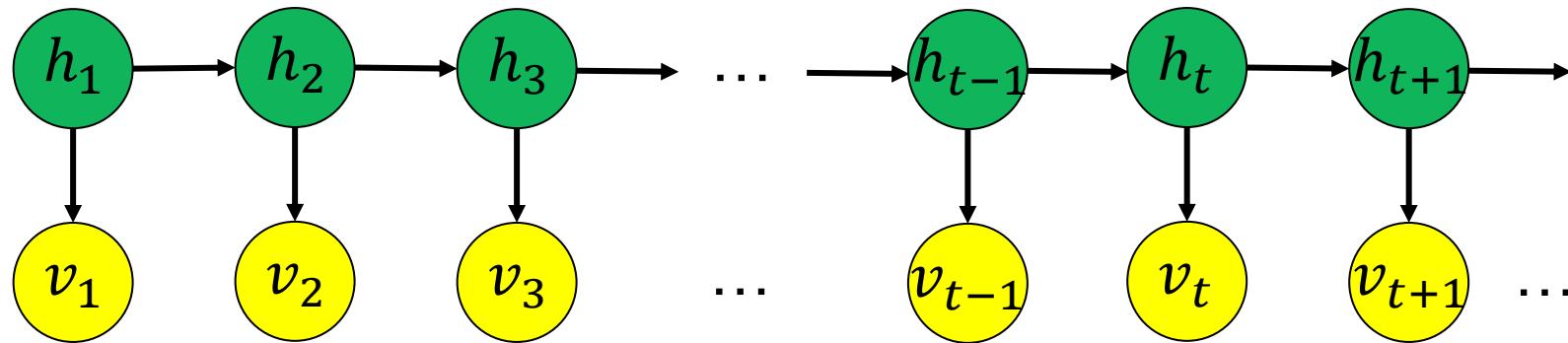
← 序列s中从状态*i*生成第*j*个单词的次数的数学期望
← 序列s中状态*i*出现次数的数学期望

$$= \frac{\sum_{v_t=j} \gamma_t(i)}{\sum_{t=1}^N \gamma_t(i)}$$

$$\xi_t(i, j) = P(h_t = i, h_{t+1} = j | v_1, v_2, \dots, v_N)$$

$$\gamma_t(i) = P(h_t = i | v_1, v_2, \dots, v_N)$$

隐马尔科夫模型：参数的无监督学习



$$\hat{\pi}_i = P(h_1 = i) = \frac{\#(h_1 = i)}{\#(\text{句子})}$$

$= \mathbb{E}_{s:h_1=i}$ ← 序列 s 初始状态为 i 的概率

$$= \gamma_1(i)$$

$$\xi_t(i, j) = P(h_t = i, h_{t+1} = j | v_1, v_2, \dots, v_N)$$

$$\gamma_t(i) = P(h_t = i | v_1, v_2, \dots, v_N)$$

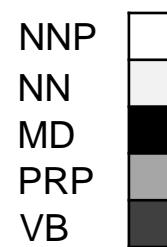
隐马尔科夫模型：参数的无监督学习

	NNP	NN	MD	PRP	VB
NNP	■	□	■	□	■
NN	□	□	□	■	□
MD	□	□	□	■	■
PRP	□	■	■	□	□
VB	■	■	□	□	□

a_{ij}

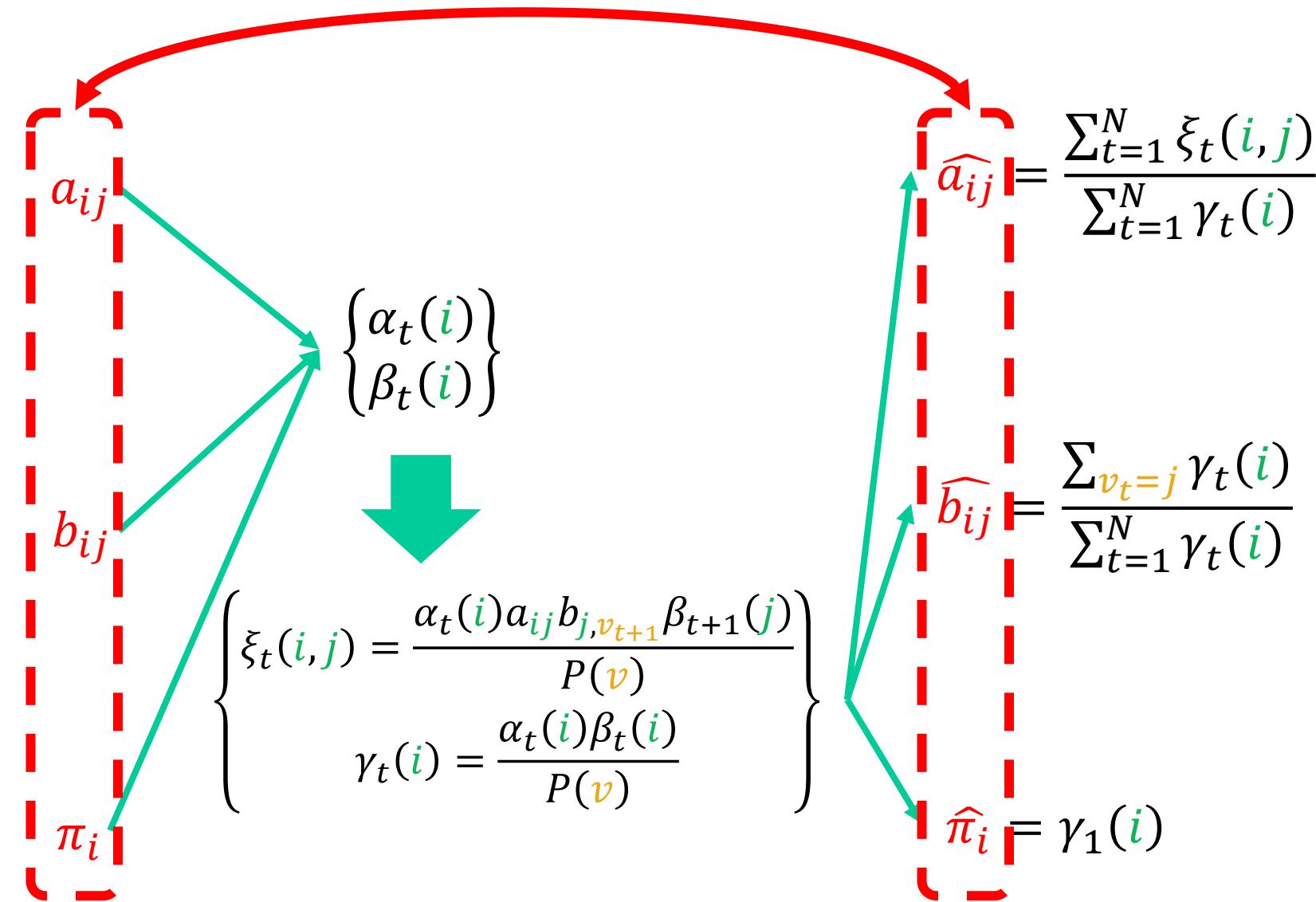
am	NNP	■	□	■	□	■	□	■	□	■	□
is	NN	□	□	□	□	□	□	□	□	□	■
are	MD	□	■	□	□	□	□	□	□	□	□
a	PRP	□	■	■	■	■	□	□	□	□	□
an	VB	■	■	■	■	■	□	□	□	□	□
the											
help											
old											
sugar											
price											
am											
is											
are											
a											
an											
the											
help											
old											
sugar											
price											

b_{ij}



π_i

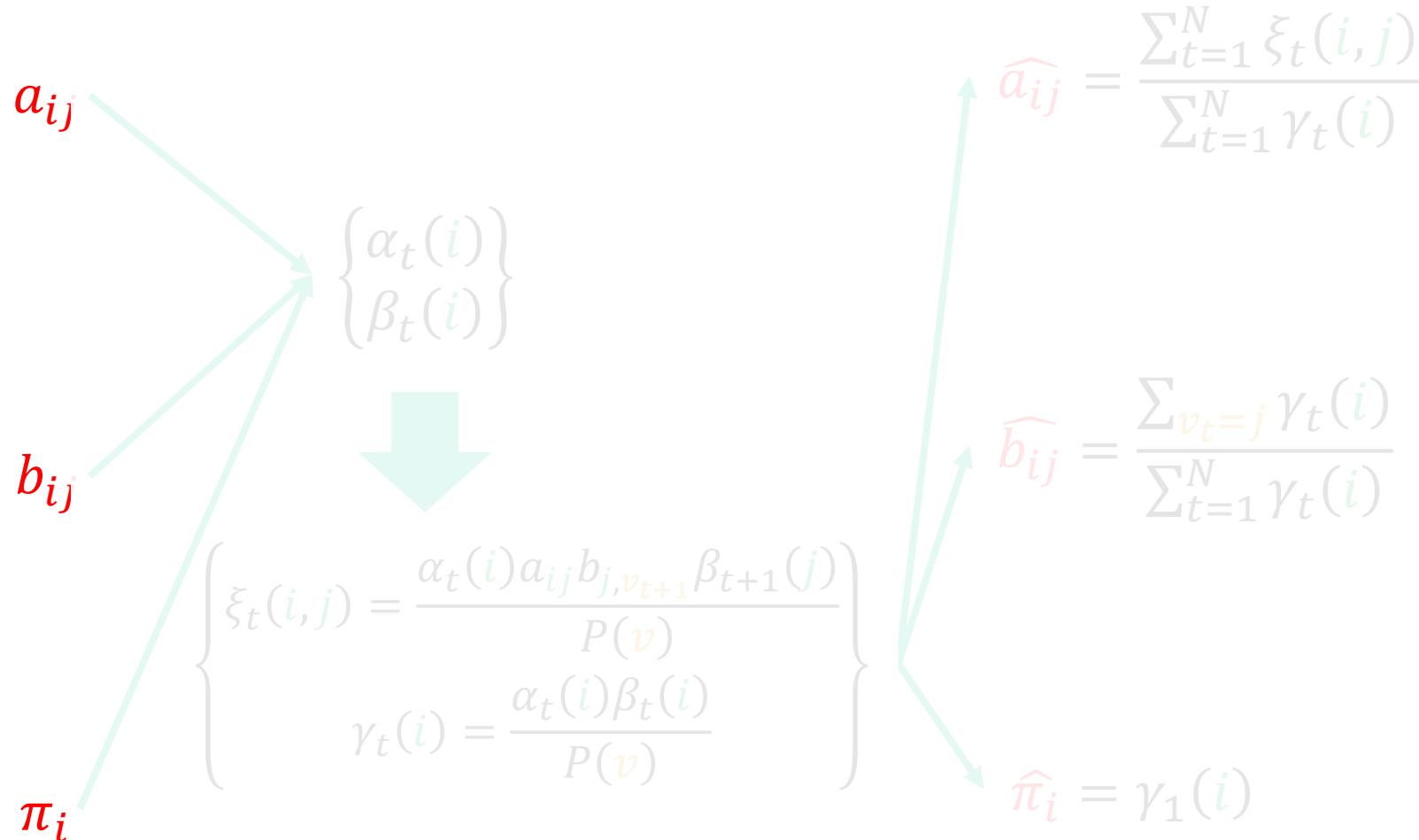
隐马尔科夫模型：参数的无监督学习



隐马尔科夫模型：参数的无监督学习

E步 (expectation) :
给定参数算各种数学期望

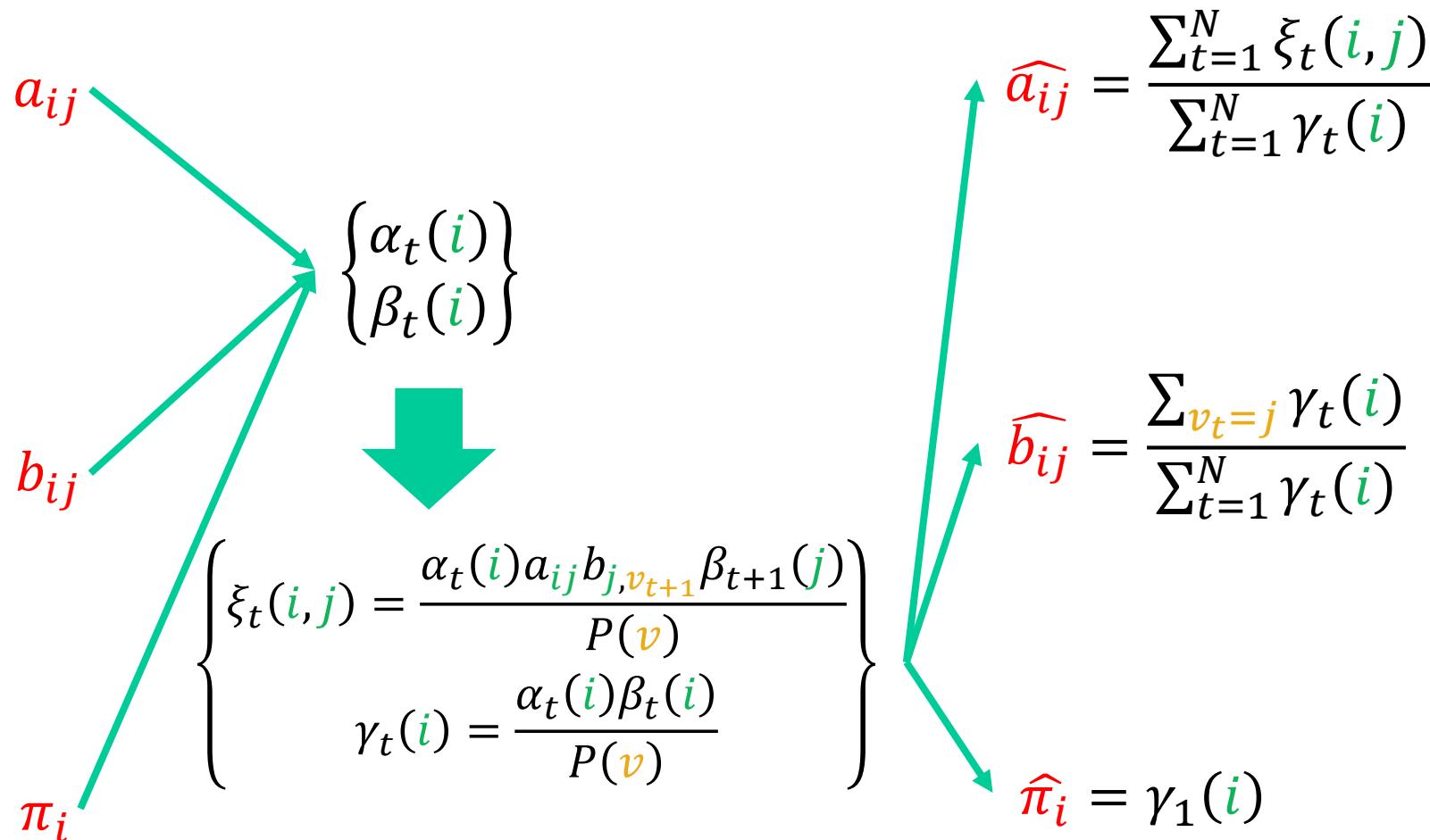
M步 (maximization) :
给定数学期望算各种参数



隐马尔科夫模型：参数的无监督学习

E步 (expectation) :
给定参数算各种数学期望

M步 (maximization) :
给定数学期望算各种参数



隐马尔科夫模型：参数的无监督学习

随机初始化 a_{ij} 、 b_{ij} 、 π_i

不断迭代以下两步直至收敛：

E步 (expectation) : 给定参数算 $\xi_t(i, j)$ 和 $\gamma_t(i)$

$$\left\{ \begin{array}{l} \xi_t(i, j) = \frac{\alpha_t(i) a_{ij} b_{j,v_{t+1}} \beta_{t+1}(j)}{P(v)} \\ \gamma_t(i) = \frac{\alpha_t(i) \beta_t(i)}{P(v)} \end{array} \right\}$$

M步 (maximization) : 给定数学期望算参数的新估值

$$\widehat{a}_{ij} = \frac{\sum_{t=1}^N \xi_t(i, j)}{\sum_{t=1}^N \gamma_t(i)}$$

$$\widehat{b}_{ij} = \frac{\sum_{v_t=j} \gamma_t(i)}{\sum_{t=1}^N \gamma_t(i)} \quad \widehat{\pi}_i = \gamma_1(i)$$

将新的估值赋值给 a_{ij} 、 b_{ij} 、 π_i

这就是前向-后向算法，也叫 Baum-Welch 算法

隐马尔科夫模型：（超纲）思考题

- 为什么这样迭代的算法是正确的？
- 为什么算法会收敛？
- 能保证找到全局最优解么？

隐马尔科夫模型：一些实战经验（以WSJ数据集为例）

有监督的POS tagging模型：

不顾上下文盲猜某个单词最常见的词性 HMM (Brants, 2000)	90%~94%
	96.5%

无监督的前向-后向算法：

HMM + Baum-Welch (Johnson, 2007) ~40%

词性标注工具推荐

- ▶ **StanfordCoreNLP**: 斯坦福的NLP核心包。具备各种nlp功能，包括词性标注。
Github地址: <https://github.com/Lynten/stanford-corenlp>
官网: <https://stanfordnlp.github.io/CoreNLP/>
- ▶ **NLTK**: NLTK是一个高效的Python构建的平台,用来处理人类自然语言数据。
Github地址: <https://github.com/nltk/nltk>
官网: <http://www.nltk.org/>
- ▶ **SpaCy**: 工业级的自然语言处理工具，遗憾的是不支持中文。
Github地址: <https://github.com/explosion/spaCy>
官网: <https://spacy.io/>
- ▶ **Jieba**: “结巴” 中文分词：做最好的 Python 中文分词组件，可以进行词性标注。
Github地址: <https://github.com/fxsjy/jieba>
- ▶ **SnowNLP**: SnowNLP是一个python写的类库，可以方便的处理中文文本内容。
Github地址: <https://github.com/isnowfy/snownlp>
- ▶ **THULAC**: THULAC (THU Lexical Analyzer for Chinese) 由清华大学自然语言处理与社会人文计算实验室研制推出的一套中文词法分析工具包，具有中文分词和词性标注功能。
Github地址: <https://github.com/thunlp/THULAC>
官网: <http://thulac.thunlp.org/>
- ▶ **Hanlp**: HanLP是一系列模型与算法组成的NLP工具包，由大快搜索主导并完全开源，目标是普及自然语言处理在生产环境中的应用。
Github地址: <https://github.com/hankcs/pyhanlp>
官网: <http://hanlp.linrunsoft.com/>