2.4 词性标注及隐马尔科夫模型

林洲汉
上海交大电院

2024年秋季学期

目录

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?

用人话说。。。

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。

什么是词性标注?

词性划分具有层次性

- 词
  - 实词
    - 名词
    - 代词
    - 动词
      - 能愿动词
      - 趋向动词
      - ...
    - 形容词
    - 数词
    - 量词
  - 虚词

在你们实际将接触到的数据集中，
1. 词性的类别数目远超这几类，
2. 不同数据集归纳出来的类别集合也不一样

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*	WP$	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
(	opening parenthesis
)	closing parenthesis
*	negator
,	comma
–	dash
.	sentence terminator
:	colon
ABL	pre-qualifier
ABN	pre-quantifier
ABX	pre-quantifier, double conjunction
AP	post-determiner
AT	article
BE/BED/BEDZ/BEG/BEM/BEN/BER/BEZ	
CC	coordinating conjunction
CD	cardinal numeral
CS	subordinating conjunction
DO/DOD/DOZ	
DT	singular determiner,
DTI	singular or plural determiner
DTS	plural determiner
DTX	determiner, double conjunction
EX	existential there
HV/HVD/HVG/HVN/HVZ	
IN	preposition
JJ	adjective
JJR	comparative adjective
JJS	semantically superlative adj.
JJT	morphologically superlative adj.
MD	modal auxiliary
NN	(common) singular or mass noun
NN$	possessive singular common noun
NNS	plural common noun
NNS$	possessive plural noun
NP	singular proper noun
NP$	possessive singular proper noun

Tag	Description
RP	adverb or particle
TO	infinitive marker
UH	interjection, exclamation
VB	verb, base form
VBD	verb, past tense
VBG	verb, present participle, gerund
VBN	verb, past participle
VBZ	verb, 3rd singular present
WDT	wh- determiner
WP$	possessive wh- pronoun
WPO	objective wh- pronoun
WPS	nominative wh- pronoun
WQL	how
WRB	wh- adverb
NPS	plural proper noun
NPS$	possessive plural proper noun
NR	adverbial noun
NR$	possessive adverbial noun
NRS	plural adverbial noun
OD	ordinal numeral
PN	nominal pronoun
PN$	possessive nominal pronoun
PP$	possessive personal pronoun
PP$$	second possessive personal pronoun
PPL	singular reflexive personal pronoun
PPLS	plural reflexive pronoun
PPO	objective personal pronoun
PPS	3rd. sg. nominative pronoun
PPSS	other nominative pronoun
QL	qualifier
QLP	post-qualifier
RB	adverb
RBR	comparative adverb
RBT	superlative adverb
RN	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

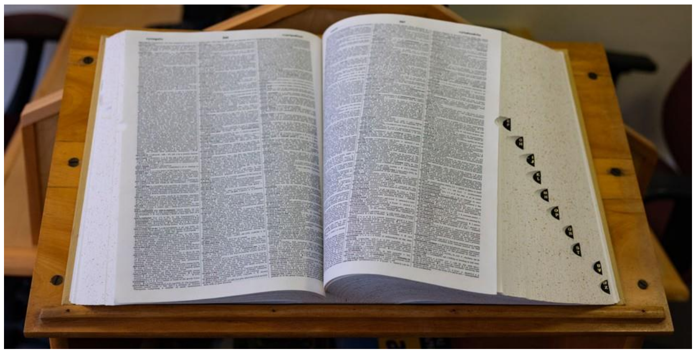
- **基于规则的词性标注方法**
  - 基于规则的词性标注方法是人们提出较早的一种词性标注方法，其基本思想是按兼类词搭配关系和上下文语境建造词类消歧规则。早期的词类标注规则一般由人工构建。
  - 随着标注语料库规模的增大，可利用的资源也变得越来越多，这时候以人工提取规则的方法显然变得不现实，于是乎，人们提出了基于机器学习的规则自动提出方法。
- **基于统计模型的词性标注方法**
  - 统计方法将词性标注看作是一个序列标注问题。其基本思想是：给定带有各自标注的词的序列，我们可以确定下一个词最可能的词性。
  - 现在已经有隐马尔可夫模型（HMM）或条件随机场（CRF）等统计模型了，这些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。

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。

目录

- 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



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*	WP$	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).

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 / may I）
- 动名词 VB -> NN
- VB -> NNP

二阶关联
- 年龄短语 CD -> NNS/NN -> JJ

… …

…

… …

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

NNP → NNP → , → CD → NNS → JJ → , → MD → VB → NNP

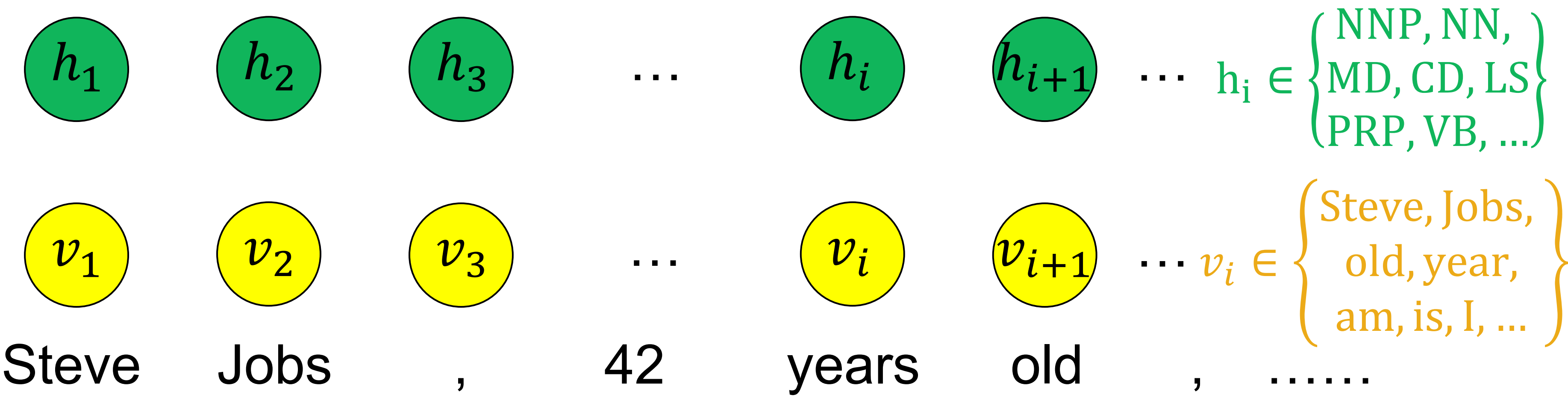
Steve	Jobs	,	42	years	old	,	will	join	Apple	…
								create	NeXT	
								fund	Pixar	
Donald	Trump		71					be	POA	

……

- 我们只知道单词，不知道他们背后的词性
- 我们知道词性之间组织起来的规则
- 对于同一个词性，有很多单词与之对应，形成很多语法正确的句子。

隐马尔科夫模型（Hidden Markov Model），
用概率的语言统一实现了这一切。

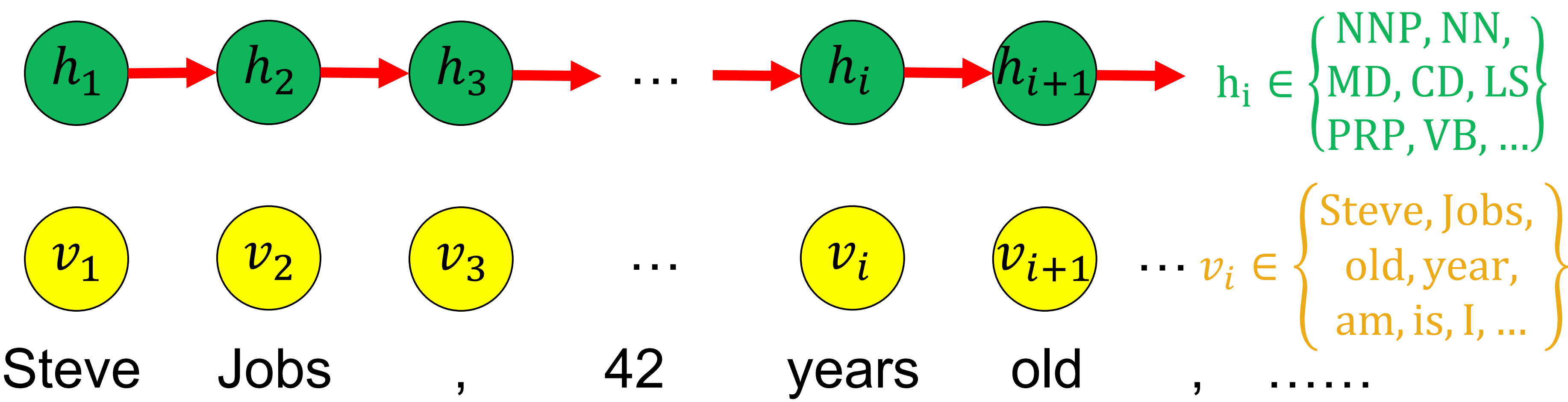
隐马尔科夫模型



- **我们只知道单词，不知道他们背后的词性**
- 我们知道词性之间组织起来的规则
- 对于同一个词性，有很多单词与之对应，形成很多语法正确的句子。

隐马尔科夫模型（Hidden Markov Model），
用概率的语言统一实现了这一切。

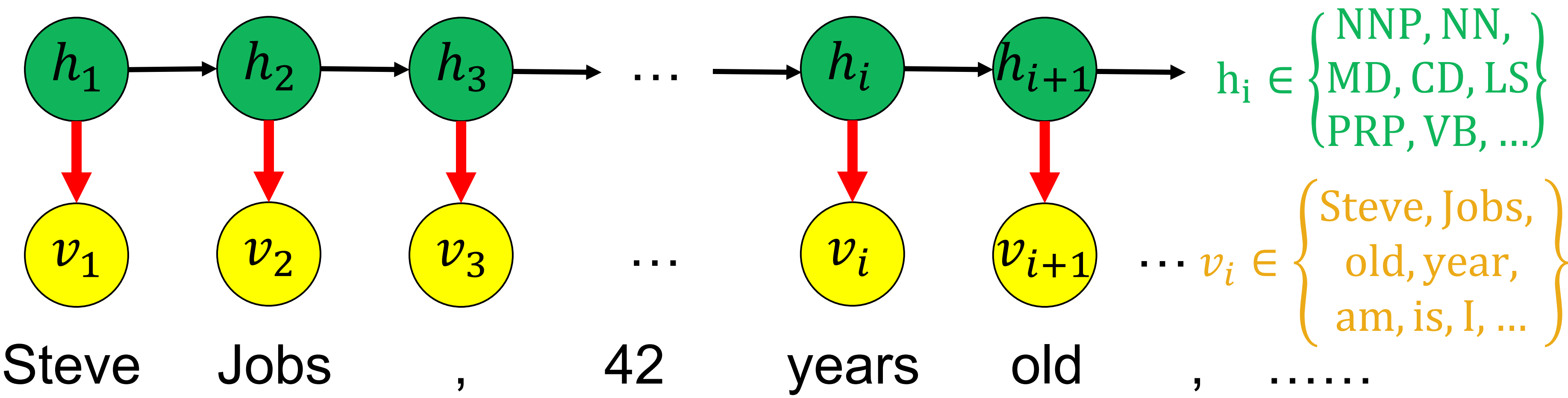
隐马尔科夫模型



- 我们只知道单词，不知道他们背后的词性
- 我们知道词性之间组织起来的规则
- 对于同一个词性，有很多单词与之对应，形成很多语法正确的句子。

隐马尔科夫模型（Hidden Markov Model），
用概率的语言统一实现了这一切。

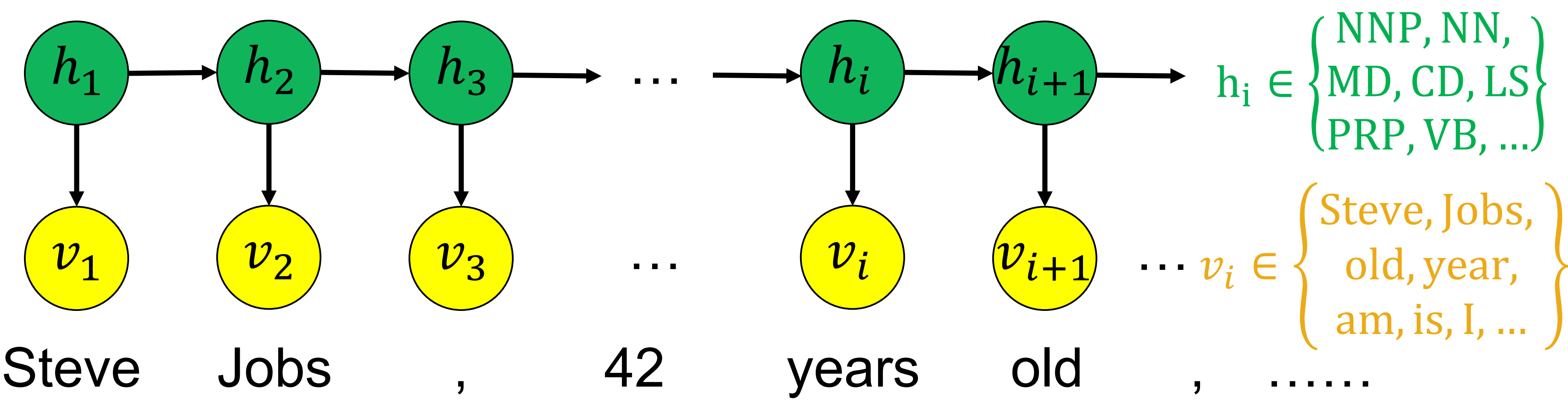
隐马尔科夫模型



- 我们只知道单词，不知道他们背后的词性
- 我们知道词性之间组织起来的规则
- 对于同一个词性，有很多单词与之对应，形成很多语法正确的句子。

隐马尔科夫模型（Hidden Markov Model），
用概率的语言统一实现了这一切。

隐马尔科夫模型



已知句子中的单词，预测每个单词的词性，变成了这样一个任务：

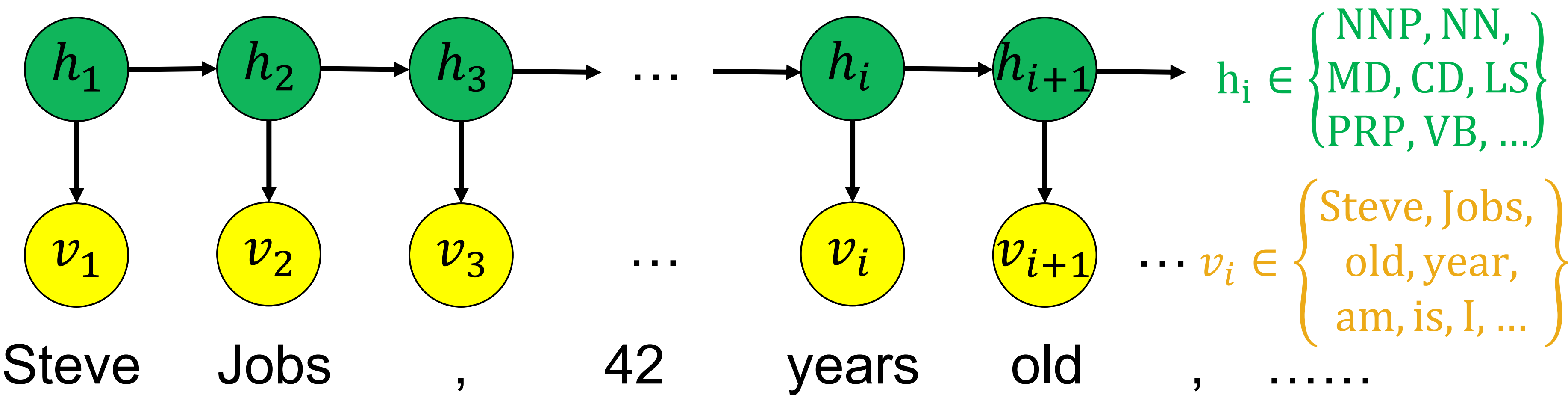
$$\underset{h_1,h_2,\ldots,h_i,\ldots,h_N}{\operatorname{argmax}} P(h_1, h_2, \ldots, h_i, \ldots, h_N \mid v_1, v_2, \ldots, v_i, \ldots, v_N)$$

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

$$P(v,\ h)$$

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

隐马尔科夫模型



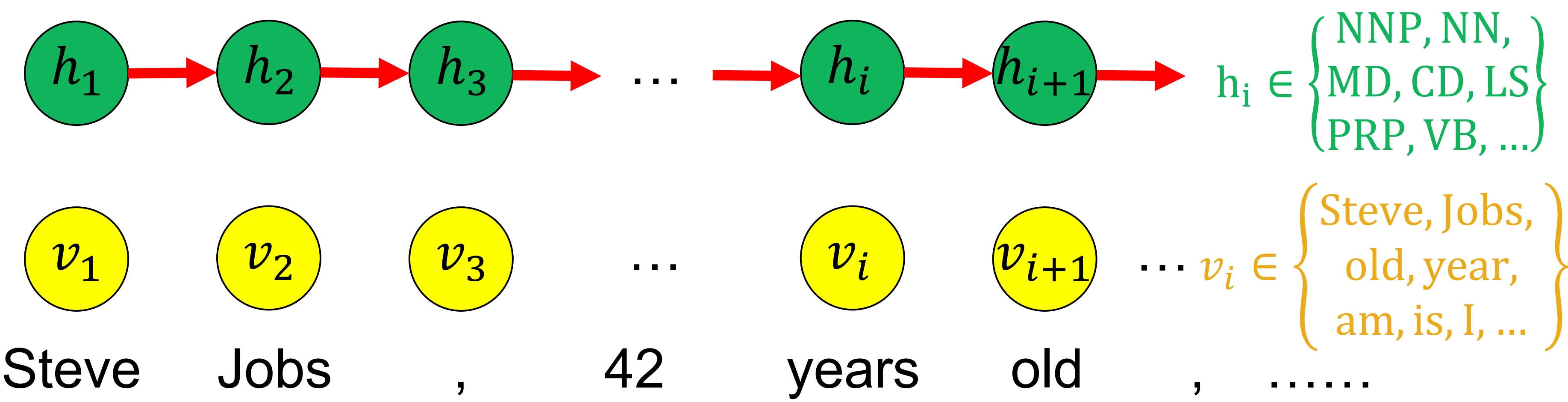
已知句子中的单词，预测每个单词的词性，变成了这样一个任务：

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

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

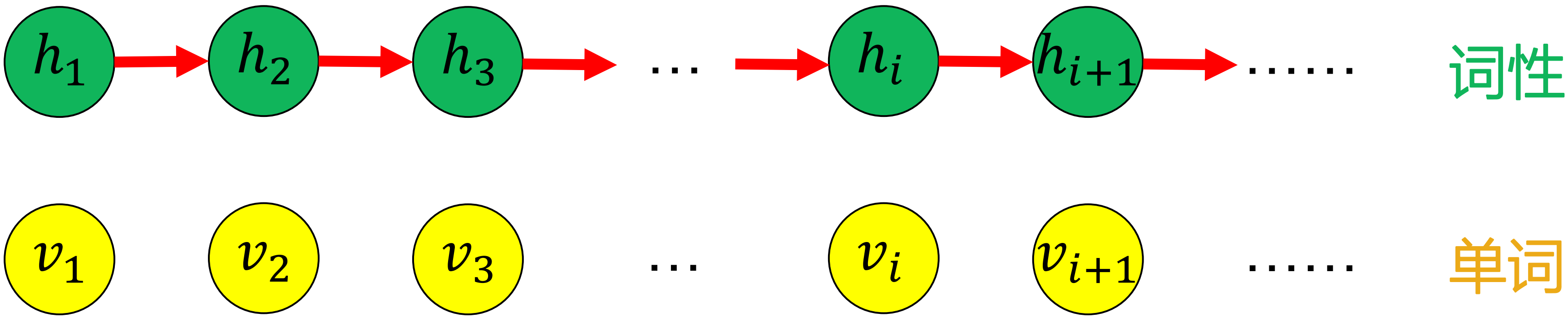
$$= \underset{h}{\operatorname{argmax}} \frac{P(v|h)P(h)}{P(v)} \qquad = \underset{h}{\operatorname{argmax}} \boxed{P(v|h)} \cdot \boxed{P(h)}$$

隐马尔科夫模型



$$\boxed{P(h)}$$

隐马尔科夫模型



Steve Jobs , 42 years old , ……

规则:

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

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

… ⋮

二阶关联
- 年龄短语 — CD -> NNS/NN -> JJ

…… ……

…

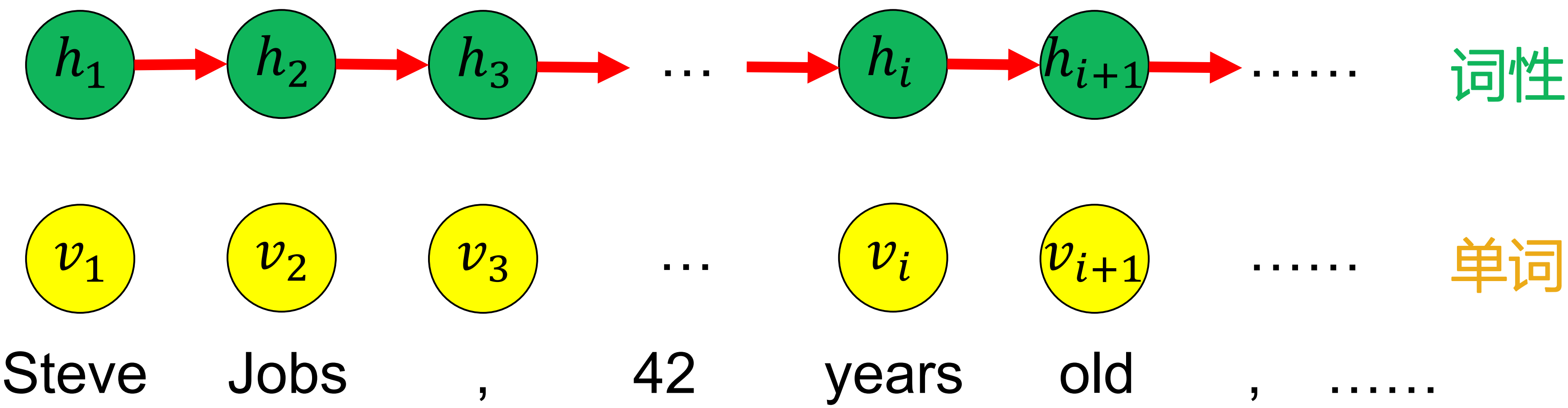
…… ……

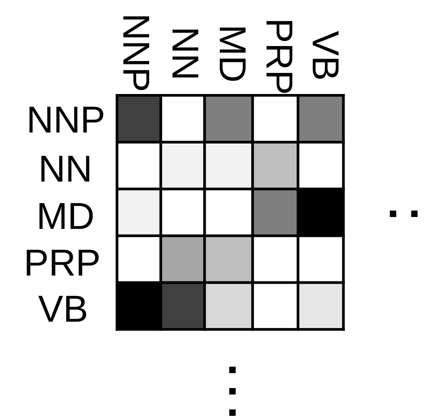
■ 存在该关系

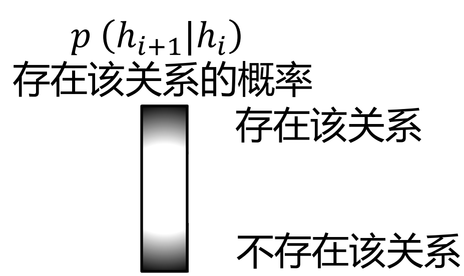
□ 不存在该关系

$\boxed{P(h)}$

隐马尔科夫模型

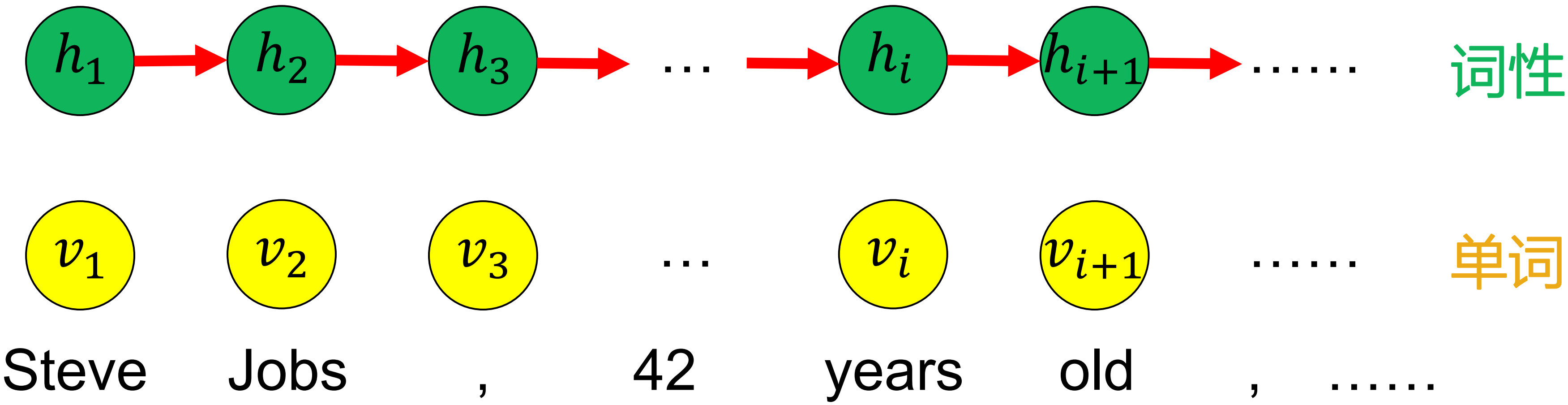






$$\boxed{P(h)}$$

隐马尔科夫模型

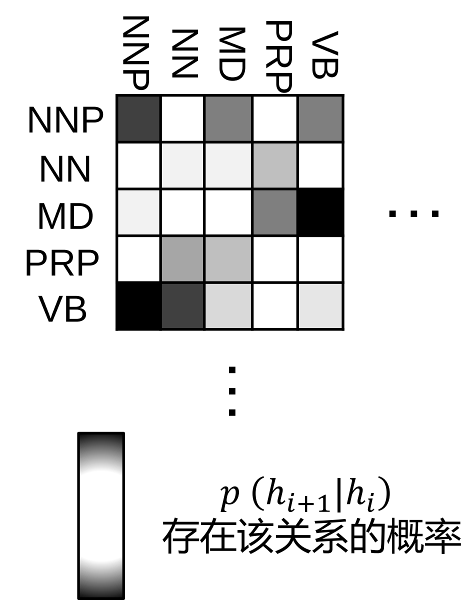


$$\boxed{P(h)}$$

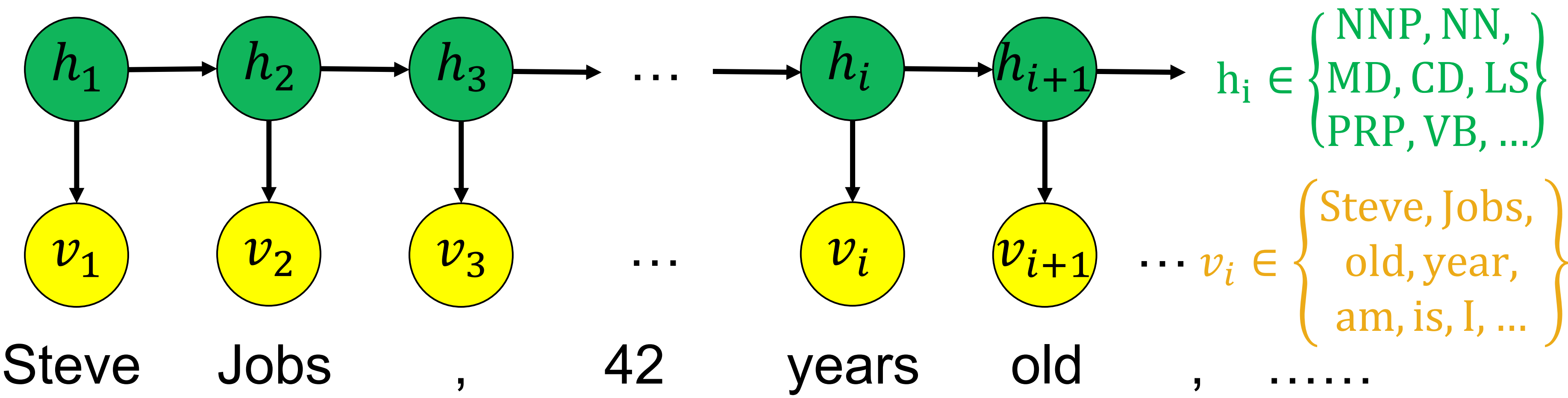
$$= P(h_1, h_2, \dots, h_i, \dots, h_N)$$

$$= P(h_1)P(h_2|h_1)\dots P(h_{i+1}|h_i)\dots P(h_N|h_{N-1})$$

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



隐马尔科夫模型



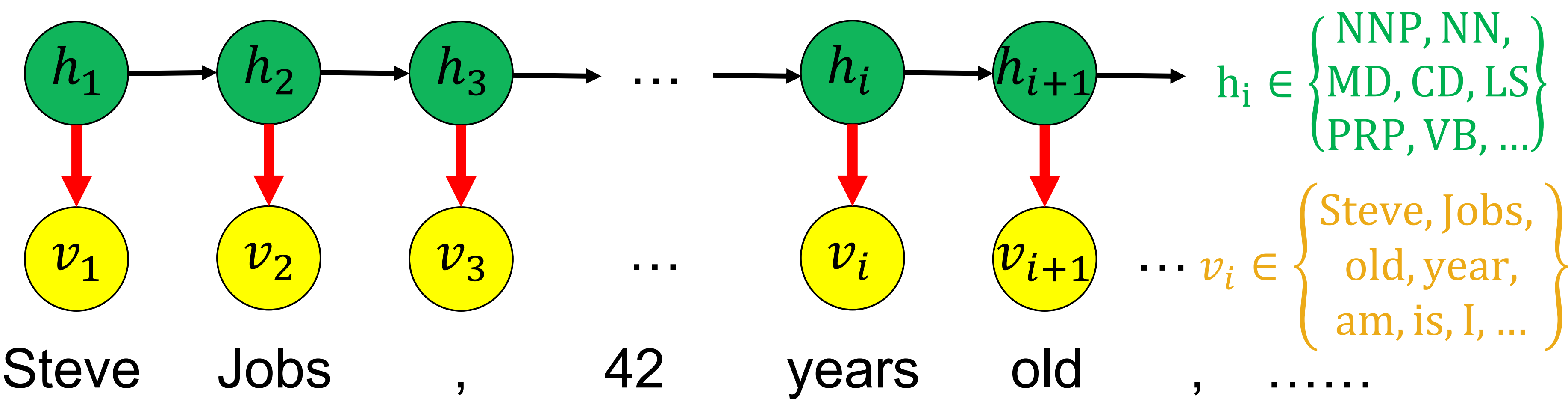
已知句子中的单词，预测每个单词的词性，变成了这样一个任务：

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

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

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

隐马尔科夫模型



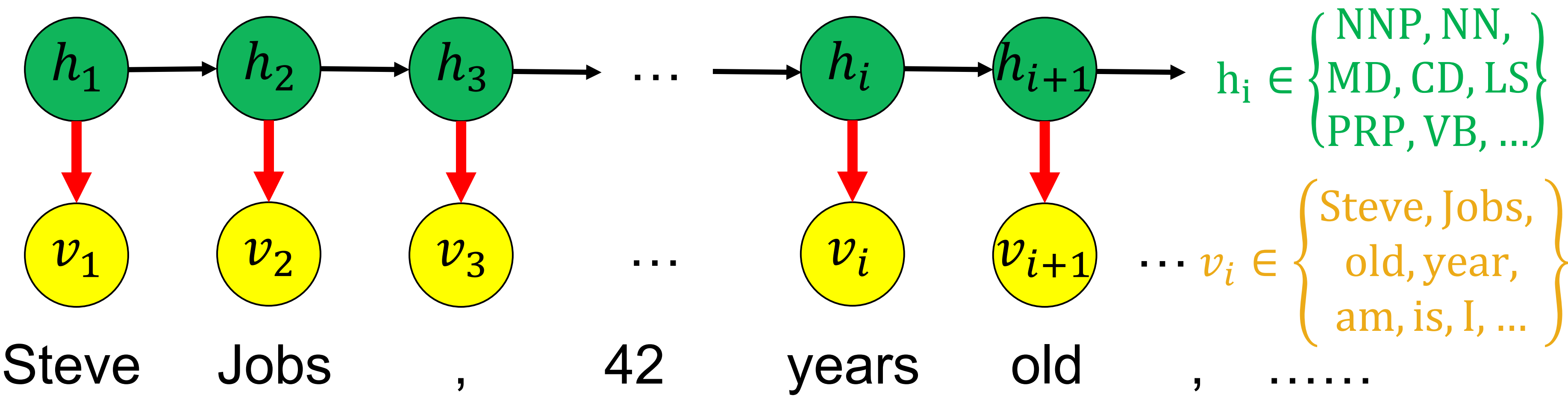
已知句子中的单词，预测每个单词的词性，变成了这样一个任务：

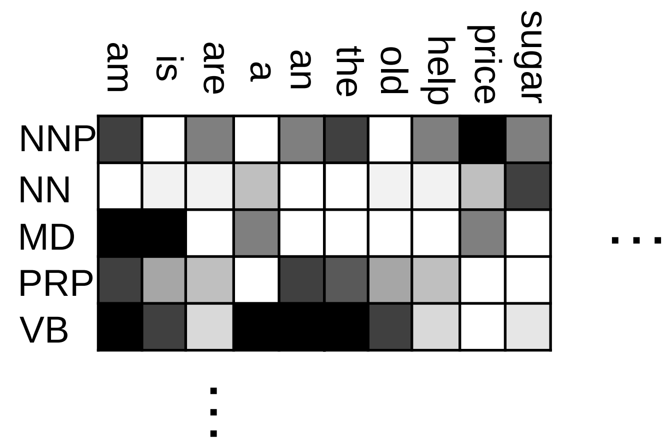
$$\underset{h_1,h_2,\ldots,h_i,\ldots,h_N}{\operatorname{argmax}} P(h_1, h_2, \ldots, h_i, \ldots, h_N \mid v_1, v_2, \ldots, v_i, \ldots, v_N)$$

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

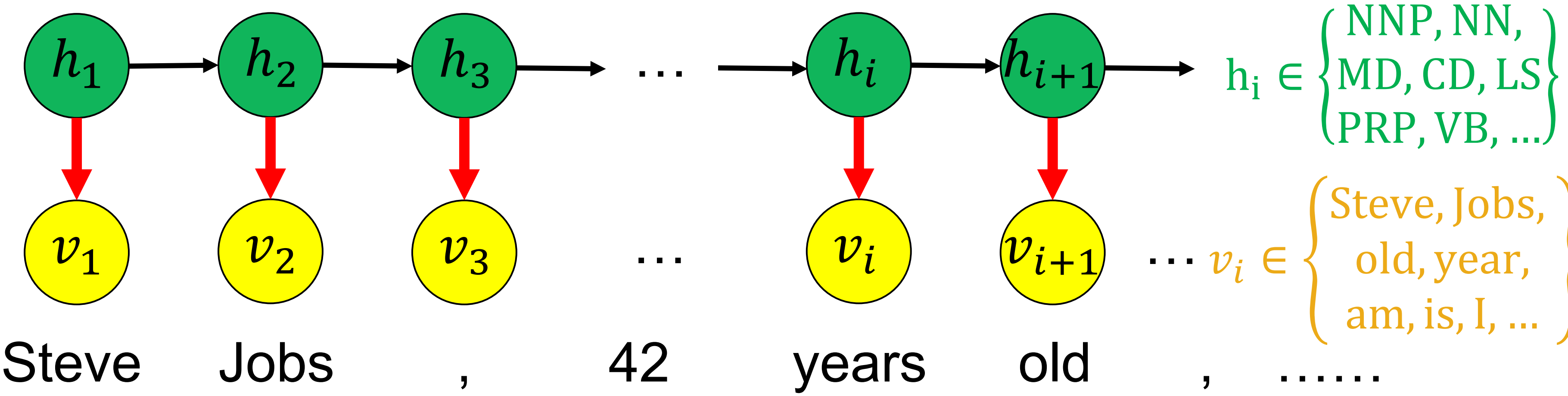
$$= \underset{h}{\operatorname{argmax}} \frac{P(v|h)P(h)}{P(v)} \qquad = \underset{h}{\operatorname{argmax}} \boxed{P(v|h)} \cdot P(h)$$

隐马尔科夫模型





$$\boxed{P(v|h)}$$

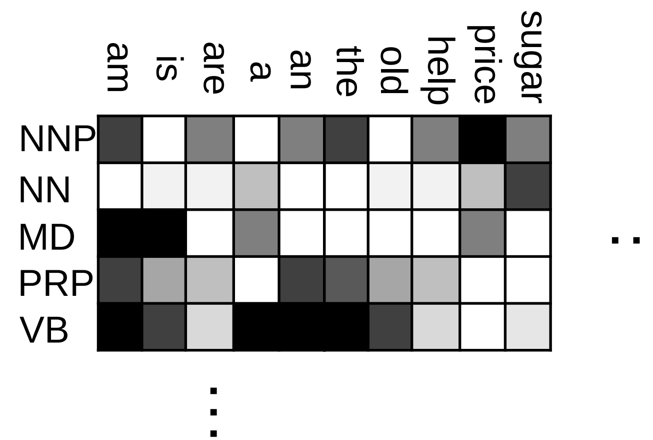


$$\boxed{P(v|h)}$$

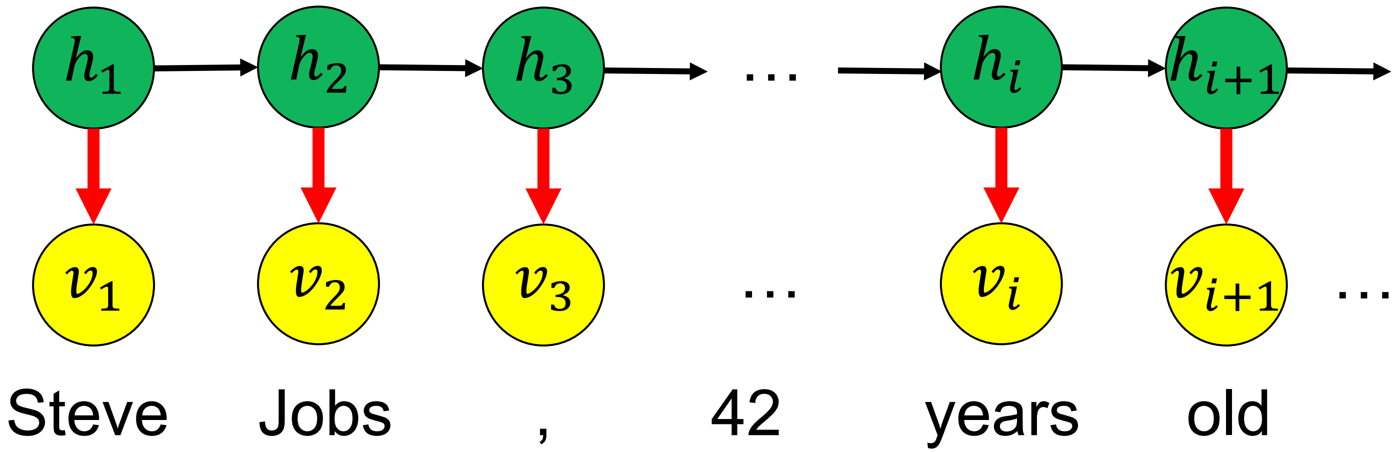
$$= P(v_1, v_2, \ldots, v_i, \ldots, v_N \mid h_1, h_2, \ldots, h_i, \ldots, h_N)$$

$$= P(v_1|h_1)P(v_2|h_2) \ldots P(v_i|h_i) \ldots P(v_N|h_N)$$

$$= \prod_{j=1}^{N} P(v_j|h_j)$$



隐马尔科夫模型



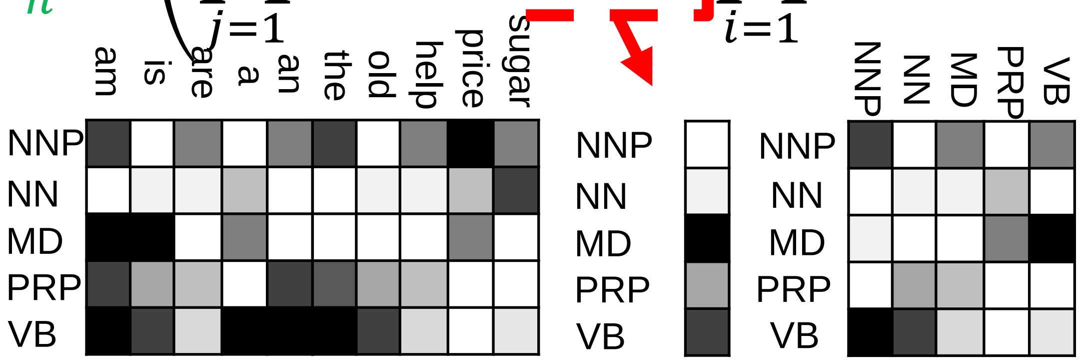
$h_i \in \{\text{NNP, NN, MD, CD, LS, PRP, VB, …}\}$

$\cdots\ v_i \in \{\text{Steve, Jobs, old, year, am, is, I, …}\}$

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

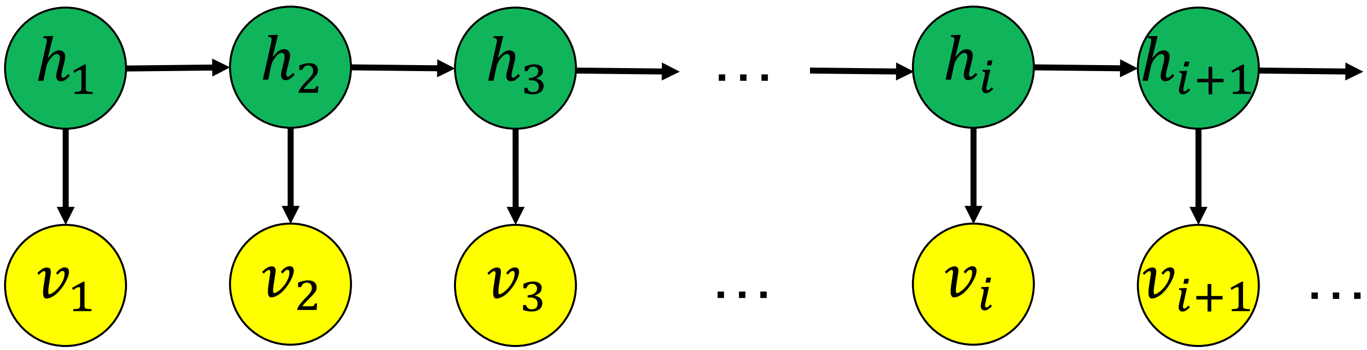
$$= \underset{h}{\operatorname{argmax}} \left(\prod_{j=1}^{N} P(v_j|h_j) \boxed{P(h_1)} \prod_{i=1}^{N-1} P(h_{i+1}|h_i) \right)$$

参数:



…

隐马尔科夫模型



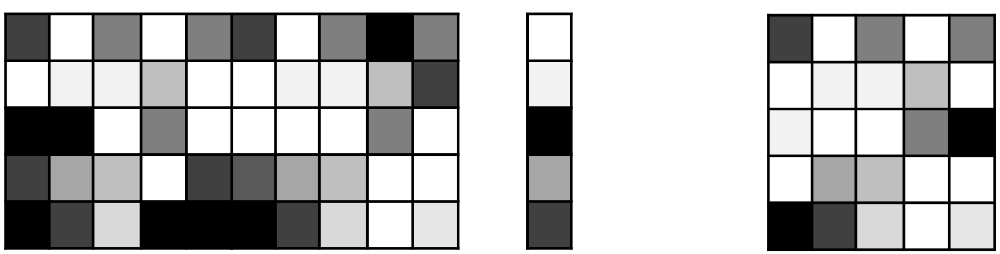
隐马尔科夫模型的定义

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

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

依照左式对联合概率分布$P(v, h)$ 建模，即被称为HMM模型。

参数:



v, h都应是离散随机变量的离散序列。

小测验

已知：

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

	NNP	,	CD	NNS	JJ
NNP	0.2	0.2	0.01	0.01	0.01
,	0.2	0	0.3	0.2	0.1
CD	0.15	0.01	0.1	0.4	0.2
NNS	0.2	0.1	0.01	0.01	0.1
JJ	0.1	0.1	0.1	0.2	0.4

NNP	0.05
,	0.3
CD	0.01
NNS	0.2
JJ	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**

$$P(h,v) = P(\mathbf{NNP},\mathbf{NNP},\mathbf{,},\mathbf{CD},\mathbf{NNS},\mathbf{JJ},\text{Steve},\text{Jobs},\text{,},\text{42},\text{years},\text{old})$$

$$= P(\mathbf{NNP})P(\mathbf{NNP}|\mathbf{NNP})P(\mathbf{,}|\mathbf{NNP})\,P(\mathbf{CD}|\mathbf{,})\,P(\mathbf{NNS}|\mathbf{CD})P(\mathbf{JJ}|\mathbf{NNS})$$

$$P(\text{Steve}|\mathbf{NNP})P(\text{Jobs}|\mathbf{NNP})\,P(\text{,}|\mathbf{,})\,P(\text{42}|\mathbf{CD})P(\text{years}|\mathbf{NNS})\,P(\text{old}|\mathbf{JJ})$$

$$= 0.05\times 0.2\times 0.2\times 0.3\times 0.4\times 0.1$$

$$\times 0.1\times 0.05\times 0.7\times 0.9\times 0.1\times 0.4$$

$$= 0.000000003024$$

已知：

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

	NNP	,	CD	NNS	JJ
NNP	0.2	0.2	0.01	0.01	0.01
,	0.2	0	0.3	0.2	0.1
CD	0.15	0.01	0.1	0.4	0.2
NNS	0.2	0.1	0.01	0.01	0.1
JJ	0.1	0.1	0.1	0.2	0.4

NNP	0.05
,	0.3
CD	0.01
NNS	0.2
JJ	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**

$$P(h,v) = P(\textbf{NNP},\textbf{NNP},\textbf{,},\textbf{CD},\textbf{NNS},\textbf{JJ},\text{Steve},\text{Jobs},\text{,},\text{42},\text{years},\text{old})$$

$$= P(\textbf{NNP})P(\textbf{NNS}|\textbf{NNP})P(\textbf{,}|\textbf{NNS})\,P(\textbf{CD}|\textbf{,})\,P(\textbf{NNS}|\textbf{CD})P(\textbf{JJ}|\textbf{NNS})$$
$$P(\text{Steve}|\textbf{NNP})P(\text{Jobs}|\textbf{NNS})\,P(\text{,}|\textbf{,})\,P(\text{42}|\textbf{CD})P(\text{years}|\textbf{NNS})\,P(\text{old}|\textbf{JJ})$$

$$= 0.05 \times 0.01 \times 0.1 \times 0.3 \times 0.4 \times 0.1$$
$$\times 0.1 \times 0.2 \times 0.7 \times 0.9 \times 0.1 \times 0.4$$

$$= 0.000000003024$$

隐马尔科夫模型

已知：

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

	NNP	,	CD	NNS	JJ
NNP	0.2	0.2	0.01	0.01	0.01
,	0.2	0	0.3	0.2	0.1
CD	0.15	0.01	0.1	0.4	0.2
NNS	0.2	0.1	0.01	0.01	0.1
JJ	0.1	0.1	0.1	0.2	0.4

NNP	0.05
,	0.3
CD	0.01
NNS	0.2
JJ	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**

$$\log P(h_1, v) = -19.62 \qquad P(h_1, v) = 0.000000003024$$

对应POS为：**NNP NNS , CD NNS JJ**

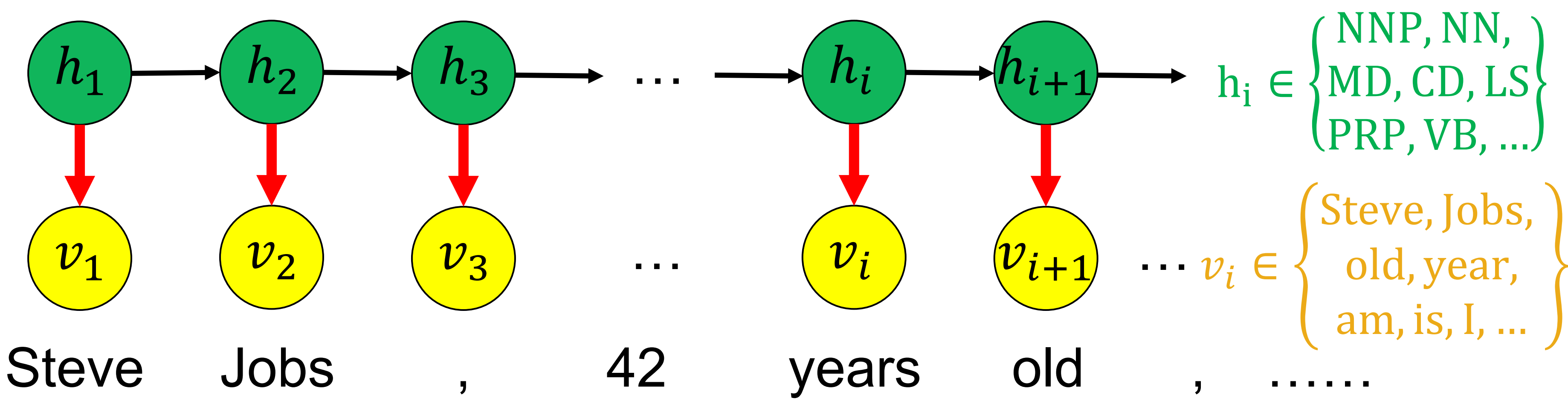
$$\log P(h_2, v) = -21.92 \qquad P(h_2, v) = 0.0000000003024$$

$$\log(MN) = \log M + \log N$$

目录

- 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

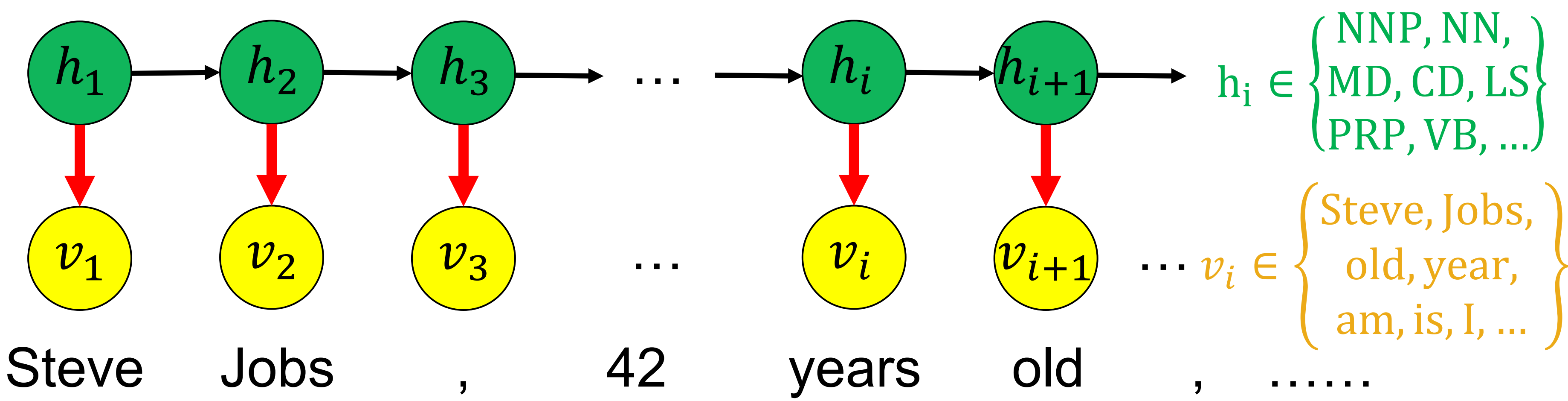
隐马尔科夫模型



$$P(h_1, v)$$

$$P(h_2, v)$$

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。

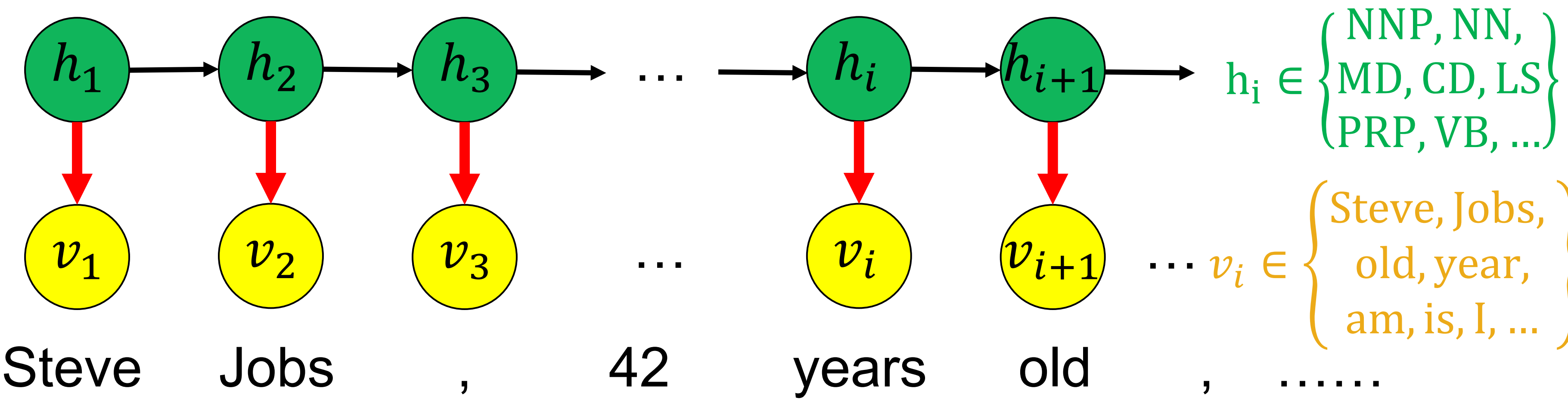


$$\boxed{\underset{h}{\operatorname{argmax}}} P(h|v) = \underset{h}{\operatorname{argmax}} P(h, v)$$

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$\underset{h}{\operatorname{argmax}}(\cdot)$，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

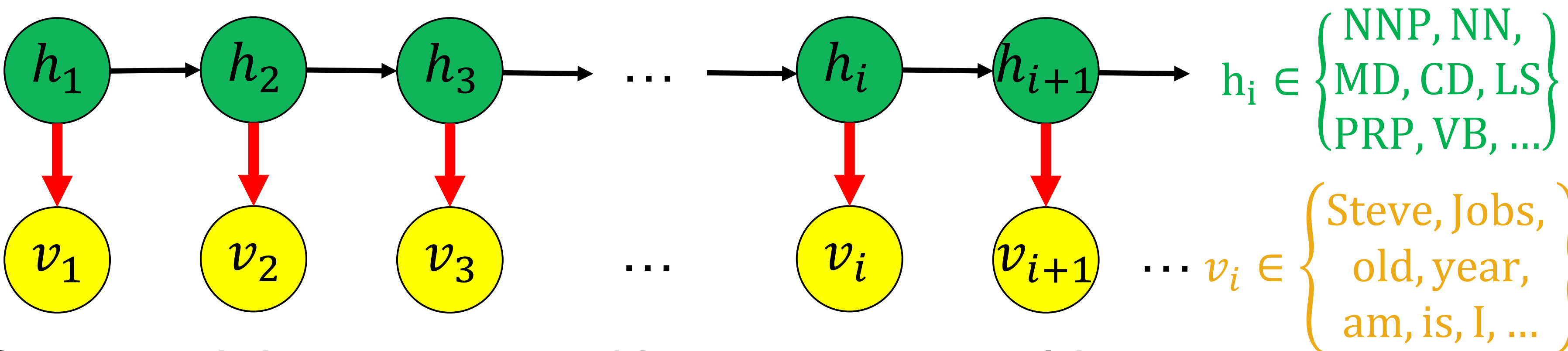


$$\underset{h}{\operatorname{argmax}} P(h|v) = \underset{h}{\operatorname{argmax}} P(h, v)$$

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

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

隐马尔科夫模型：Viterbi算法



Steve Jobs , 42 years old , ……

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

$$= \underset{h_1,h_2,\ldots,h_i,\ldots,h_N}{\operatorname{argmax}} \begin{pmatrix} P(v_1|h_1)P(v_2|h_2)\ldots P(v_i|h_i)\ldots P(v_N|h_N) \\ P(h_1)P(h_2|h_1)\ldots P(h_{i+1}|h_i)\ldots P(h_{N-1}|h_N) \end{pmatrix}$$

已知：

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

	NNP	,	CD	NNS	JJ
NNP	0.2	0.2	0.01	0.01	0.01
,	0.2	0	0.3	0.2	0.1
CD	0.15	0.01	0.1	0.4	0.2
NNS	0.2	0.1	0.01	0.01	0.1
JJ	0.1	0.1	0.1	0.2	0.4

NNP	0.05
,	0.3
CD	0.01
NNS	0.2
JJ	0.3

求联合概率分布：

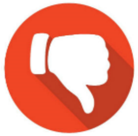
求以下句子的最佳POS标注：

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**



$\log P(h_2, v) = -21.92$ $P(h_2, v) = 0.0000000003024$

$$\log(MN) = \log M + \log N$$

隐马尔科夫模型：Viterbi算法

已知：

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

	NNP	,	CD	NNS	JJ
NNP	0.2	0.2	0.01	0.01	0.01
,	0.2	0	0.3	0.2	0.1
CD	0.15	0.01	0.1	0.4	0.2
NNS	0.2	0.1	0.01	0.01	0.1
JJ	0.1	0.1	0.1	0.2	0.4

NNP	0.05
,	0.3
CD	0.01
NNS	0.2
JJ	0.3

求以下句子的最佳POS标注：

给定： Steve Jobs , 42 years old

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

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

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

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

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

隐马尔科夫模型：Viterbi算法

已知：

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

	NNP	,	CD	NNS	JJ
NNP	0.2	0.2	0.01	0.01	0.01
,	0.2	0	0.3	0.2	0.1
CD	0.15	0.01	0.1	0.4	0.2
NNS	0.2	0.1	0.01	0.01	0.1
JJ	0.1	0.1	0.1	0.2	0.4

NNP	0.05
,	0.3
CD	0.01
NNS	0.2
JJ	0.3

求以下句子的最佳POS标注：

给定： Steve Jobs , 42 years old

	Steve
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	0.05	0.01	0.1	0.01	0.04	0.01
,	0.01	0.01	0.01	0.7	0.01	0.01
CD	0.01	0.01	0.01	0.01	0.01	0.9
NNS	0.2	0.1	0.01	0.01	0.01	0.01
JJ	0.01	0.01	0.01	0.01	0.4	0.01

	NNP	,	CD	NNS	JJ
NNP	0.2	0.2	0.01	0.01	0.01
,	0.2	0	0.3	0.2	0.1
CD	0.15	0.01	0.1	0.4	0.2
NNS	0.2	0.1	0.01	0.01	0.1
JJ	0.1	0.1	0.1	0.2	0.4

NNP	0.05
,	0.3
CD	0.01
NNS	0.2
JJ	0.3

求以下句子的最佳POS标注：

给定：Steve Jobs , 42 years old

	Steve
NNP	0.005
,	0.003
CD	0.0001
NNS	0.002
JJ	0.003

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

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	0.2	0.01	0.01	0.01	0.05
,	0.01	0.01	0.01	0.7	0.01	0.01	0.2	0	0.3	0.2	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	0.2	0.01
NNS	0.2	0.1	0.01	0.01	0.01	0.01	0.2	0.1	0.01	0.01	0.1	0.2
JJ	0.01	0.01	0.01	0.01	0.4	0.01	0.1	0.1	0.1	0.2	0.4	0.3

求以下句子的最佳POS标注：

给定： Steve Jobs , 42 years old

	Steve	
NNP	0.005	→ 5e−5
,	0.003	
CD	0.0001	
NNS	0.002	
JJ	0.003	

$$\max_{left}(left \cdot P(,|left\_tag)P(\text{Jobs}|,))$$

已知：

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

	NNP	,	CD	NNS	JJ
NNP	0.2	0.2	0.01	0.01	0.01
,	0.2	0	0.3	0.2	0.1
CD	0.15	0.01	0.1	0.4	0.2
NNS	0.2	0.1	0.01	0.01	0.1
JJ	0.1	0.1	0.1	0.2	0.4

NNP	0.05
,	0.3
CD	0.01
NNS	0.2
JJ	0.3

求以下句子的最佳POS标注：

给定：Steve Jobs , 42 years old

	Steve	Jobs
NNP	0.005	5e−5
,	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	NNS	JJ	
NNP	0.05	0.01	0.1	0.01	0.04	0.01	0.2	0.2	0.01	0.01	0.01	0.05
,	0.01	0.01	0.01	0.7	0.01	0.01	0.2	0	0.3	0.2	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	0.2	0.01
NNS	0.2	0.1	0.01	0.01	0.01	0.01	0.2	0.1	0.01	0.01	0.1	0.2
JJ	0.01	0.01	0.01	0.01	0.4	0.01	0.1	0.1	0.1	0.2	0.4	0.3

求以下句子的最佳POS标注：

给定：Steve Jobs , 42 years old

NNP 0.005 → 5e−5
, 0.003 → 1e−5
CD 0.0001
NNS 0.002
JJ 0.003

$$\max_{left}(left \cdot P(\mathbf{CD}|left\_tag)P(\text{Jobs}|\mathbf{CD}))$$

隐马尔科夫模型：Viterbi算法

已知：

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

	NNP	,	CD	NNS	JJ
NNP	0.2	0.2	0.01	0.01	0.01
,	0.2	0	0.3	0.2	0.1
CD	0.15	0.01	0.1	0.4	0.2
NNS	0.2	0.1	0.01	0.01	0.1
JJ	0.1	0.1	0.1	0.2	0.4

NNP	0.05
,	0.3
CD	0.01
NNS	0.2
JJ	0.3

求以下句子的最佳POS标注：

给定： Steve Jobs , 42 years old

	Steve	Jobs
NNP	0.005	5e−5
,	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	0.05	0.01	0.1	0.01	0.04	0.01
,	0.01	0.01	0.01	0.7	0.01	0.01
CD	0.01	0.01	0.01	0.01	0.01	0.9
NNS	0.2	0.1	0.01	0.01	0.01	0.01
JJ	0.01	0.01	0.01	0.01	0.4	0.01

	NNP	,	CD	NNS	JJ
NNP	0.2	0.2	0.01	0.01	0.01
,	0.2	0	0.3	0.2	0.1
CD	0.15	0.01	0.1	0.4	0.2
NNS	0.2	0.1	0.01	0.01	0.1
JJ	0.1	0.1	0.1	0.2	0.4

NNP	0.05
,	0.3
CD	0.01
NNS	0.2
JJ	0.3

求以下句子的最佳POS标注：

给定： Steve Jobs , 42 years old

	Steve	Jobs
NNP	0.005	5e−5
,	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	NNS	JJ	
NNP	0.05	0.01	0.1	0.01	0.04	0.01	0.2	0.2	0.01	0.01	0.01	0.05
,	0.01	0.01	0.01	0.7	0.01	0.01	0.2	0	0.3	0.2	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	0.2	0.01
NNS	0.2	0.1	0.01	0.01	0.01	0.01	0.2	0.1	0.01	0.01	0.1	0.2
JJ	0.01	0.01	0.01	0.01	0.4	0.01	0.1	0.1	0.1	0.2	0.4	0.3

求以下句子的最佳POS标注：

给定：	Steve	Jobs	,	42	years	old
NNP	0.005	5e−5				
,	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	NNS	JJ	
NNP	0.05	0.01	0.1	0.01	0.04	0.01	0.2	0.2	0.01	0.01	0.01	0.05
,	0.01	0.01	0.01	0.7	0.01	0.01	0.2	0	0.3	0.2	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	0.2	0.01
NNS	0.2	0.1	0.01	0.01	0.01	0.01	0.2	0.1	0.01	0.01	0.1	0.2
JJ	0.01	0.01	0.01	0.01	0.4	0.01	0.1	0.1	0.1	0.2	0.4	0.3

求以下句子的最佳POS标注：

给定： Steve Jobs , 42 years old

	Steve	Jobs
NNP	0.005	5e−5
,	0.003	1e−5
CD	0.0001	9e−5
NNS	0.002	1.2e−4
JJ	0.003	1.2e−5

$$\max_{left}(left \cdot P(\mathbf{NNP}|left\_tag)P(,\ |\mathbf{NNP}))$$

隐马尔科夫模型：Viterbi算法

已知：

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

	NNP	,	CD	NNS	JJ
NNP	0.2	0.2	0.01	0.01	0.01
,	0.2	0	0.3	0.2	0.1
CD	0.15	0.01	0.1	0.4	0.2
NNS	0.2	0.1	0.01	0.01	0.1
JJ	0.1	0.1	0.1	0.2	0.4

NNP	0.05
,	0.3
CD	0.01
NNS	0.2
JJ	0.3

求以下句子的最佳POS标注：

给定：	Steve	Jobs	,	42	years	old
NNP	0.005	5e−5	1e−7			
,	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	0.05	0.01	0.1	0.01	0.04	0.01
,	0.01	0.01	0.01	0.7	0.01	0.01
CD	0.01	0.01	0.01	0.01	0.01	0.9
NNS	0.2	0.1	0.01	0.01	0.01	0.01
JJ	0.01	0.01	0.01	0.01	0.4	0.01

	NNP	,	CD	NNS	JJ
NNP	0.2	0.2	0.01	0.01	0.01
,	0.2	0	0.3	0.2	0.1
CD	0.15	0.01	0.1	0.4	0.2
NNS	0.2	0.1	0.01	0.01	0.1
JJ	0.1	0.1	0.1	0.2	0.4

NNP	0.05
,	0.3
CD	0.01
NNS	0.2
JJ	0.3

求以下句子的最佳POS标注：

给定：	Steve	Jobs	,	42	years	old
NNP	0.005	5e−5	1e−7	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	**3.03e−8**

隐马尔科夫模型：Viterbi算法

已知：

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

	NNP	,	CD	NNS	JJ
NNP	0.2	0.2	0.01	0.01	0.01
,	0.2	0	0.3	0.2	0.1
CD	0.15	0.01	0.1	0.4	0.2
NNS	0.2	0.1	0.01	0.01	0.1
JJ	0.1	0.1	0.1	0.2	0.4

NNP	0.05
,	0.3
CD	0.01
NNS	0.2
JJ	0.3

求以下句子的最佳POS标注：

给定：	Steve	Jobs	,	42	years	old
NNP	0.005	5e-5	1e-7	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	**3.03e-8**

隐马尔科夫模型：Viterbi算法

已知：

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

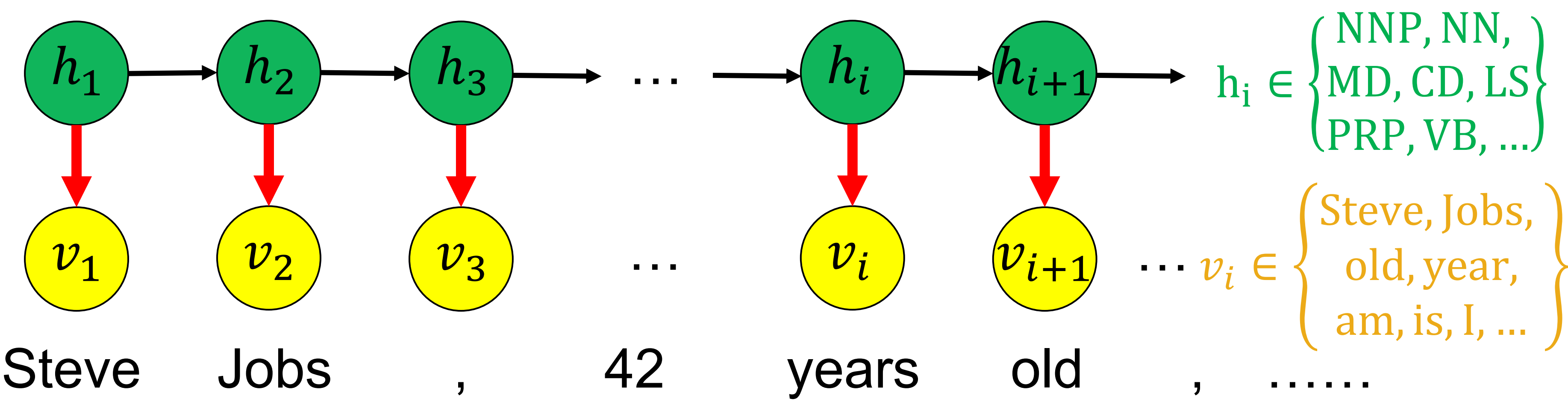
	NNP	,	CD	NNS	JJ
NNP	0.2	0.2	0.01	0.01	0.01
,	0.2	0	0.3	0.2	0.1
CD	0.15	0.01	0.1	0.4	0.2
NNS	0.2	0.1	0.01	0.01	0.1
JJ	0.1	0.1	0.1	0.2	0.4

NNP	0.05
,	0.3
CD	0.01
NNS	0.2
JJ	0.3

求以下句子的最佳POS标注：

给定：	Steve	Jobs	,	42	years	old
NNP	0.005	5e−5	1e−7	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	**3.03e−8**
	NNP	**NNP**		**CD**	**NNS**	**JJ**

动态规划！



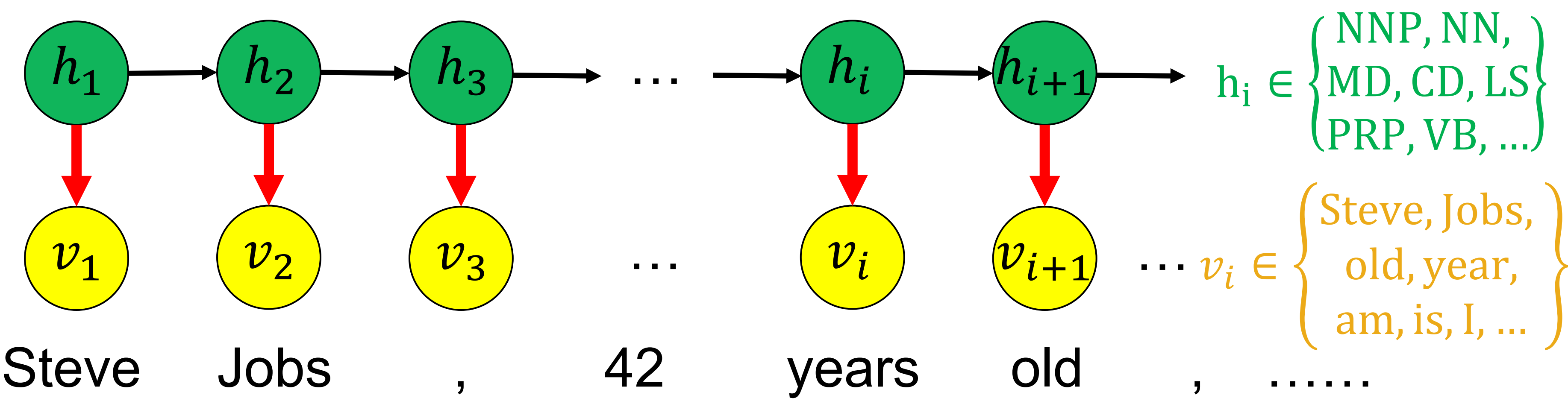
定义直接给出 • $P(h, v)$

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

$$P(h|v) = \frac{P(v|h)P(h)}{\boxed{P(v)}}$$

目录

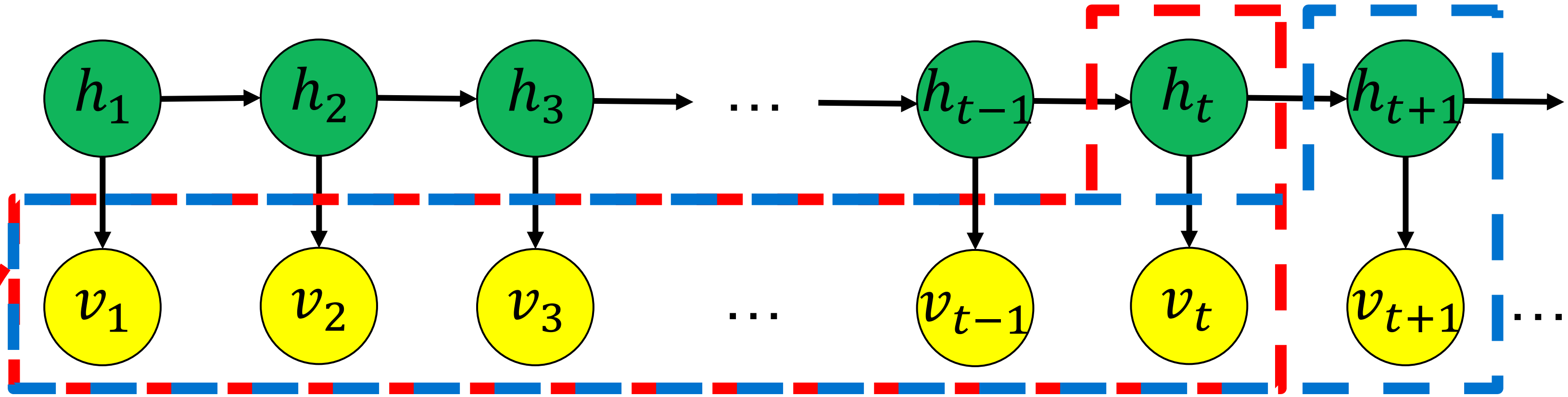
隐马尔科夫模型



定义直接给出 • $P(h, v)$

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

前向算法
后向算法 • $P(v)$



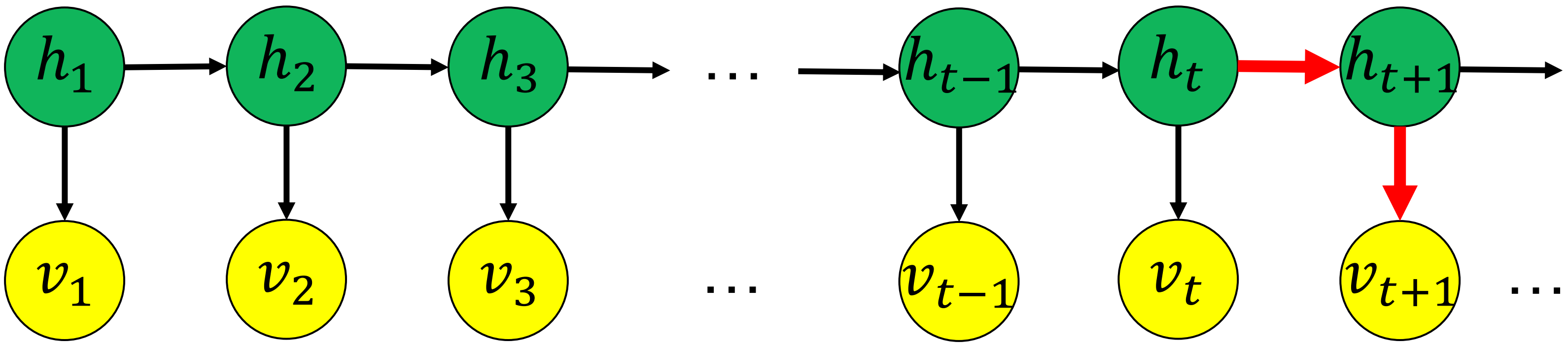
$$P(h,v) = P(v|h)\cdot P(h) = \prod_{j=1}^{N} P(v_j|h_j)\,P(h_1)\prod_{i=1}^{N-1} P(h_{i+1}|h_i)$$

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

$$\alpha_{t+1}(j) =$$

?

隐马尔科夫模型：α与β

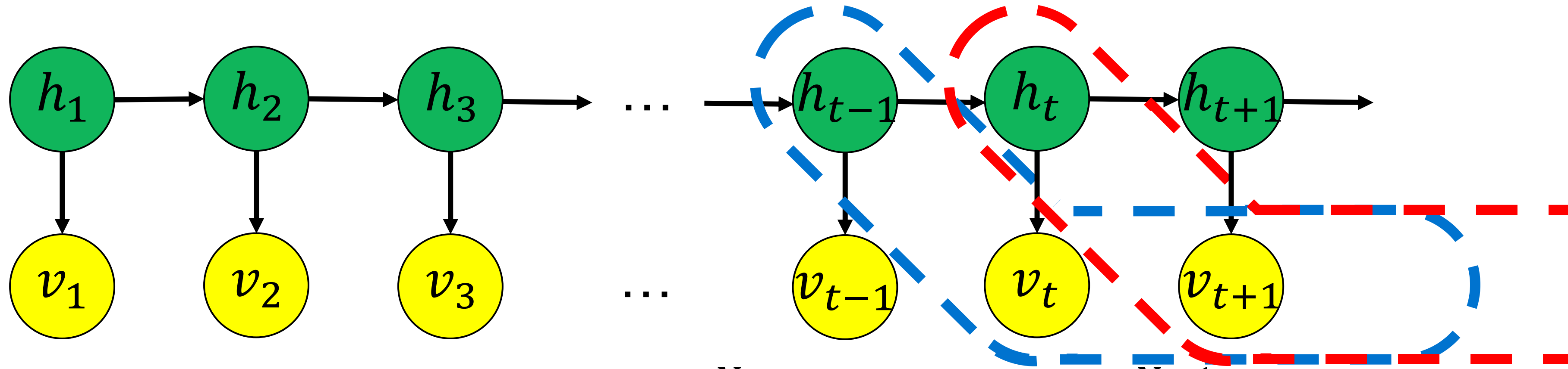


$$P(h,v)=P(v|h)\cdot P(h)=\prod_{j=1}^{N}P\left(v_j|h_j\right)P(h_1)\prod_{i=1}^{N-1}P(h_{i+1}|h_i)$$

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

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

隐马尔科夫模型：α与β



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

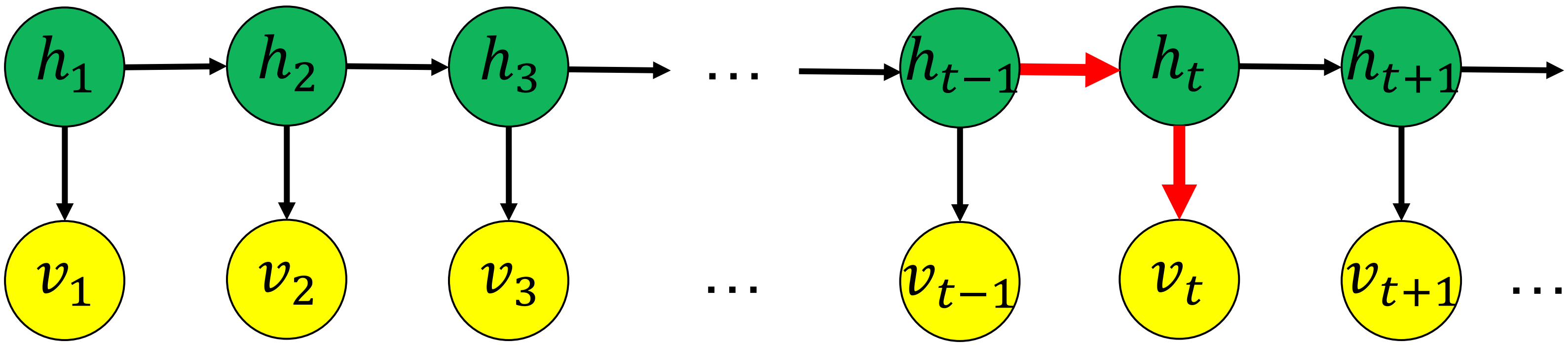
$$\alpha_t(j) = P(v_1, v_2, \cdots, v_t, h_t = j)$$

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

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

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

隐马尔科夫模型：α与β



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

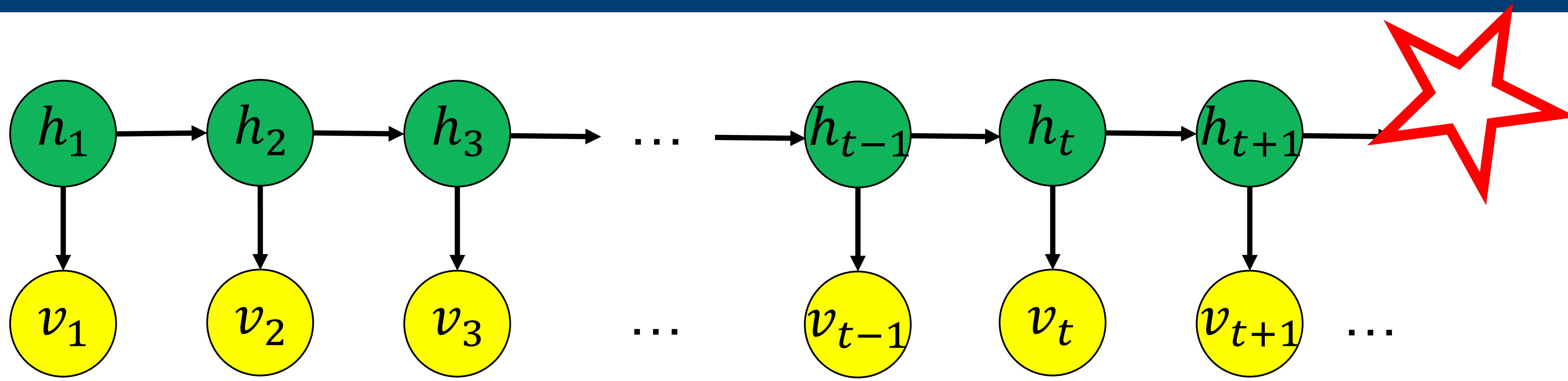
$$\alpha_t(j)=P(v_1,v_2,\cdots,v_t,h_t=j)$$

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

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

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

隐马尔科夫模型：α与β



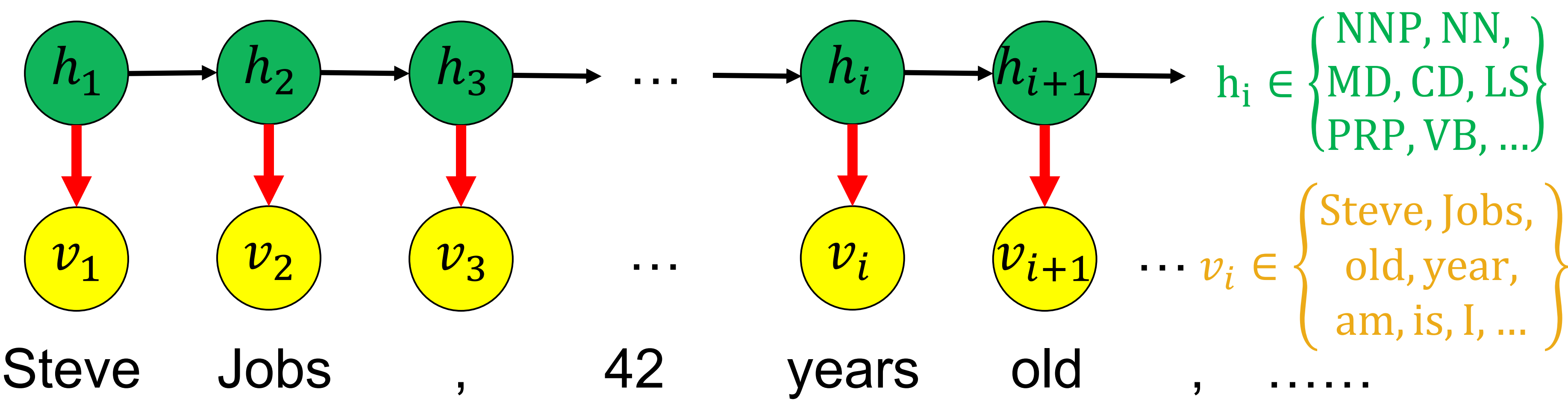
$$P(h,v) = P(v|h) \cdot P(h) = \prod_{j=1}^{N} P\left(v_j|h_j\right) P(h_1) \prod_{i=1}^{N-1} P(h_{i+1}|h_i)$$

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

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

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

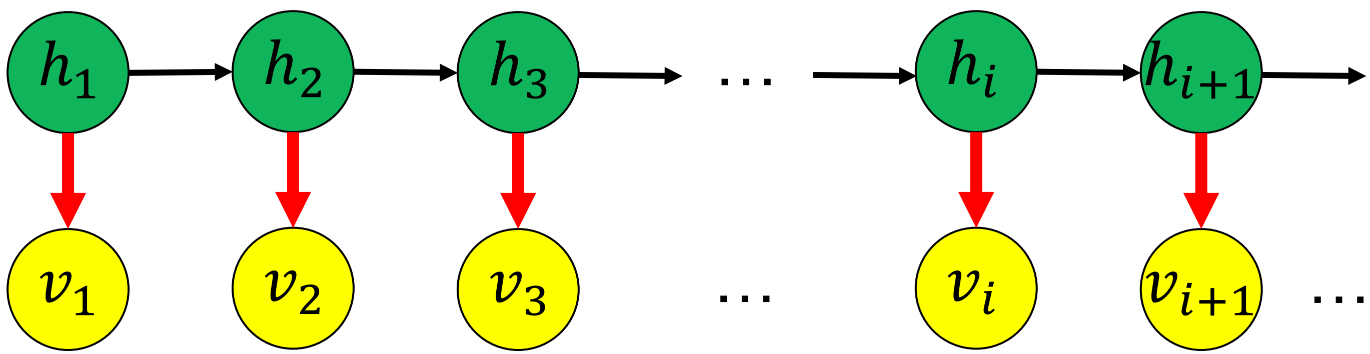
$$\beta_{t-1}(j) = \sum_{i=1}^{S} \beta_t(i) P(h_t = i|h_{t-1} = j) P(v_t|h_t = i)$$



定义直接给出 • $P(h, v)$

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

前向算法
后向算法 • $P(v)$



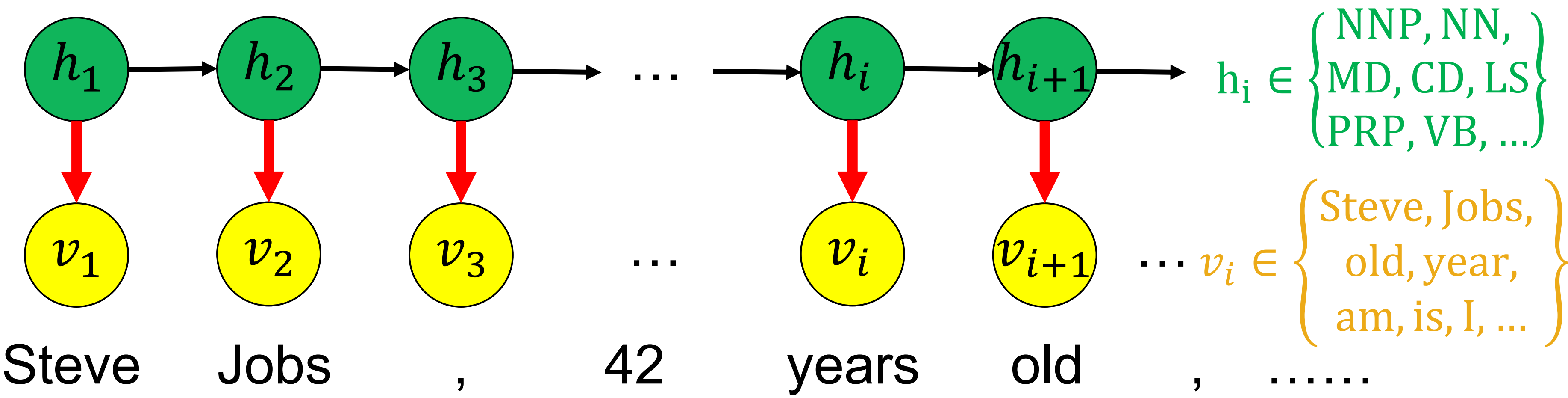
$h_i \in \begin{Bmatrix} \text{NNP, NN,} \\ \text{MD, CD, LS} \\ \text{PRP, VB, ...} \end{Bmatrix}$

$\cdots\ v_i \in \begin{Bmatrix} \text{Steve, Jobs,} \\ \text{old, year,} \\ \text{am, is, I, ...} \end{Bmatrix}$

Steve	Jobs	,	42	years	old	,	......	
NNP	**NNP**	**,**	**CD**	**NNS**	**JJ**			$P(h_1, v)$
								$+$
NNP	**NNS**	**,**	**CD**	**NNS**	**JJ**			$P(h_2, v)$
								$+$
			$\vdots$					$\vdots$

- $P(v) = \sum_{h} P(h, v)$

隐马尔科夫模型：前向算法



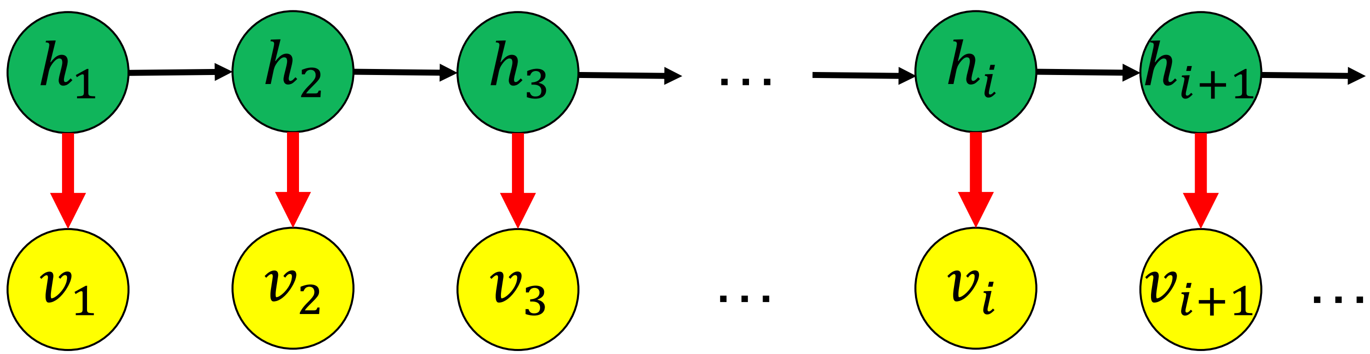
$$P(v) = \sum_{h} P(h, v)$$

$$= \sum_{h} \big(P(v|h) \cdot P(h)\big)$$

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

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：前向算法



$$\mathrm{h_i} \in \left\{ \begin{matrix} \mathrm{NNP, NN,} \\ \mathrm{MD, CD, LS} \\ \mathrm{PRP, VB, \ldots} \end{matrix} \right\}$$

$$\cdots\ v_i \in \left\{ \begin{matrix} \mathrm{Steve, Jobs,} \\ \mathrm{old, year,} \\ \mathrm{am, is, I, \ldots} \end{matrix} \right\}$$

Steve Jobs , 42 years old , ……

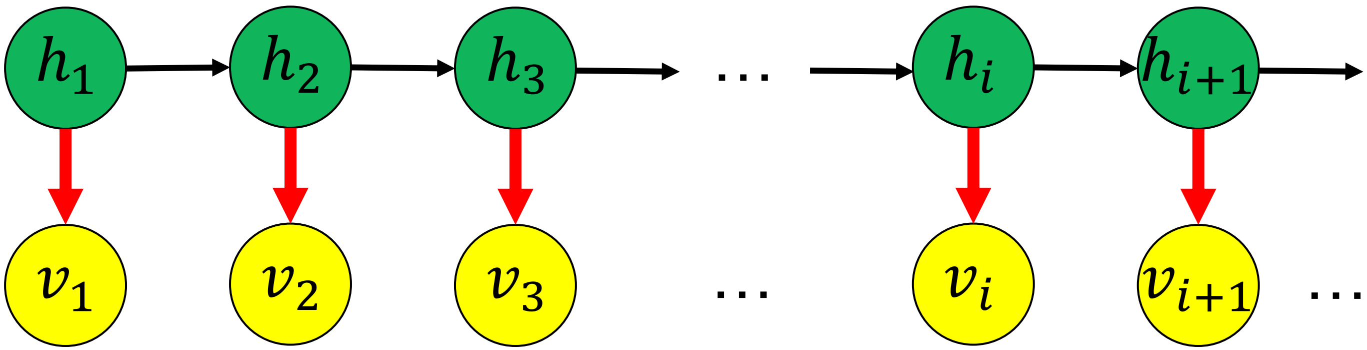
$$P(v) = \sum_h \left(\prod_{j=1}^{N} P(v_j|h_j)\, P(h_1) \prod_{i=1}^{N-1} P(h_{i+1}|h_i) \right)$$

前向算法

$$= \sum_{h_1,h_2,\ldots,h_i,\ldots,h_N} \left(\begin{matrix} P(v_1|h_1)P(v_2|h_2) \ldots P(v_i|h_i) \ldots P(v_N|h_N) \\ P(h_1)P(h_2|h_1) \ldots P(h_{i+1}|h_i) \ldots P(h_N|h_{N-1}) \end{matrix} \right)$$

$$= \sum_{h_N} \left\{ P(v_N|h_N) \sum_{h_{N-1}} P(h_N|h_{N-1}) \cdots \sum \cdots \sum_{h_2} \left[P(v_2|h_2)P(h_3|h_2) \left(\sum_{h_1} \begin{matrix} P(v_1|h_1) \\ P(h_1)P(h_2|h_1) \end{matrix} \right) \right] \right\}$$

隐马尔科夫模型：前向算法



$$h_i \in \begin{Bmatrix} \text{NNP, NN,} \\ \text{MD, CD, LS} \\ \text{PRP, VB, ...} \end{Bmatrix}$$

$$v_i \in \begin{Bmatrix} \text{Steve, Jobs,} \\ \text{old, year,} \\ \text{am, is, I, ...} \end{Bmatrix}$$

Steve Jobs , 42 years old , ...

后向算法

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

$$= \sum_{h_1, h_2, \ldots, h_i, \ldots, h_N} \begin{pmatrix} P(v_1|h_1)P(v_2|h_2) \ldots P(v_i|h_i) \ldots P(v_N|h_N) \\ P(h_1)P(h_2|h_1) \ldots P(h_{i+1}|h_i) \ldots P(h_N|h_{N-1}) \end{pmatrix}$$

$$= \sum_{h_N} \left\{ P(v_N|h_N) \sum_{h_{N-1}} P(h_N|h_{N-1}) \cdots \sum \cdots \sum_{h_2} \left[P(v_2|h_2)P(h_3|h_2) \left(\sum_{h_1} \begin{matrix} P(v_1|h_1) \\ P(h_1)P(h_2|h_1) \end{matrix} \right) \right] \right\}$$

隐马尔科夫模型：前向算法

已知：

	jobs	years	steve	,	old	42
NNP	0.3	0.03	0.4	0.03	0.2	0.04
,	0.01	0.02	0.02	0.9	0.02	0.03
CD	0.02	0.02	0.02	0.02	0.02	0.9
NNS	0.5	0.3	0.05	0.05	0.05	0.05
JJ	0.02	0.02	0.02	0.02	0.9	0.02

	NNP	,	CD	NNS	JJ
NNP	0.45	0.45	0.05	0.02	0.03
,	0.3	0.01	0.4	0.2	0.09
CD	0.15	0.01	0.1	0.5	0.24
NNS	0.45	0.25	0.07	0.03	0.2
JJ	0.1	0.1	0.1	0.3	0.4

NNP	0.06
,	0.35
CD	0.01
NNS	0.23
JJ	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	0.3	0.03	0.4	0.03	0.2	0.04
,	0.01	0.02	0.02	0.9	0.02	0.03
CD	0.02	0.02	0.02	0.02	0.02	0.9
NNS	0.5	0.3	0.05	0.05	0.05	0.05
JJ	0.02	0.02	0.02	0.02	0.9	0.02

	NNP	,	CD	NNS	JJ
NNP	0.45	0.45	0.05	0.02	0.03
,	0.3	0.01	0.4	0.2	0.09
CD	0.15	0.01	0.1	0.5	0.24
NNS	0.45	0.25	0.07	0.03	0.2
JJ	0.1	0.1	0.1	0.3	0.4

NNP	0.06
,	0.35
CD	0.01
NNS	0.23
JJ	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

$$S \rightarrow \left\{ \begin{array}{c} \textbf{NNP} \\ \textbf{,} \\ \textbf{CD} \\ \textbf{NNS} \\ \textbf{JJ} \end{array} \right.$$

$$\alpha_2(\textbf{CD})$$

第二个位置的隐变量是**CD**，且**CD**生成的是Jobs的概率。

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

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

已知：

	jobs	years	steve	,	old	42
NNP	0.3	0.03	0.4	0.03	0.2	0.04
,	0.01	0.02	0.02	0.9	0.02	0.03
CD	0.02	0.02	0.02	0.02	0.02	0.9
NNS	0.5	0.3	0.05	0.05	0.05	0.05
JJ	0.02	0.02	0.02	0.02	0.9	0.02

	NNP	,	CD	NNS	JJ
NNP	0.45	0.45	0.05	0.02	0.03
,	0.3	0.01	0.4	0.2	0.09
CD	0.15	0.01	0.1	0.5	0.24
NNS	0.45	0.25	0.07	0.03	0.2
JJ	0.1	0.1	0.1	0.3	0.4

NNP	0.06
,	0.35
CD	0.01
NNS	0.23
JJ	0.35

使用前向算法求以下句子出现的概率：

给定： Steve Jobs , 42 yea

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

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

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

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

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

隐马尔科夫模型：前向算法

已知：

	jobs	years	steve	,	old	42
NNP	0.3	0.03	0.4	0.03	0.2	0.04
,	0.01	0.02	0.02	0.9	0.02	0.03
CD	0.02	0.02	0.02	0.02	0.02	0.9
NNS	0.5	0.3	0.05	0.05	0.05	0.05
JJ	0.02	0.02	0.02	0.02	0.9	0.02

	NNP	,	CD	NNS	JJ
NNP	0.45	0.45	0.05	0.02	0.03
,	0.3	0.01	0.4	0.2	0.09
CD	0.15	0.01	0.1	0.5	0.24
NNS	0.45	0.25	0.07	0.03	0.2
JJ	0.1	0.1	0.1	0.3	0.4

NNP	0.06
,	0.35
CD	0.01
NNS	0.23
JJ	0.35

使用前向算法求以下句子出现的概率：

给定： Steve Jobs , 42 years

NNP	0.024
,	0.007
CD	0.0002
NNS	0.0115
JJ	0.007

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

已知：

	jobs	years	steve	,	old	42	NNP	,	CD	NNS	JJ	
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

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

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

（Viterbi算法）

隐马尔科夫模型：前向算法

已知：

	jobs	years	steve	,	old	42
NNP	0.3	0.03	0.4	0.03	0.2	0.04
,	0.01	0.02	0.02	0.9	0.02	0.03
CD	0.02	0.02	0.02	0.02	0.02	0.9
NNS	0.5	0.3	0.05	0.05	0.05	0.05
JJ	0.02	0.02	0.02	0.02	0.9	0.02

	NNP	,	CD	NNS	JJ
NNP	0.45	0.45	0.05	0.02	0.03
,	0.3	0.01	0.4	0.2	0.09
CD	0.15	0.01	0.1	0.5	0.24
NNS	0.45	0.25	0.07	0.03	0.2
JJ	0.1	0.1	0.1	0.3	0.4

NNP	0.06
,	0.35
CD	0.01
NNS	0.23
JJ	0.35

使用前向算法求以下句子出现的概率：

给定： Steve Jobs , 42 years

NNP	0.024
,	0.007
CD	0.0002
NNS	0.0115
JJ	0.007

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

隐马尔科夫模型：前向算法

已知：

	jobs	years	steve	,	old	42
NNP	0.3	0.03	0.4	0.03	0.2	0.04
,	0.01	0.02	0.02	0.9	0.02	0.03
CD	0.02	0.02	0.02	0.02	0.02	0.9
NNS	0.5	0.3	0.05	0.05	0.05	0.05
JJ	0.02	0.02	0.02	0.02	0.9	0.02

	NNP	,	CD	NNS	JJ
NNP	0.45	0.45	0.05	0.02	0.03
,	0.3	0.01	0.4	0.2	0.09
CD	0.15	0.01	0.1	0.5	0.24
NNS	0.45	0.25	0.07	0.03	0.2
JJ	0.1	0.1	0.1	0.3	0.4

NNP	0.06
,	0.35
CD	0.01
NNS	0.23
JJ	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

	Steve	Jobs
NNP	0.024	5.64e-3
,	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	0.3	0.03	0.4	0.03	0.2	0.04
,	0.01	0.02	0.02	0.9	0.02	0.03
CD	0.02	0.02	0.02	0.02	0.02	0.9
NNS	0.5	0.3	0.05	0.05	0.05	0.05
JJ	0.02	0.02	0.02	0.02	0.9	0.02

	NNP	,	CD	NNS	JJ
NNP	0.45	0.45	0.05	0.02	0.03
,	0.3	0.01	0.4	0.2	0.09
CD	0.15	0.01	0.1	0.5	0.24
NNS	0.45	0.25	0.07	0.03	0.2
JJ	0.1	0.1	0.1	0.3	0.4

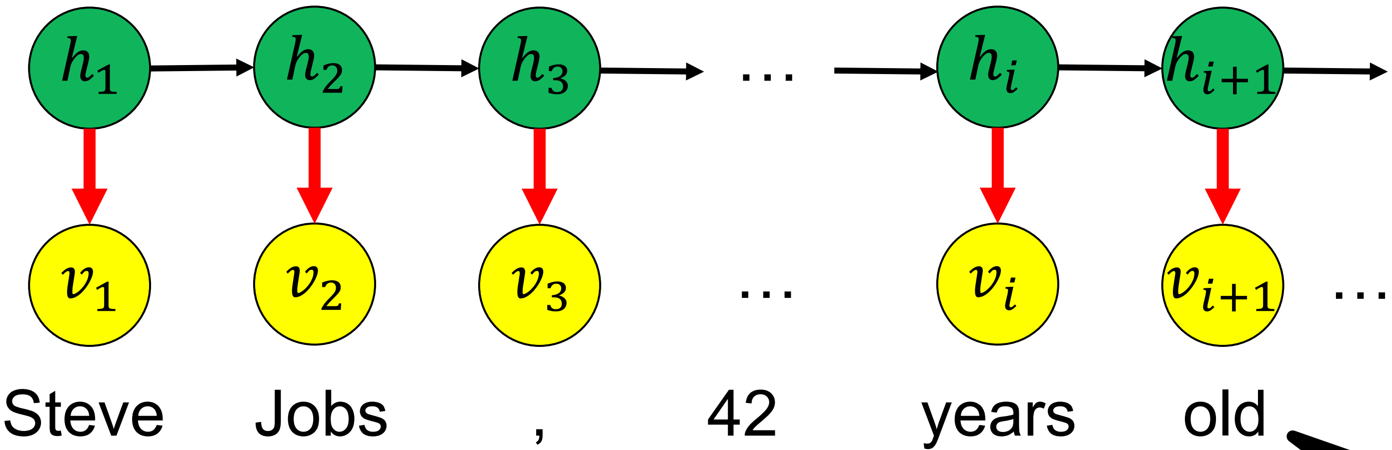
NNP	0.06
,	0.35
CD	0.01
NNS	0.23
JJ	0.35

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	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

$\sum$

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



$\mathrm{h_i} \in \{\text{NNP, NN, MD, CD, LS, PRP, VB, ...}\}$

$\cdots\ v_i \in \{\text{Steve, Jobs, old, year, am, is, I, ...}\}$

后向算法

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

$$= \sum_{h_1, h_2, \ldots, h_i, \ldots, h_N} \begin{pmatrix} P(v_1|h_1)P(v_2|h_2) \ldots P(v_i|h_i) \ldots P(v_N|h_N) \\ P(h_1)P(h_2|h_1) \ldots P(h_{i+1}|h_i) \ldots P(h_N|h_{N-1}) \end{pmatrix}$$

$$= \sum_{h_N} \left\{ P(v_N|h_N) \sum_{h_{N-1}} P(h_N|h_{N-1}) \cdots \sum \cdots \sum_{h_2} \left[P(v_2|h_2)P(h_3|h_2) \left(\sum_{h_1} \begin{matrix} P(v_1|h_1) \\ P(h_1)P(h_2|h_1) \end{matrix} \right) \right] \right\}$$

已知：

	jobs	years	steve	,	old	42
NNP	0.3	0.03	0.4	0.03	0.2	0.04
,	0.01	0.02	0.02	0.9	0.02	0.03
CD	0.02	0.02	0.02	0.02	0.02	0.9
NNS	0.5	0.3	0.05	0.05	0.05	0.05
JJ	0.02	0.02	0.02	0.02	0.9	0.02

	NNP	,	CD	NNS	JJ
NNP	0.45	0.45	0.05	0.02	0.03
,	0.3	0.01	0.4	0.2	0.09
CD	0.15	0.01	0.1	0.5	0.24
NNS	0.45	0.25	0.07	0.03	0.2
JJ	0.1	0.1	0.1	0.3	0.4

NNP	0.06
,	0.35
CD	0.01
NNS	0.23
JJ	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

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

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

隐马尔科夫模型：后向算法

已知：

	jobs	years	steve	,	old	42
NNP	0.3	0.03	0.4	0.03	0.2	0.04
,	0.01	0.02	0.02	0.9	0.02	0.03
CD	0.02	0.02	0.02	0.02	0.02	0.9
NNS	0.5	0.3	0.05	0.05	0.05	0.05
JJ	0.02	0.02	0.02	0.02	0.9	0.02

	NNP	,	CD	NNS	JJ
NNP	0.45	0.45	0.05	0.02	0.03
,	0.3	0.01	0.4	0.2	0.09
CD	0.15	0.01	0.1	0.5	0.24
NNS	0.45	0.25	0.07	0.03	0.2
JJ	0.1	0.1	0.1	0.3	0.4

0.06
0.35
0.01
0.23
0.35

使用后向算法求以下句子出现的概率：

eve 给定：Jobs , 42 years old

NNP
,
CD
NNS
JJ

$$\sum_{right}\left(right \cdot P(right\_tag|\mathbf{NNP})P(old|right\_tag)\right)$$

1
1
1
1
1

已知：

	jobs	years	steve	,	old	42
NNP	0.3	0.03	0.4	0.03	0.2	0.04
,	0.01	0.02	0.02	0.9	0.02	0.03
CD	0.02	0.02	0.02	0.02	0.02	0.9
NNS	0.5	0.3	0.05	0.05	0.05	0.05
JJ	0.02	0.02	0.02	0.02	0.9	0.02

	NNP	,	CD	NNS	JJ
NNP	0.45	0.45	0.05	0.02	0.03
,	0.3	0.01	0.4	0.2	0.09
CD	0.15	0.01	0.1	0.5	0.24
NNS	0.45	0.25	0.07	0.03	0.2
JJ	0.1	0.1	0.1	0.3	0.4

0.06
0.35
0.01
0.23
0.35

使用后向算法求以下句子出现的概率：

eve 给定：Jobs , 42 years old

NNP
,
CD
NNS
JJ

0.128

$$\sum_{right} \left(right \cdot P(right\_tag|,)P(\text{old}|right\_tag) \right)$$

1
1
1
1
1

隐马尔科夫模型：后向算法

已知：

	jobs	years	steve	,	old	42
NNP	0.3	0.03	0.4	0.03	0.2	0.04
,	0.01	0.02	0.02	0.9	0.02	0.03
CD	0.02	0.02	0.02	0.02	0.02	0.9
NNS	0.5	0.3	0.05	0.05	0.05	0.05
JJ	0.02	0.02	0.02	0.02	0.9	0.02

	NNP	,	CD	NNS	JJ
NNP	0.45	0.45	0.05	0.02	0.03
,	0.3	0.01	0.4	0.2	0.09
CD	0.15	0.01	0.1	0.5	0.24
NNS	0.45	0.25	0.07	0.03	0.2
JJ	0.1	0.1	0.1	0.3	0.4

NNP	0.06
,	0.35
CD	0.01
NNS	0.23
JJ	0.35

使用后向算法求以下句子出现的概率：

eve 给定：Jobs , 42 years old

NNP
,
CD
NNS
JJ

0.128
0.1592

$$\sum_{right}\left(right \cdot P(right\_tag|\mathbf{CD})P(old|right\_tag)\right)$$

1
1
1
1
1

隐马尔科夫模型：后向算法

已知：

	jobs	years	steve	,	old	42
NNP	0.3	0.03	0.4	0.03	0.2	0.04
,	0.01	0.02	0.02	0.9	0.02	0.03
CD	0.02	0.02	0.02	0.02	0.02	0.9
NNS	0.5	0.3	0.05	0.05	0.05	0.05
JJ	0.02	0.02	0.02	0.02	0.9	0.02

	NNP	,	CD	NNS	JJ
NNP	0.45	0.45	0.05	0.02	0.03
,	0.3	0.01	0.4	0.2	0.09
CD	0.15	0.01	0.1	0.5	0.24
NNS	0.45	0.25	0.07	0.03	0.2
JJ	0.1	0.1	0.1	0.3	0.4

NNP	0.06
,	0.35
CD	0.01
NNS	0.23
JJ	0.35

使用后向算法求以下句子出现的概率：

eve 给定：Jobs , 42 years old

NNP
,
CD
NNS
JJ

0.128
0.1592
0.2732

$$\sum_{right} \left(right \cdot P(right\_tag|\mathbf{NNS}) P(\mathrm{old}|right\_tag) \right)$$

1
1
1
1
1

隐马尔科夫模型：后向算法

已知：

	jobs	years	steve	,	old	42
NNP	0.3	0.03	0.4	0.03	0.2	0.04
,	0.01	0.02	0.02	0.9	0.02	0.03
CD	0.02	0.02	0.02	0.02	0.02	0.9
NNS	0.5	0.3	0.05	0.05	0.05	0.05
JJ	0.02	0.02	0.02	0.02	0.9	0.02

	NNP	,	CD	NNS	JJ
NNP	0.45	0.45	0.05	0.02	0.03
,	0.3	0.01	0.4	0.2	0.09
CD	0.15	0.01	0.1	0.5	0.24
NNS	0.45	0.25	0.07	0.03	0.2
JJ	0.1	0.1	0.1	0.3	0.4

0.06
0.35
0.01
0.23
0.35

使用后向算法求以下句子出现的概率：

eve 给定：

	Jobs	,	42	years	old
NNP				0.128	1
,				0.1592	1
CD				0.2732	1
NNS				0.2779	1
JJ					1

$$\sum_{right}\left(right \cdot P(right\_tag|\mathbf{JJ})P(old|right\_tag)\right)$$

隐马尔科夫模型：后向算法

已知：

	jobs	years	steve	,	old	42	NNP	,	CD	NNS	JJ	
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：

eve 给定：Jobs , 42 years old

	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	0.3	0.03	0.4	0.03	0.2	0.04
,	0.01	0.02	0.02	0.9	0.02	0.03
CD	0.02	0.02	0.02	0.02	0.02	0.9
NNS	0.5	0.3	0.05	0.05	0.05	0.05
JJ	0.02	0.02	0.02	0.02	0.9	0.02

	NNP	,	CD	NNS	JJ
NNP	0.45	0.45	0.05	0.02	0.03
,	0.3	0.01	0.4	0.2	0.09
CD	0.15	0.01	0.1	0.5	0.24
NNS	0.45	0.25	0.07	0.03	0.2
JJ	0.1	0.1	0.1	0.3	0.4

0.06
0.35
0.01
0.23
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	0.3	0.03	0.4	0.03	0.2	0.04
,	0.01	0.02	0.02	0.9	0.02	0.03
CD	0.02	0.02	0.02	0.02	0.02	0.9
NNS	0.5	0.3	0.05	0.05	0.05	0.05
JJ	0.02	0.02	0.02	0.02	0.9	0.02

	NNP	,	CD	NNS	JJ
NNP	0.45	0.45	0.05	0.02	0.03
,	0.3	0.01	0.4	0.2	0.09
CD	0.15	0.01	0.1	0.5	0.24
NNS	0.45	0.25	0.07	0.03	0.2
JJ	0.1	0.1	0.1	0.3	0.4

NNP	0.06
,	0.35
CD	0.01
NNS	0.23
JJ	0.35

使用后向算法求以下句子出现的概率：

给定：

	jobs	,	42	years	old
NNP			5.34e-3	0.128	1
,			2.08e-2	0.1592	1
CD				0.2732	1
NNS				0.2779	1
JJ				0.399	1

$$\sum_{right}\left(right \cdot P(right\_tag|\mathbf{CD})P(years|right\_tag)\right)$$

隐马尔科夫模型：后向算法

已知：

	jobs	years	steve	,	old	42
NNP	0.3	0.03	0.4	0.03	0.2	0.04
,	0.01	0.02	0.02	0.9	0.02	0.03
CD	0.02	0.02	0.02	0.02	0.02	0.9
NNS	0.5	0.3	0.05	0.05	0.05	0.05
JJ	0.02	0.02	0.02	0.02	0.9	0.02

	NNP	,	CD	NNS	JJ	
NNP	0.45	0.45	0.05	0.02	0.03	0.06
,	0.3	0.01	0.4	0.2	0.09	0.35
CD	0.15	0.01	0.1	0.5	0.24	0.01
NNS	0.45	0.25	0.07	0.03	0.2	0.23
JJ	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：

	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	NNS	JJ	
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	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	0.3	0.03	0.4	0.03	0.2	0.04
,	0.01	0.02	0.02	0.9	0.02	0.03
CD	0.02	0.02	0.02	0.02	0.02	0.9
NNS	0.5	0.3	0.05	0.05	0.05	0.05
JJ	0.02	0.02	0.02	0.02	0.9	0.02

	NNP	,	CD	NNS	JJ
NNP	0.45	0.45	0.05	0.02	0.03
,	0.3	0.01	0.4	0.2	0.09
CD	0.15	0.01	0.1	0.5	0.24
NNS	0.45	0.25	0.07	0.03	0.2
JJ	0.1	0.1	0.1	0.3	0.4

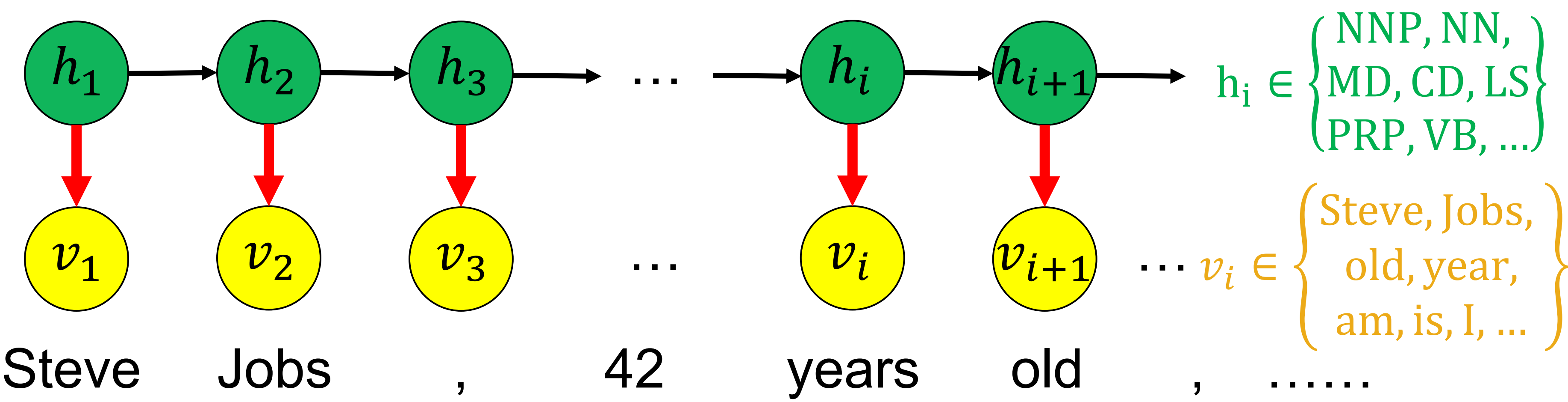
NNP	0.06
,	0.35
CD	0.01
NNS	0.23
JJ	0.35

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

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

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



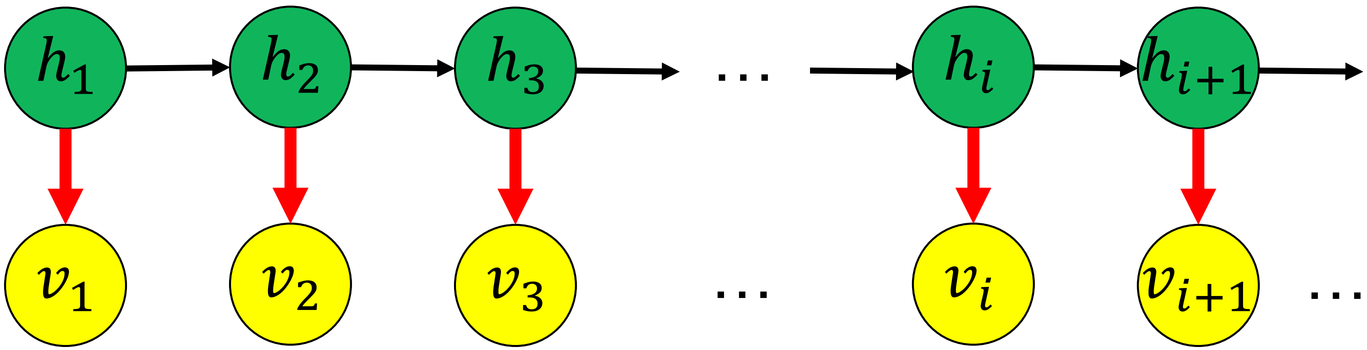
定义直接给出 • $P(h, v)$

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

前向算法
后向算法 • $P(v)$

目录

隐马尔科夫模型



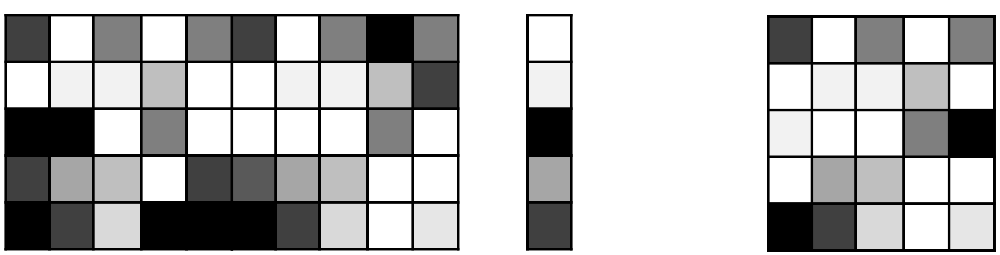
隐马尔科夫模型的定义

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

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

依照左式对联合概率分布$P(v,h)$建模，即被称为HMM模型。

参数：

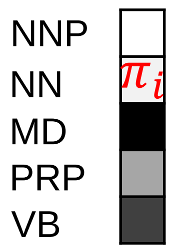


v,h都应是离散随机变量的离散序列。

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

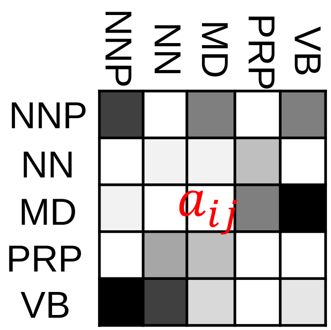
给定POS的标注数据集，
这些参数即可通过数数得到：

$P(h_1)$



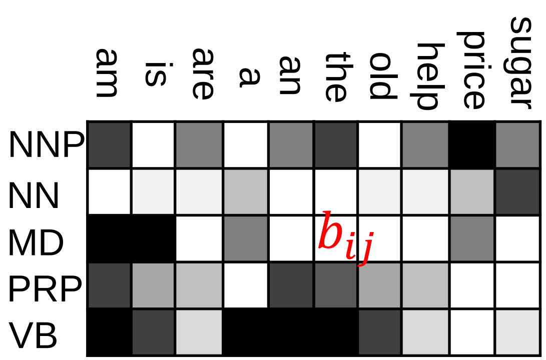
$$\widehat{\pi}_i = P(h_1=i)=\frac{\#(h_1=i)}{\#(\text{句子})}$$

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



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

$P(v_j|h_j)$



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

隐马尔科夫模型：思考题

这样数数的方式，计算非常简便直观。
可是，为什么这样算出来的值是对的？
　　存不存在更好的算法？

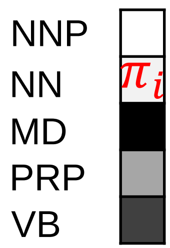
目录

隐马尔科夫模型：参数的有监督学习

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

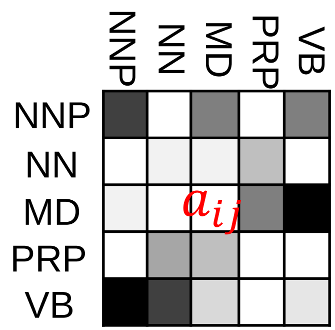
给定POS的标注数据集，
这些参数即可通过数数得到：

$P(h_1)$



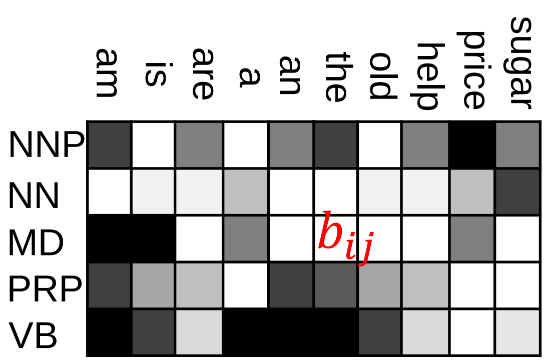
$$\hat{\pi}_i=P(h_1=i)=\frac{\#(h_1=i)}{\#(\text{句子})}$$

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



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

$P(v_j|h_j)$



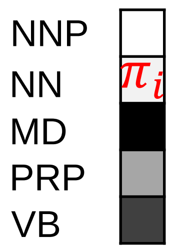
$$\widehat{b_{ij}}=P(v_t=j|h_t=i)=\frac{\#(tag=i,word=j)}{\#(tag=i)}$$

隐马尔科夫模型：参数的无监督学习

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

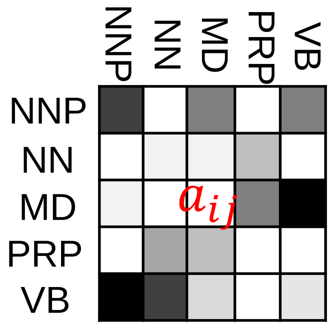
给定POS的无标注标注数据集，
这些参数如何计算？

$P(h_1)$



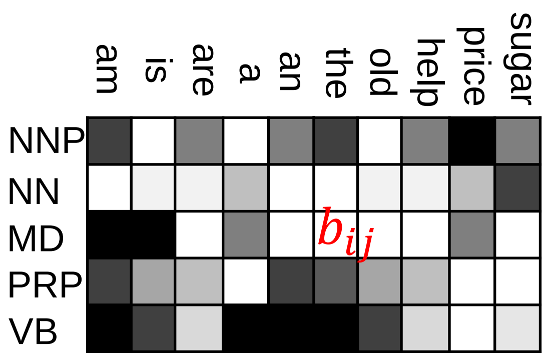
$$\hat{\pi}_i = P(h_1 = i) = \frac{\cancel{\#(h_1 = i)}}{\#(句子)}$$

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



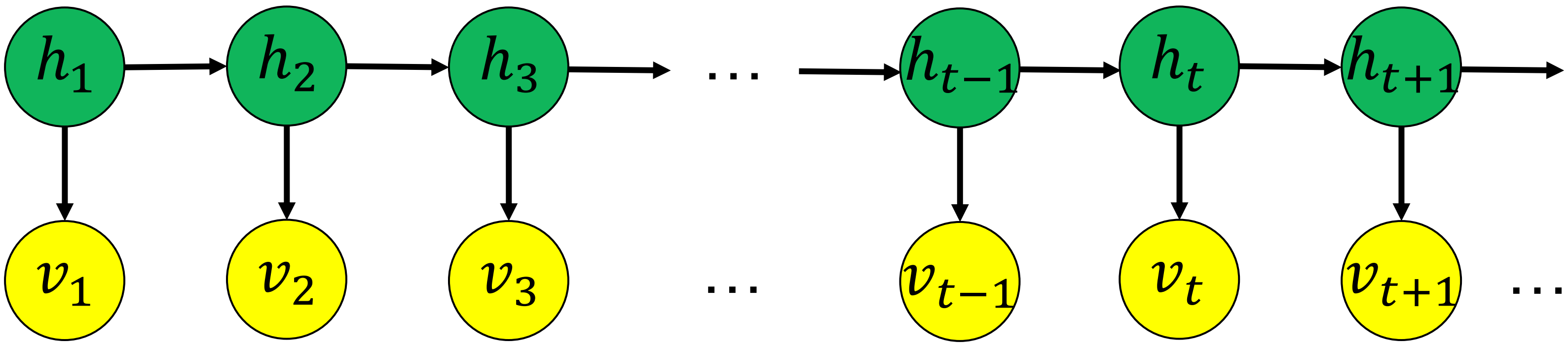
$$\widehat{a_{ij}} = P(h_{t+1} = j|h_t = i) = \frac{\cancel{\#(tags = (i,j))}}{\cancel{\#(tag = i)}}$$

$P(v_j|h_j)$



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

隐马尔科夫模型：参数的无监督学习



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

$$= \frac{\sum_{s=1}^{M} \mathbb{E}_{s:i\to j}}{\sum_{s=1}^{M} \mathbb{E}_{s:i}}$$

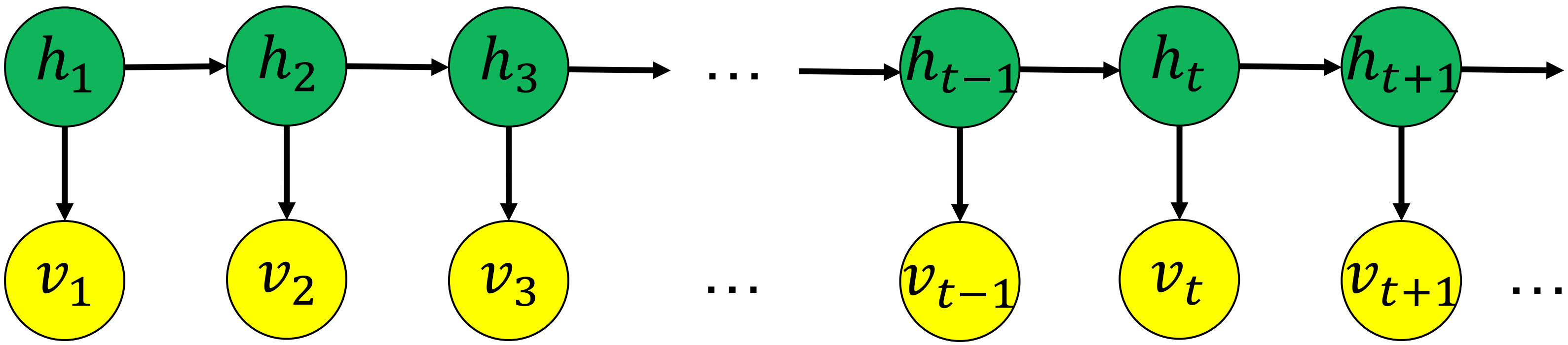
序列s中从状态i转移到j的次数的数学期望

序列s中状态i出现次数的数学期望

$$\mathbb{E}_{s:i\to j} = \sum_{t=1}^{N} \xi_t(i,j)$$

$$\mathbb{E}_{s:i} = \sum_{j=1}^{S} \sum_{t=1}^{N} \xi_t(i,j)$$

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



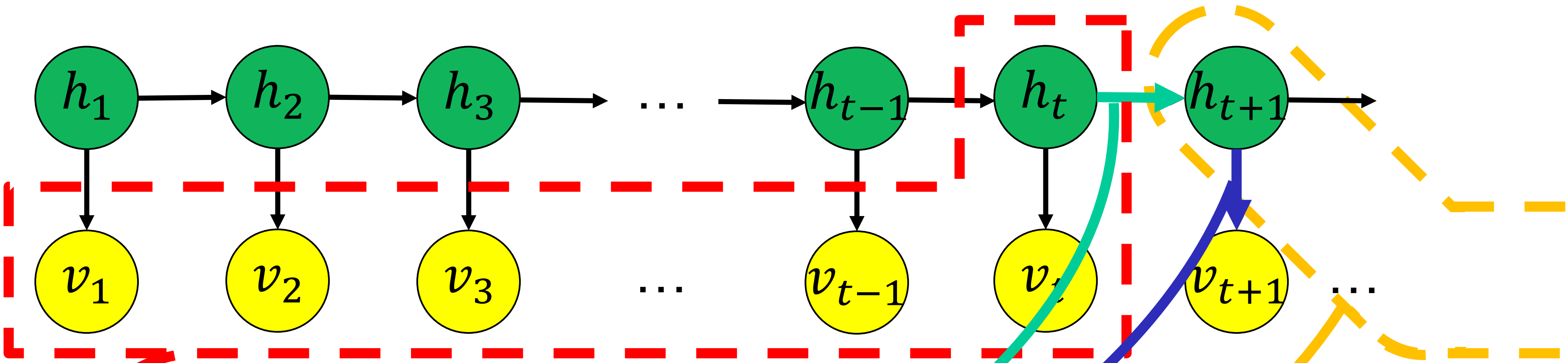
$$\xi_t(i,j) = P(h_t = i, h_{t+1} = j | v_1, v_2, \cdots, v_N)$$

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

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

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

隐马尔科夫模型：参数的无监督学习



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

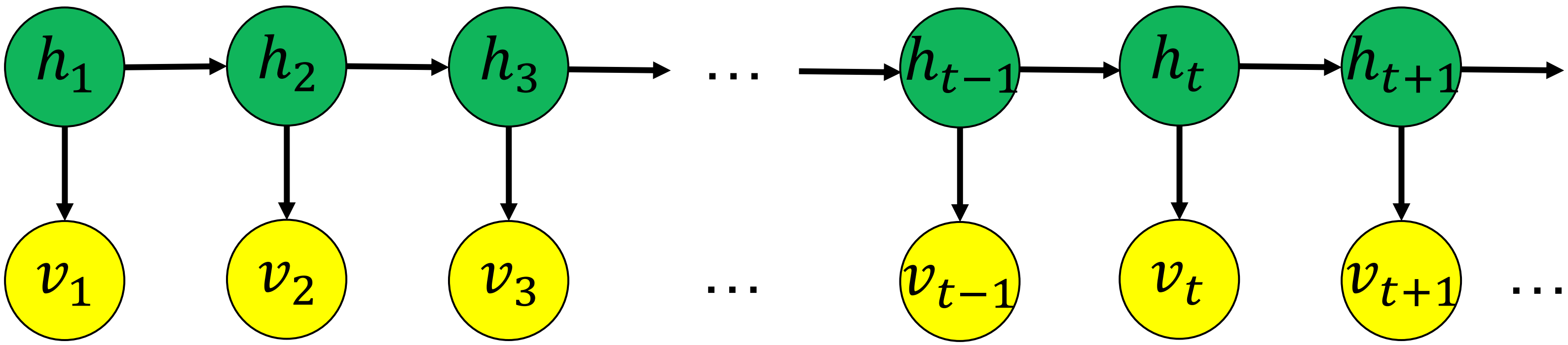
$$= \frac{P(h_t = i, h_{t+1} = j, v)}{P(v)}$$

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

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

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

隐马尔科夫模型：参数的无监督学习



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

$$= \frac{\sum_{s=1}^{M} \mathbb{E}_{s:i\to j}}{\sum_{s=1}^{M} \mathbb{E}_{s:i}}$$

$\longleftarrow$ 序列s中从状态i转移到j的次数的数学期望

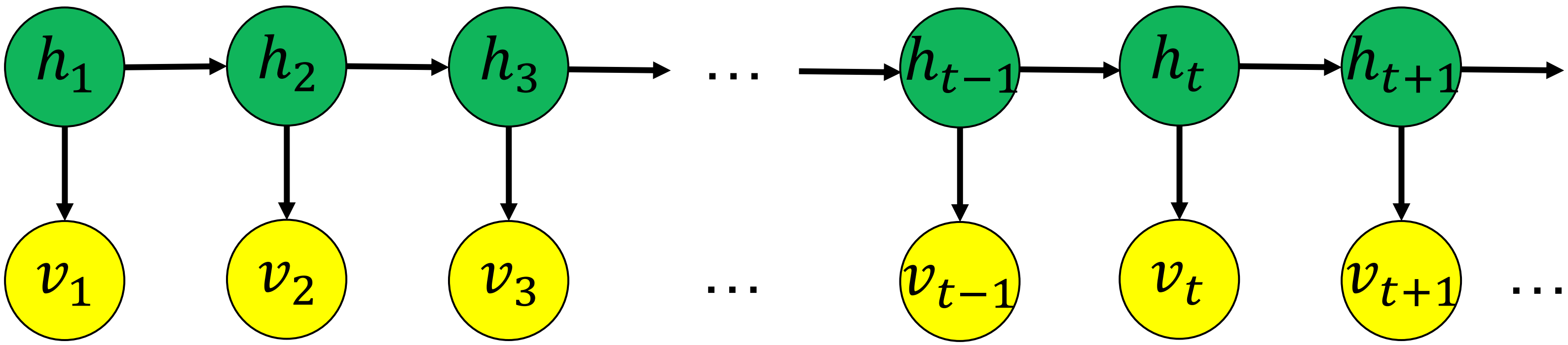
$\longleftarrow$ 序列s中状态i出现次数的数学期望

$$\mathbb{E}_{s:i\to j} = \sum_{t=1}^{N} \xi_t(i,j)$$

$$\mathbb{E}_{s:i} = \sum_{j=1}^{S} \sum_{t=1}^{N} \xi_t(i,j)$$

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

隐马尔科夫模型：参数的无监督学习



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

$$= \frac{\sum_{s=1}^{M} \mathbb{E}_{s:i\to j}}{\sum_{s=1}^{M} \mathbb{E}_{s:i}}$$

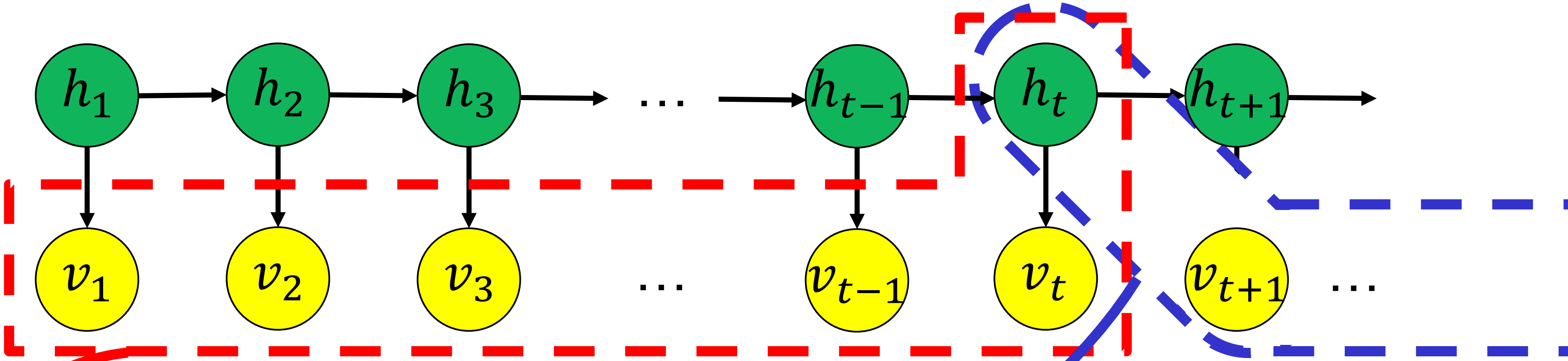
序列s中从状态i生成第j个单词的次数的数学期望

序列s中状态i出现次数的数学期望

$$\mathbb{E}_{s:i\to j} = \sum_{v_t = j} \gamma_t(i)$$

$$\mathbb{E}_{s:i} = \sum_{t=1}^{N} \gamma_t(i)$$

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



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

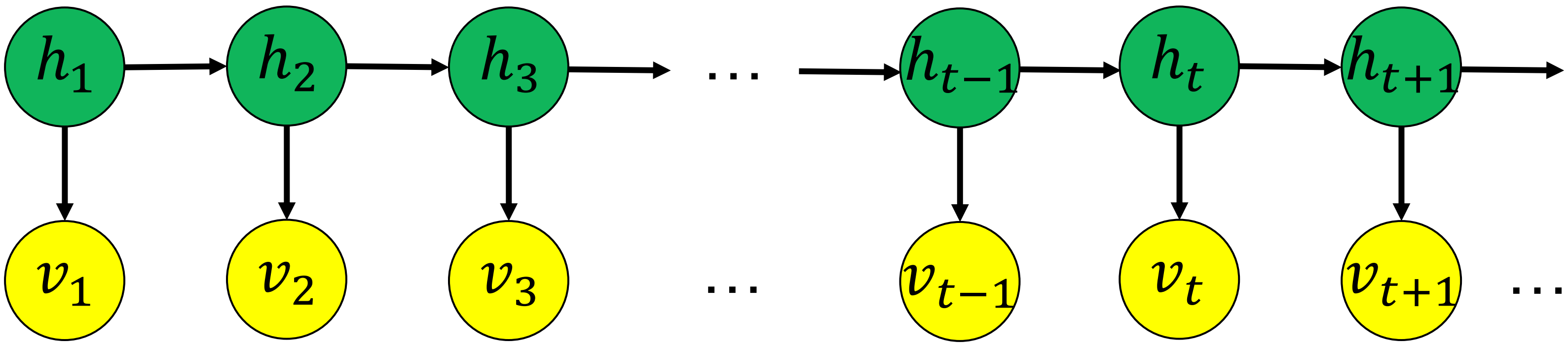
$$= \frac{P(h_t = i, v)}{P(v)}$$

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

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

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

隐马尔科夫模型：参数的无监督学习



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

$$= \frac{\sum_{s=1}^{M} \mathbb{E}_{s:i \to j}}{\sum_{s=1}^{M} \mathbb{E}_{s:i}}$$

$\leftarrow$ 序列s中从状态i生成第j个单词的次数的数学期望

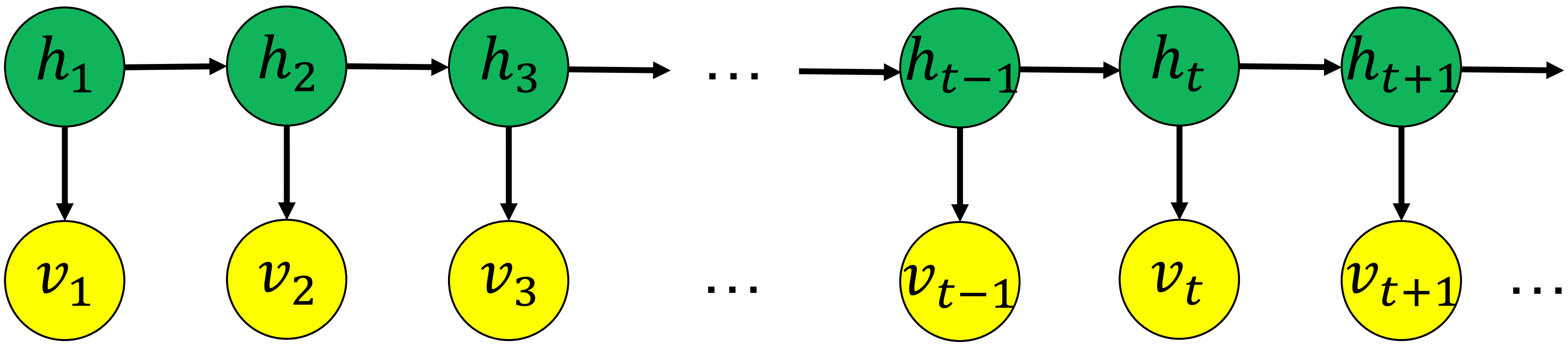
$\leftarrow$ 序列s中状态i出现次数的数学期望

$$\mathbb{E}_{s:i \to j} = \sum_{v_t = j} \gamma_t(i)$$

$$\mathbb{E}_{s:i} = \sum_{t=1}^{N} \gamma_t(i)$$

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

隐马尔科夫模型：参数的无监督学习

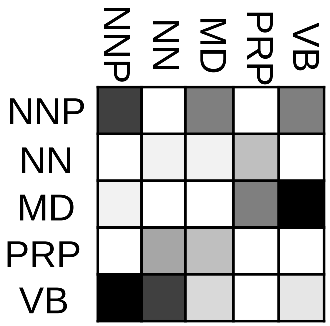


$$\widehat{\pi}_i = P(h_1 = i) = \frac{\#(h_1 = i)}{\#(\text{句子})}$$

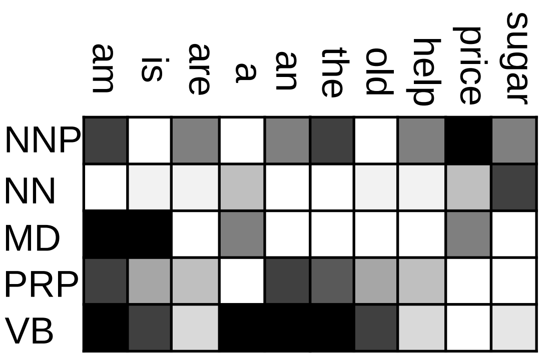
$$= \sum_{s=1}^{\mathrm{M}} \mathbb{E}_{s:h_1=i} \quad \longleftarrow \quad \text{序列}s\text{中初始状态为}i\text{的概率}$$

$$\mathbb{E}_{s:h_1=i} = \gamma_1(i) \quad \longleftarrow \quad \gamma_t(i) = P(h_t = i | v_1, v_2, \cdots, v_N)$$

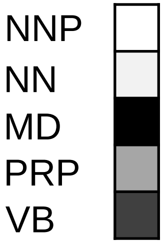
隐马尔科夫模型：参数的无监督学习



a_{ij}



b_{ij}



π_i

隐马尔科夫模型：参数的无监督学习

a_{ij}

b_{ij}

π_i

$$\begin{Bmatrix} \alpha_t(i) \\ \beta_t(i) \end{Bmatrix}$$

$$\begin{Bmatrix} \xi_t(i,j) = \dfrac{\alpha_t(i)a_{ij}b_{j,v_{t+1}}\beta_{t+1}(j)}{P(v)} \\ \gamma_t(i) = \dfrac{\alpha_t(i)\beta_t(i)}{P(v)} \end{Bmatrix}$$

$$\widehat{a_{ij}} = \frac{\sum_{s=1}^{\mathrm{M}}\sum_{t=1}^{N}\xi_t(i,j)}{\sum_{s=1}^{\mathrm{M}}\sum_{j=1}^{S}\sum_{t=1}^{N}\xi_t(i,j)}$$

$$\widehat{b_{ij}} = \frac{\sum_{s=1}^{\mathrm{M}}\sum_{v_t=j}\gamma_t(i)}{\sum_{s=1}^{\mathrm{M}}\sum_{t=1}^{N}\gamma_t(i)}$$

$$\widehat{\pi_i} = \sum_{s=1}^{\mathrm{M}}\gamma_1(i)$$

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

a_{ij}

b_{ij}

π_i

$$\begin{Bmatrix} \alpha_t(i) \\ \beta_t(i) \end{Bmatrix}$$

$$\begin{Bmatrix} \xi_t(i,j) = \dfrac{\alpha_t(i) a_{ij} b_{j,v_{t+1}} \beta_{t+1}(j)}{P(v)} \\ \gamma_t(i) = \dfrac{\alpha_t(i)\beta_t(i)}{P(v)} \end{Bmatrix}$$

$$\widehat{a_{ij}} = \frac{\sum_{s=1}^{M} \sum_{t=1}^{N} \xi_t(i,j)}{\sum_{s=1}^{M} \sum_{j=1}^{S} \sum_{t=1}^{N} \xi_t(i,j)}$$

$$\widehat{b_{ij}} = \frac{\sum_{s=1}^{M} \sum_{v_t=j} \gamma_t(i)}{\sum_{s=1}^{M} \sum_{t=1}^{N} \gamma_t(i)}$$

$$\widehat{\pi_i} = \sum_{s=1}^{M} \gamma_1(i)$$

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

a_{ij}

b_{ij}

π_i

$$\begin{Bmatrix} \alpha_t(i) \\ \beta_t(i) \end{Bmatrix}$$

$$\begin{Bmatrix} \xi_t(i,j) = \dfrac{\alpha_t(i) a_{ij} b_{j,v_{t+1}} \beta_{t+1}(j)}{P(v)} \\ \gamma_t(i) = \dfrac{\alpha_t(i)\beta_t(i)}{P(v)} \end{Bmatrix}$$

$$\widehat{a_{ij}} = \frac{\sum_{s=1}^{\mathrm{M}} \sum_{t=1}^{N} \xi_t(i,j)}{\sum_{s=1}^{\mathrm{M}} \sum_{j=1}^{S} \sum_{t=1}^{N} \xi_t(i,j)}$$

$$\widehat{b_{ij}} = \frac{\sum_{s=1}^{\mathrm{M}} \sum_{v_t=j} \gamma_t(i)}{\sum_{s=1}^{\mathrm{M}} \sum_{t=1}^{N} \gamma_t(i)}$$

$$\widehat{\pi_i} = \sum_{s=1}^{\mathrm{M}} \gamma_1(i)$$

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$\gamma_t(i)$

$$\begin{Bmatrix} \xi_t(i,j) = \dfrac{\alpha_t(i) a_{ij} b_{j,v_{t+1}} \beta_{t+1}(j)}{P(v)} \\ \gamma_t(i) = \dfrac{\alpha_t(i)\beta_t(i)}{P(v)} \end{Bmatrix}$$

M步（maximization）：给定数学期望算参数的新估值

$$\widehat{a_{ij}} = \frac{\sum_{s=1}^{M}\sum_{t=1}^{N}\xi_t(i,j)}{\sum_{s=1}^{M}\sum_{j=1}^{S}\sum_{t=1}^{N}\xi_t(i,j)} \quad \widehat{b_{ij}} = \frac{\sum_{s=1}^{M}\sum_{v_t=j}\gamma_t(i)}{\sum_{s=1}^{M}\sum_{t=1}^{N}\gamma_t(i)} \quad \widehat{\pi_i} = \sum_{s=1}^{M}\gamma_1(i)$$

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	90%~94%
HMM（Brants, 2000）	96.5%

无监督的前向-后向算法：

HMM + Baum-Welch（Johnson, 2007）	~40%