

2018/11/05 19:00 修正

# 計算言語学

## 構造分類(系列ラベリング)

東京大学生産技術研究所

吉永 直樹

site: <http://www.tkl.iis.u-tokyo.ac.jp/~ynaga/class/cl/>

# 言語をカテゴリ(クラス)の構造に分類する

## 品詞タグ付け (part-of-speech tagging)

Lucy in the sky with diamonds  
NNP IN DT NN IN NNS

単語に対する品詞の系列

## 固有表現認識 (named entity recognition)

Ringo Star has joined the Beatles  
PERSON Non-NE ORGANIZATION

固有表現のチャンクとその分類

## 単語分割 (word segmentation)

中 国 人 参 政 權  
0 1 1 0 0 1

単語境界の有無の系列

## 依存構造解析 (dependency parsing)

Ringo Star has joined the Beatles



依存構造木 or 結合操作の系列

個々のラベルの間に依存関係がある

# 言語をカテゴリ(クラス)の構造に分類する

## 品詞タグ付け (part-of-speech tagging)    固有表現認識 (named entity recognition)

Lucy in the sky with diamonds  
NNP IN DT NN IN NNS

単語に対する品詞の系列

Ringo Star has joined the Beatles  
PERSON Non-NE ORGANIZATION

固有表現のチャンクとその分類

## 単語分割 (word segmentation)

中 国 人 参 政 權  
0 1 1 0 0 1

単語境界の有無の系列

## 依存構造解析 (dependency parsing)

Ringo Star has joined the Beatles  
~~~~~  
↑↑↑↑↑

依存構造木 or 結合操作の系列

個々のラベルの間に依存関係がある

# 品詞 (Parts-of-speech, POS)

形態的・意味的・統語的な役割に基づく単語の分類

- Dionysius Thrax の先駆的定義 (対象: 古ギリシャ語)
  - 名詞 / 動詞 / 副詞 / 代名詞 / 前置詞 / 接続詞 / 分詞 / 冠詞

Open class POS  
(新語が現れる)

ex. *iPhone, google*

Closed class POS  
(語彙ほぼ固定)

- 様々な言語処理タスクの特徴量として有用  
構文解析, 固有表現抽出, 情報検索, 音声合成 etc.

言語(コーパス)ごとに品詞セットには揺れがあるが  
**言語普遍の品詞セット**を作ろうとする試みも

# Universal POS tags [Nivre+ 2016]

# Universal dependencies のために設計 16(+1) 品詞

- Open class POS: NOUN (名詞), VERB (動詞), ADJ (形容詞),  
ADV (副詞), PROPN (固有名詞), INTJ (間投詞)  
主に内容語
  - Closed class POS: ADP (前/後置詞), AUX (助動詞), DET (限定詞),  
NUM (数詞), PART (分詞), PRON (代名詞),  
CCONJ (等位接続詞), SCONJ (従位接続詞)  
主に機能語
  - Other: PUNCT (句読点), SYM (記号), X (code-switching 等への対処)

# Penn TreeBank タグセット [Marcus+ 1994]

- 英語における標準的な品詞タグセット
  - Brownコーパスで定義された87品詞を45に縮退
  - 形態素が含む情報も含有 (例: NN/NNS, JJ/JJR, VB/BBD)

| Tag  | Description                   | Example             | Tag   | Description           | Example            | Tag  | Description          | Example            |
|------|-------------------------------|---------------------|-------|-----------------------|--------------------|------|----------------------|--------------------|
| CC   | coordinating conjunction      | <i>and, but, or</i> | PDT   | predeterminer         | <i>all, both</i>   | VBP  | verb non-3sg present | <i>eat</i>         |
| CD   | cardinal number               | <i>one, two</i>     | POS   | possessive ending     | 's                 | VBZ  | verb 3sg pres        | <i>eats</i>        |
| DT   | determiner                    | <i>a, the</i>       | PRP   | personal pronoun      | <i>I, you, he</i>  | WDT  | wh-determ.           | <i>which, that</i> |
| EX   | existential ‘there’           | <i>there</i>        | PRP\$ | possess. pronoun      | <i>your, one's</i> | WP   | wh-pronoun           | <i>what, who</i>   |
| FW   | foreign word                  | <i>mea culpa</i>    | RB    | adverb                | <i>quickly</i>     | WP\$ | wh-possess.          | <i>whose</i>       |
| IN   | preposition/<br>subordin-conj | <i>of, in, by</i>   | RBR   | comparative<br>adverb | <i>faster</i>      | WRB  | wh-adverb            | <i>how, where</i>  |
| JJ   | adjective                     | <i>yellow</i>       | RBS   | superlatv. adverb     | <i>fastest</i>     | \$   | dollar sign          | \$                 |
| JJR  | comparative adj               | <i>bigger</i>       | RP    | particle              | <i>up, off</i>     | #    | pound sign           | #                  |
| JJS  | superlative adj               | <i>wildest</i>      | SYM   | symbol                | <i>+, %, &amp;</i> | "    | left quote           | ‘ or “             |
| LS   | list item marker              | <i>1, 2, One</i>    | TO    | “to”                  | <i>to</i>          | ”    | right quote          | ’ or ”             |
| MD   | modal                         | <i>can, should</i>  | UH    | interjection          | <i>ah, oops</i>    | (    | left paren           | [, (, {, <         |
| NN   | sing or mass noun             | <i>llama</i>        | VB    | verb base form        | <i>eat</i>         | )    | right paren          | ], ), }, >         |
| NNS  | noun, plural                  | <i>llamas</i>       | VBD   | verb past tense       | <i>ate</i>         | ,    | comma                | ,                  |
| NNP  | proper noun, sing.            | <i>IBM</i>          | VBG   | verb gerund           | <i>eating</i>      | .    | sent-end punc        | . ! ?              |
| NNPS | proper noun, plu.             | <i>Carolinas</i>    | VBN   | verb past part.       | <i>eaten</i>       | :    | sent-mid punc        | : ; ... - -        |

# Penn TreeBank タグセット [Marcus+ 1994]

- 英語における標準的な品詞タグセット
    - Brownコーパスで定義された87品詞を45に縮退
    - 形態素が含む情報も含有(例: NN/NNS, JJ/JJR, VB/BBD)

| Tag  | Description                   | Example             | Tag   | Description           |                |    |               |             |
|------|-------------------------------|---------------------|-------|-----------------------|----------------|----|---------------|-------------|
| CC   | coordinating conjunction      | <i>and, but, or</i> | PDT   | predeterminer         |                |    |               |             |
| CD   | cardinal number               | <i>one, two</i>     | POS   | possessive ending     |                |    |               |             |
| DT   | determiner                    | <i>a, the</i>       | PRP   | personal pronoun      |                |    |               |             |
| EX   | existential ‘there’           | <i>there</i>        | PRP\$ | possess. pronoun      |                |    |               |             |
| FW   | foreign word                  | <i>mea culpa</i>    | RB    | adverb                |                |    |               |             |
| IN   | preposition/<br>subordin-conj | <i>of, in, by</i>   | RBR   | comparative<br>adverb |                |    |               |             |
| JJ   | adjective                     | <i>yellow</i>       | RBS   | superlatv. adverb     | <i>fastest</i> | \$ | dollar sign   | \$          |
| JJR  | comparative adj               | <i>bigger</i>       | RP    | particle              | <i>up, off</i> | #  | pound sign    | #           |
| JJS  | superlative adj               | <i>wildest</i>      | SYM   | symbol                |                |    |               |             |
| LS   | list item marker              | <i>1, 2, One</i>    | TO    | “to”                  |                |    |               |             |
| MD   | modal                         | <i>can, should</i>  | UH    | interjection          |                |    |               |             |
| NN   | sing or mass noun             | <i>llama</i>        | VB    | verb base form        | <i>eat</i>     | (  | left paren    | L, (, {, <  |
| NNS  | noun, plural                  | <i>llamas</i>       | VBD   | verb past tense       | <i>ate</i>     | )  | right paren   | ], ), }, >  |
| NNP  | proper noun, sing.            | <i>IBM</i>          | VBG   | verb gerund           | <i>eating</i>  | ,  | comma         | ,           |
| NNPS | proper noun, plu.             | <i>Carolinas</i>    | VBN   | verb past part.       | <i>eaten</i>   | .  | sent-end punc | . ! ?       |
|      |                               |                     |       |                       | :              |    | sent-mid punc | : ; ... - - |

# 品詞タグ付け

与えられた文中の単語の品詞の曖昧性を解消

入力 (単語列) Lucy in the sky with diamonds

出力 (品詞列) NNP IN DT NN IN NNS

コーパス中で間接的に定義

- タスクの難しさは言語(品詞タグ)やコーパスにより異なる

| Type (種類数):  | English (PTB) | English (Brown) | Japanese (Kyoto) |
|--------------|---------------|-----------------|------------------|
| 一意 (1 tag)   | 44,432 (86%)  | 45,799 (85%)    | 37,170 (94%)     |
| 曖昧 (2+ tags) | 7,025 (14%)   | 8,050 (15%)     | 2,261 (6%)       |

Token (出現数):

|              |               |               |               |
|--------------|---------------|---------------|---------------|
| 一意 (1 tag)   | 577,421 (45%) | 384,349 (33%) | 446,427 (46%) |
| 曖昧 (2+ tags) | 711,780 (55%) | 786,646 (67%) | 526,467 (54%) |

名詞・動詞間の曖昧性

高頻度の機能語(助詞・接尾辞),  
固有名詞の曖昧性や注釈ミス

# 系列ラベリング問題

- 入出力が系列で表現される、構造分類の一種
  - 入力の一要素(単語)に対し、一つのラベルが対応

品詞タグ付け (part-of-speech tagging)

Lucy in the sky with diamonds  
NNP IN DT NN IN NNS

- 入力の複数の要素(句)に対し、一つのラベルが対応

固有表現認識 (named entity recognition)

Ringo Star has joined the Beatles  
PERSON Non-NE ORGANIZATION  
B-PER I-PER O O B-ORG I-ORG

Semi-markov CRF [Sarawagi+2004] などでモデル化可能だが  
実際にはBIOタグ付けを用いて単純化する場合が多い

# 教師あり学習に基づく構造分類

- タスク: 入力構造  $X$  をクラス構造  $Y$  に分類する

- 出力構造: ラベル列, 木, グラフ

- 例) 品詞分析: 単語列 → 品詞列

- 生成モデル: 同時確率  $P(X, Y)$  をモデル化

$$\begin{aligned}\hat{Y} &= \operatorname{argmax}_{Y \in \mathcal{Y}} P(Y|X) = \operatorname{argmax}_{Y \in \mathcal{Y}} \frac{P(X|Y)P(Y)}{P(Y)} \\ &= \operatorname{argmax}_{Y \in \mathcal{Y}} \frac{\textcolor{red}{P(X|Y)}}{\textcolor{red}{\text{尤度}}} \frac{\textcolor{blue}{P(Y)}}{\textcolor{blue}{\text{事前確率}}}\end{aligned}$$

- 識別モデル: 条件付き確率  $P(Y|X)$  を直接モデル化

$$\hat{Y} = \operatorname{argmax}_{Y \in \mathcal{Y}} P(Y|X)$$

# 教師あり学習に基づく構造分類

- タスク: 入力構造  $X$  をクラス構造  $Y$  に分類する
  - 出力構造: ラベル列, 木, グラフ  
例) 品詞分析: 単語列 → 品詞列
- 生成モデル: 同時確率  $P(X, Y)$  をモデル化

$$\begin{aligned}\hat{Y} &= \operatorname{argmax}_{Y \in \mathcal{Y}} P(Y|X) = \operatorname{argmax}_{Y \in \mathcal{Y}} \frac{P(X|Y)P(Y)}{P(Y)} \\ &= \operatorname{argmax}_{Y \in \mathcal{Y}} \frac{\textcolor{red}{P(X|Y)}}{\textcolor{blue}{P(Y)}}\end{aligned}$$

尤度 事前確率

- 識別:  $Y$  を構成する ラベル間の依存性 をどう考えるか?

$$\hat{Y} = \operatorname{argmax}_{Y \in \mathcal{Y}} P(Y|X)$$

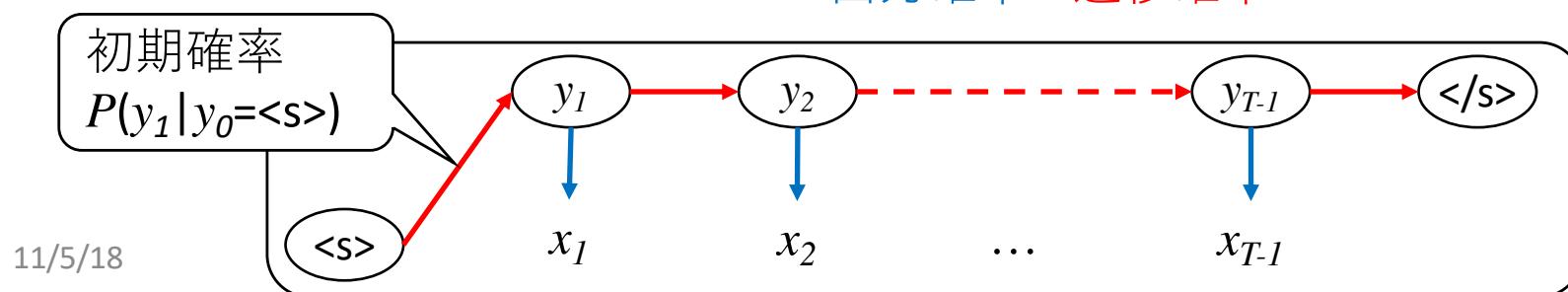
# Hidden Markov Model (HMM, 隠れマルコフモデル)

- マルコフ連鎖に基づく単語/ラベル列の生成モデル
  - 単語列/ラベル列の確率をマルコフ連鎖で単純化

$$\begin{aligned}
 P(Y = y_1^T | X = x_1^T) &= \frac{P(x_1^T | y_1^T)P(y_1^T)}{P(x_1^T)} = P(x_1^T | y_1^T)P(y_1^T) \\
 &= P(x_1^T | y_1^T)P(y_1)P(y_2 | y_1^1)P(y_3 | y_1^2) \dots P(y_T | y_1^{T-1}) \\
 &\approx P(x_1^T | y_1^T) \prod_{t=1}^T P(y_t | y_{t-1}) \quad \text{ラベル列へのマルコフ性の仮定(マルコフ連鎖)}
 \end{aligned}$$

出力列に対する独立性の仮定

$$\begin{aligned}
 &= P(x_1 | y_1^T)P(x_2 | x_1^1 y_1^T) \dots P(x_T | x_1^{T-1} y_1^T) \prod_{t=1}^T P(y_t | y_{t-1}) \\
 &\approx \prod_{t=1}^T \underbrace{P(x_t | y_t)}_{\text{出力確率}} \underbrace{P(y_t | y_{t-1})}_{\text{遷移確率}}
 \end{aligned}$$



# 隠れマルコフモデルの推定

- HMMのパラメタ(遷移確率, 出力確率)は注釈付きコーパスから**最尤推定**により計算

## 注釈付きコーパス

Lucy in the sky with diamonds  
NNP IN DT NN IN NNS

Paint the sky with stars  
VB DT NN IN NNS

Bill will sky a ball  
NNP MD VB DT NN

## 遷移確率

$$\hat{P}(y|y') = \frac{C(y, y')}{\sum_y C(y, y')}$$

$$\hat{P}(\text{NNP}|\text{<s>}) = \frac{2}{3} = 0.6\dot{6}$$

$$\hat{P}(\text{IN}|\text{NNP}) = \frac{1}{2} = 0.\dot{5}$$

$$\hat{P}(\text{DT}|\text{IN}) = \frac{1}{3} = 0.3\dot{3}$$

...

$$\hat{P}(\text{</s>}|\text{NN}) = \frac{1}{3} = 0.3\dot{3}$$

## 出力確率

$$\hat{P}(x|y) = \frac{C(x, y)}{\sum_x C(x, y)}$$

$$\hat{P}(\text{Lucy}|\text{NNP}) = \frac{1}{2} = 0.5$$

$$\hat{P}(\text{in}|\text{IN}) = \frac{1}{3} = 0.3\dot{3}$$

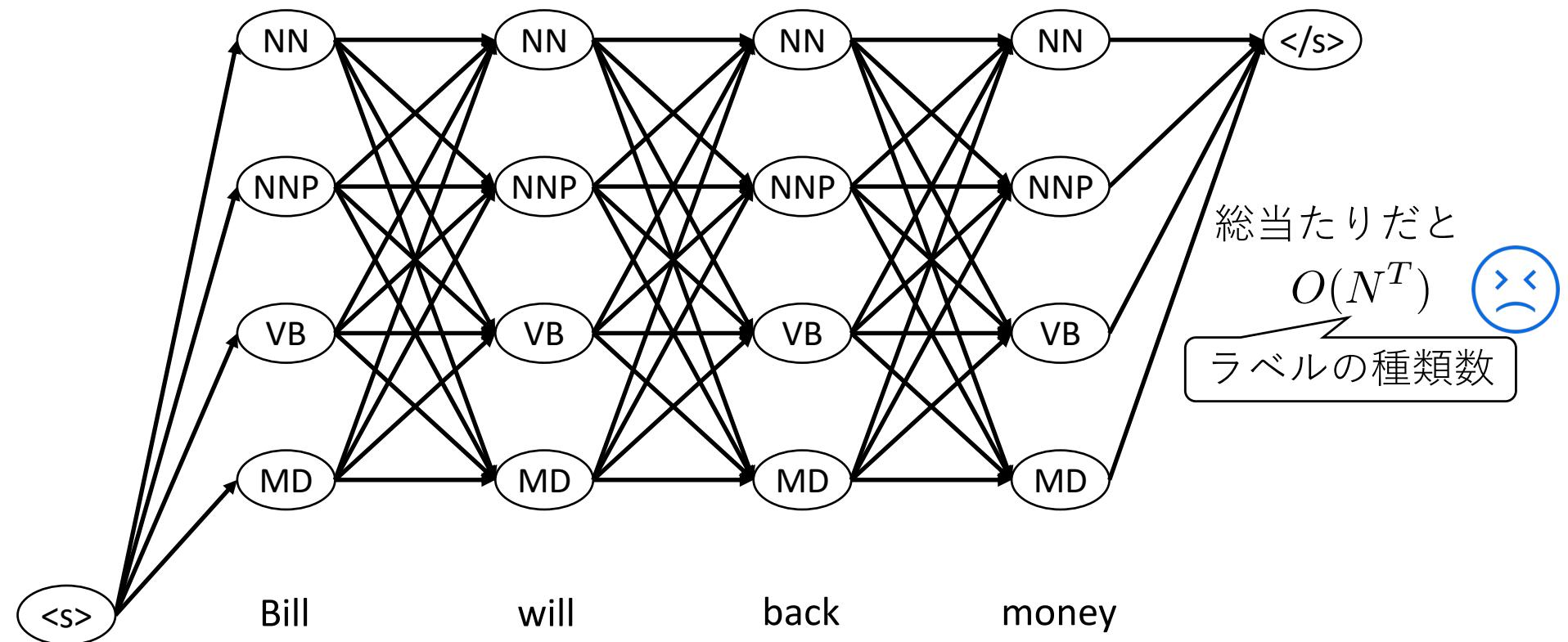
$$\hat{P}(\text{the}|\text{DET}) = \frac{2}{3} = 0.6\dot{6}$$

...

# 復号化問題 (decoding): 最尤ラベル列の推定

- 構造分類 = 超多クラス分類
  - 可能なラベル列は系列長  $T$  に対して **指数爆発**

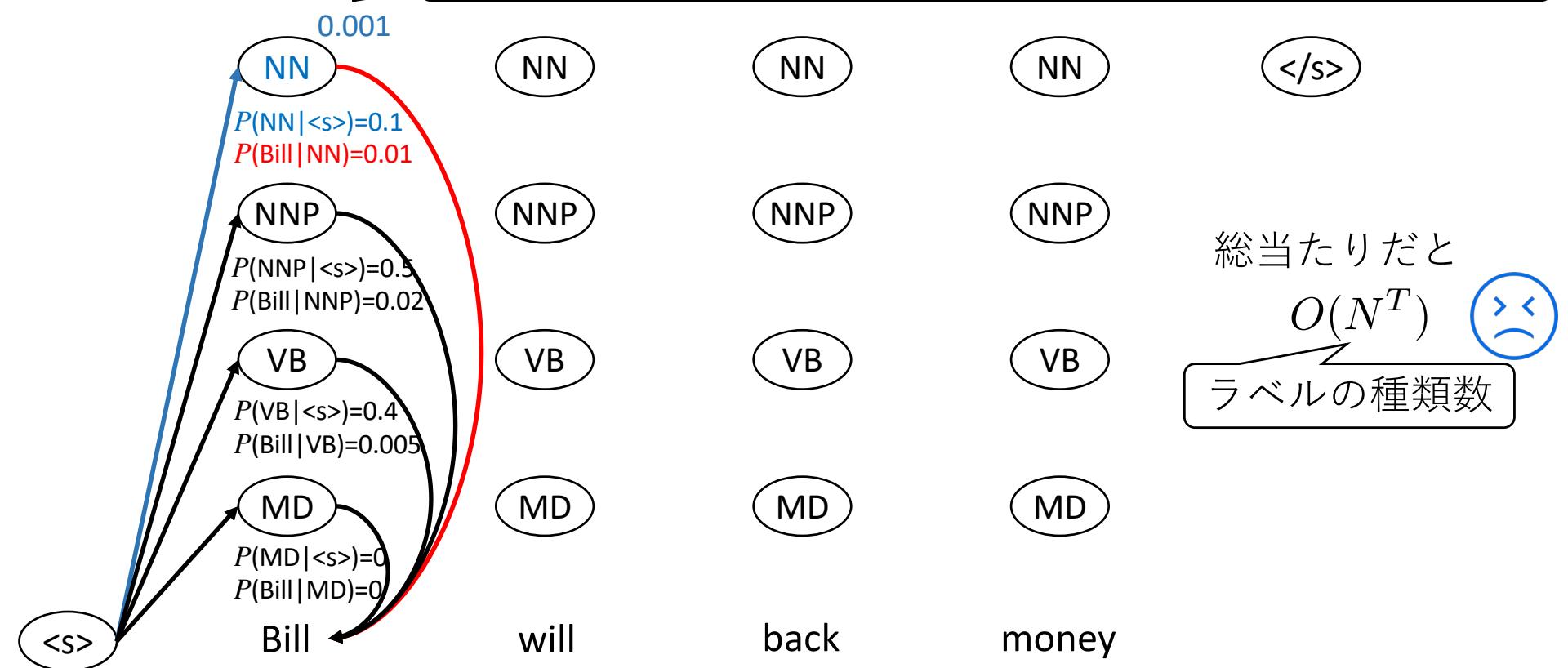
$$\hat{y}_1^T = \operatorname{argmax}_{y_1^T \in \mathcal{Y}} \hat{P}(y_1^T | x_1^T) \approx \operatorname{argmax}_{y_1^T \in \mathcal{Y}} \prod_{t=1}^T \hat{P}(x_t | y_t) \hat{P}(y_t | y_{t-1})$$



# Viterbi アルゴリズム

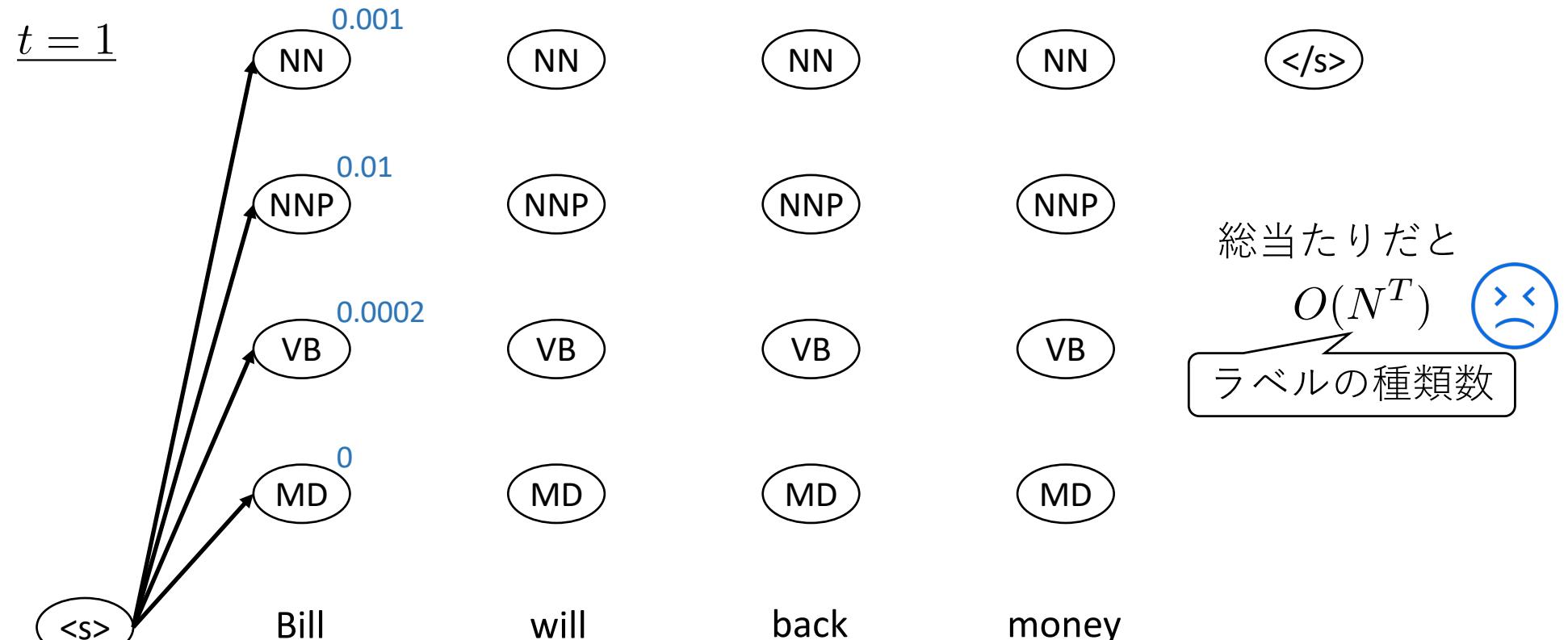
- 動的計画法に基づき効率的に最尤ラベル列を推定
  - 時刻  $t$  の各状態に至る最尤ラベル列と確率を構成的に計算

$$P(\text{Bill}, \text{NN}) = P(\text{NN}|\langle s \rangle)P(\text{Bill}|\text{NN}) = 0.1 \cdot 0.01 = 0.001$$



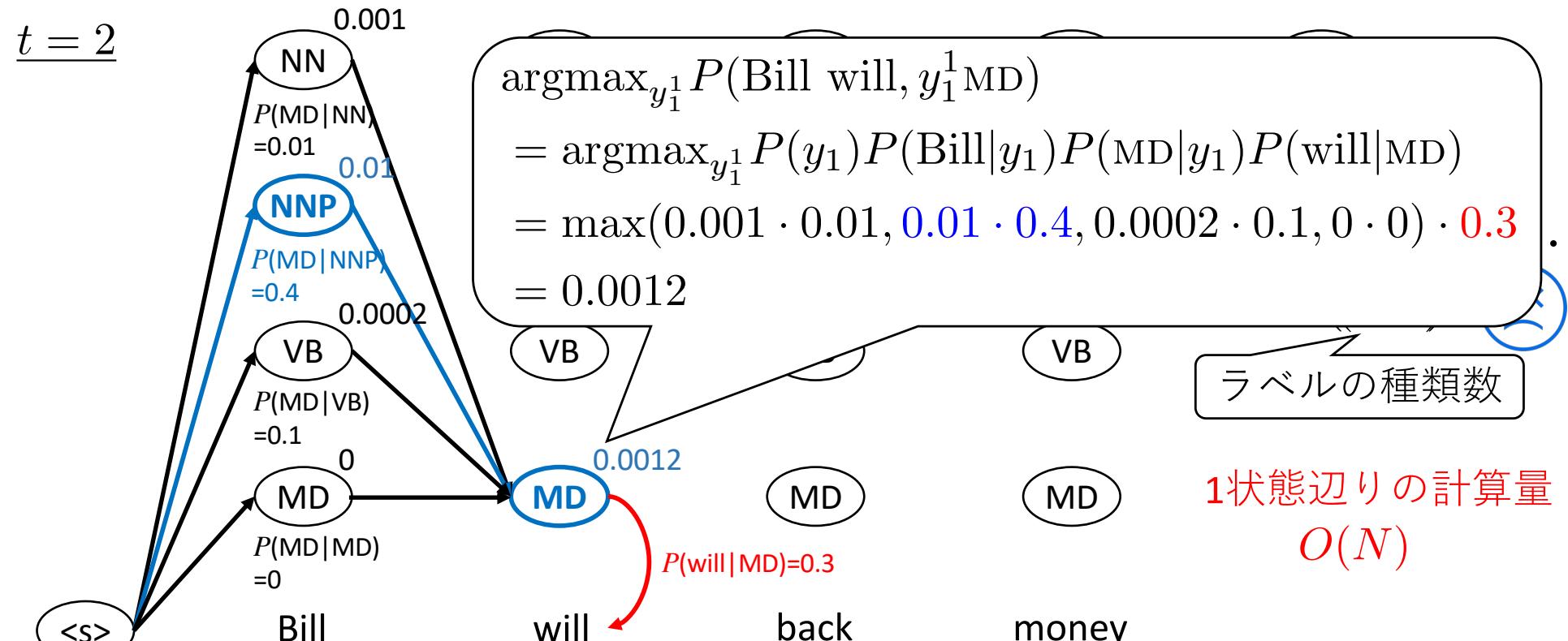
# Viterbi アルゴリズム

- 動的計画法に基づき効率的に最尤ラベル列を推定
  - 時刻  $t$  の各状態に至る最尤ラベル列と確率を構成的に計算



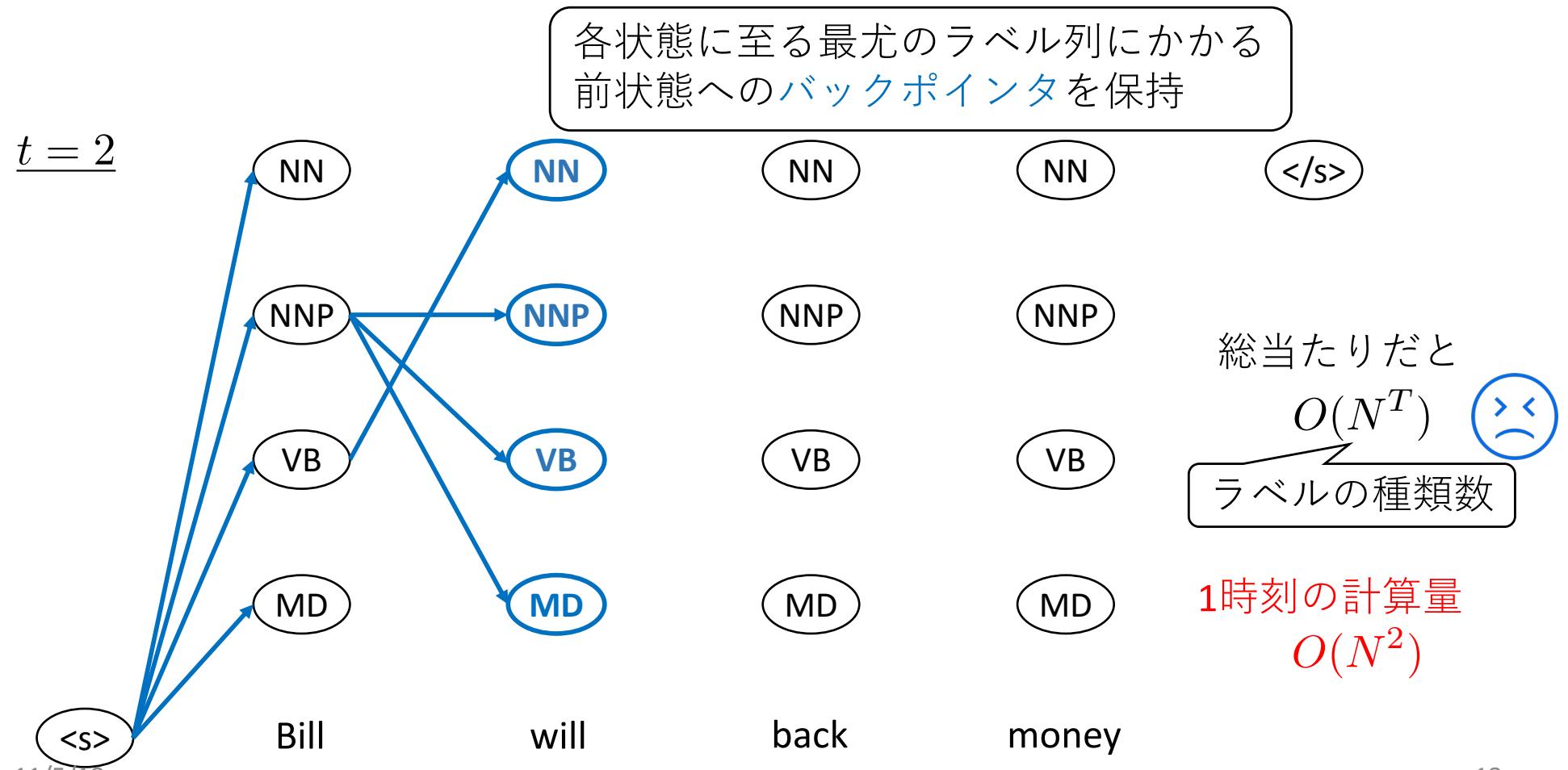
# Viterbi アルゴリズム

- 動的計画法に基づき効率的に最尤ラベル列を推定
  - 時刻  $t$  の各状態に至る最尤ラベル列と確率を構成的に計算



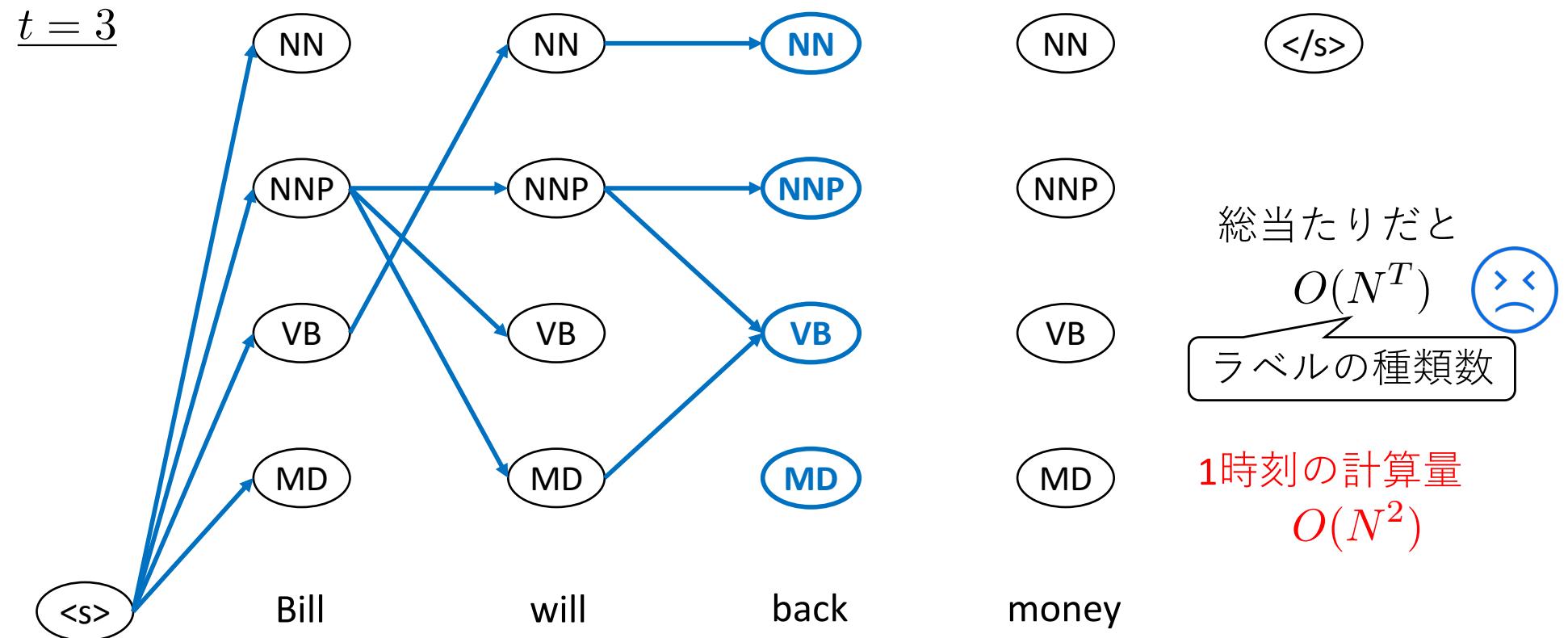
# Viterbi アルゴリズム

- 動的計画法に基づき効率的に最尤ラベル列を推定
  - 時刻  $t$  の各状態に至る最尤ラベル列と確率を構成的に計算



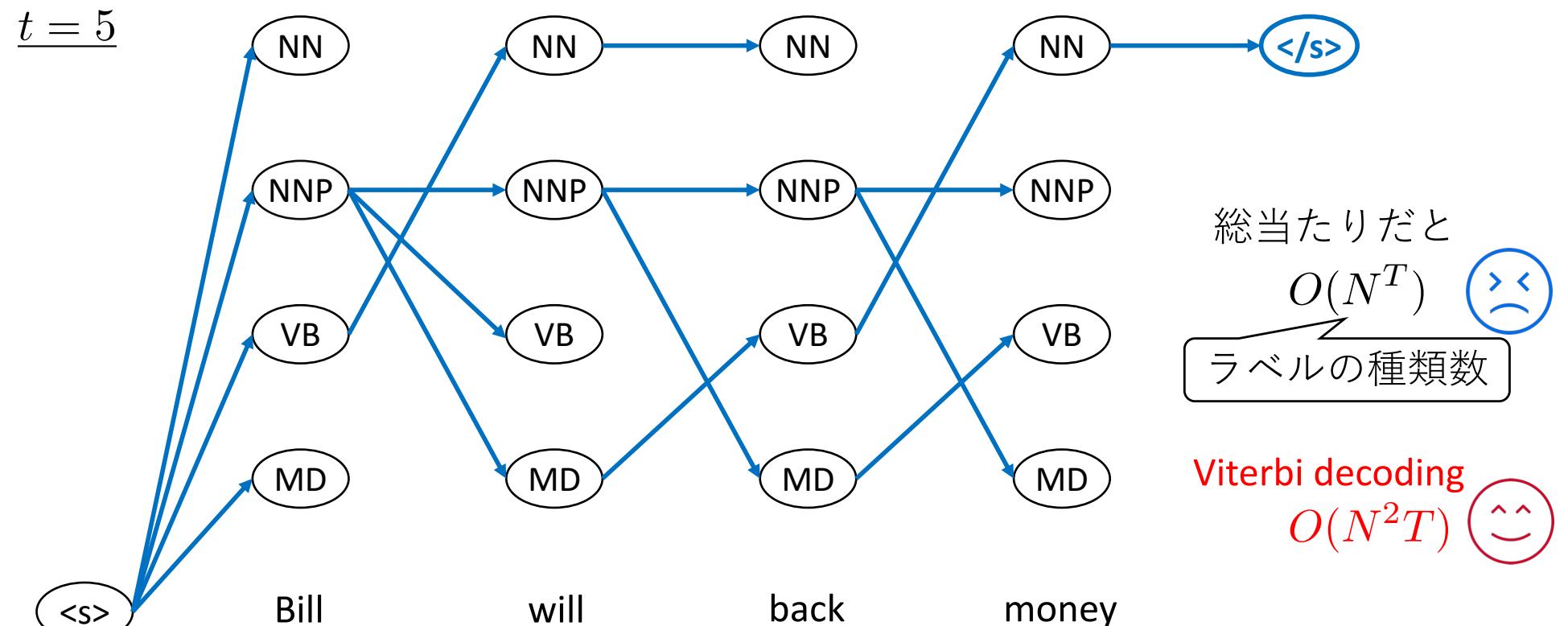
# Viterbi アルゴリズム

- 動的計画法に基づき効率的に最尤ラベル列を推定
  - 時刻  $t$  の各状態に至る最尤ラベル列と確率を構成的に計算



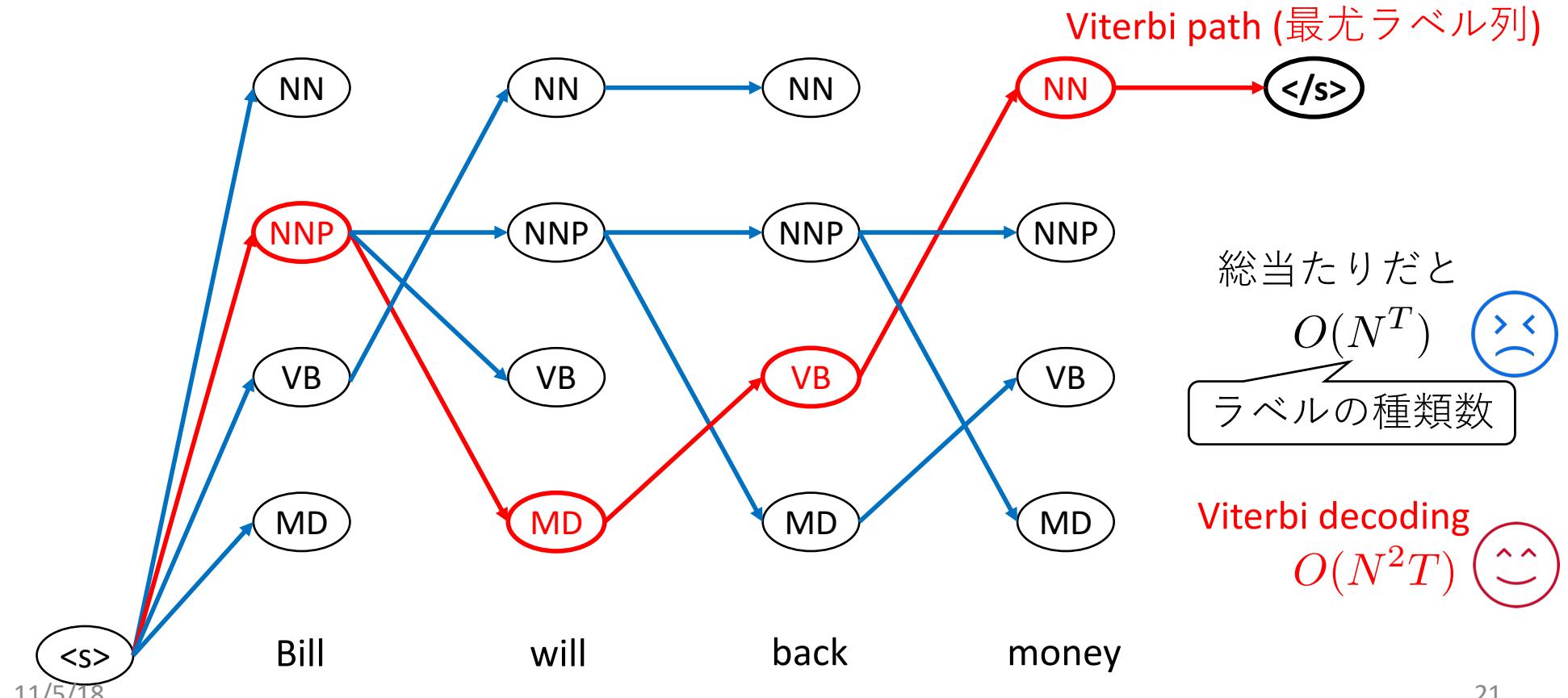
# Viterbi アルゴリズム

- 動的計画法に基づき効率的に最尤ラベル列を推定
  - 時刻  $t$  の各状態に至る最尤ラベル列と確率を構成的に計算



# Viterbi アルゴリズム

- 動的計画法に基づき効率的に最尤ラベル列を推定
  - 時刻  $t$  の各状態に至る最尤ラベル列と確率を構成的に計算
  - 文末から前状態を遡って最尤ラベル列を復元

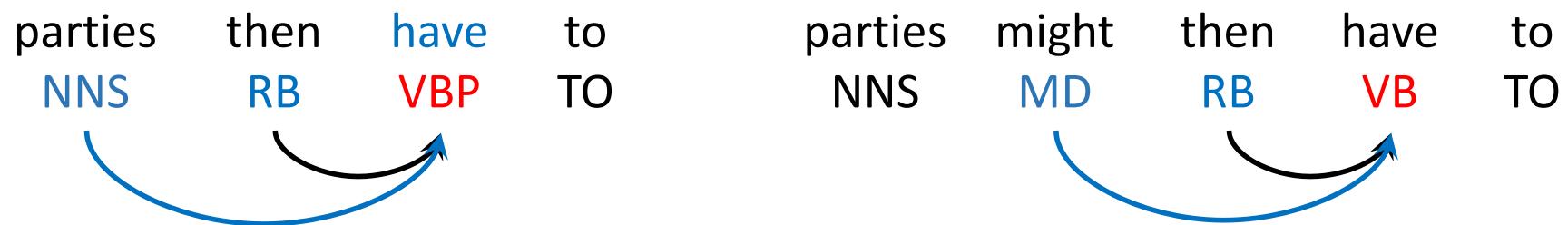


# Trigram HMM

- HMM でラベル間のマルコフ性の仮定を緩める
  - bigram → trigram

$$P(X, Y) = \prod_{t=1}^N P(y_t | y_{t-1} y_{t-2}) P(x_t | y_t)$$

- 品詞タグ付けて考慮できる問題(例)



ただし計算量は  $O(N^3 L)$ !

# 未知ラベル n-gram

- 遷移確率  $P(y_t|y_{t-n+1}^{t-1})$  は学習コーパスが小さいときや考慮する履歴  $n$  が増えると 0 になる
- 対策: 線形補間

$$P(y_t|y_{t-1}y_{t-2}) = \lambda_3 P(y_t|y_{t-1}y_{t-2}) + \lambda_2 P(y_t|y_{t-1}) + \lambda_1 P(y_t)$$

$$\lambda_3 + \lambda_2 + \lambda_1 = 1$$

- $\lambda_i$  は削除補完法 [Jelinek+ 1980] などで求める

# 未知語問題(またしても！)

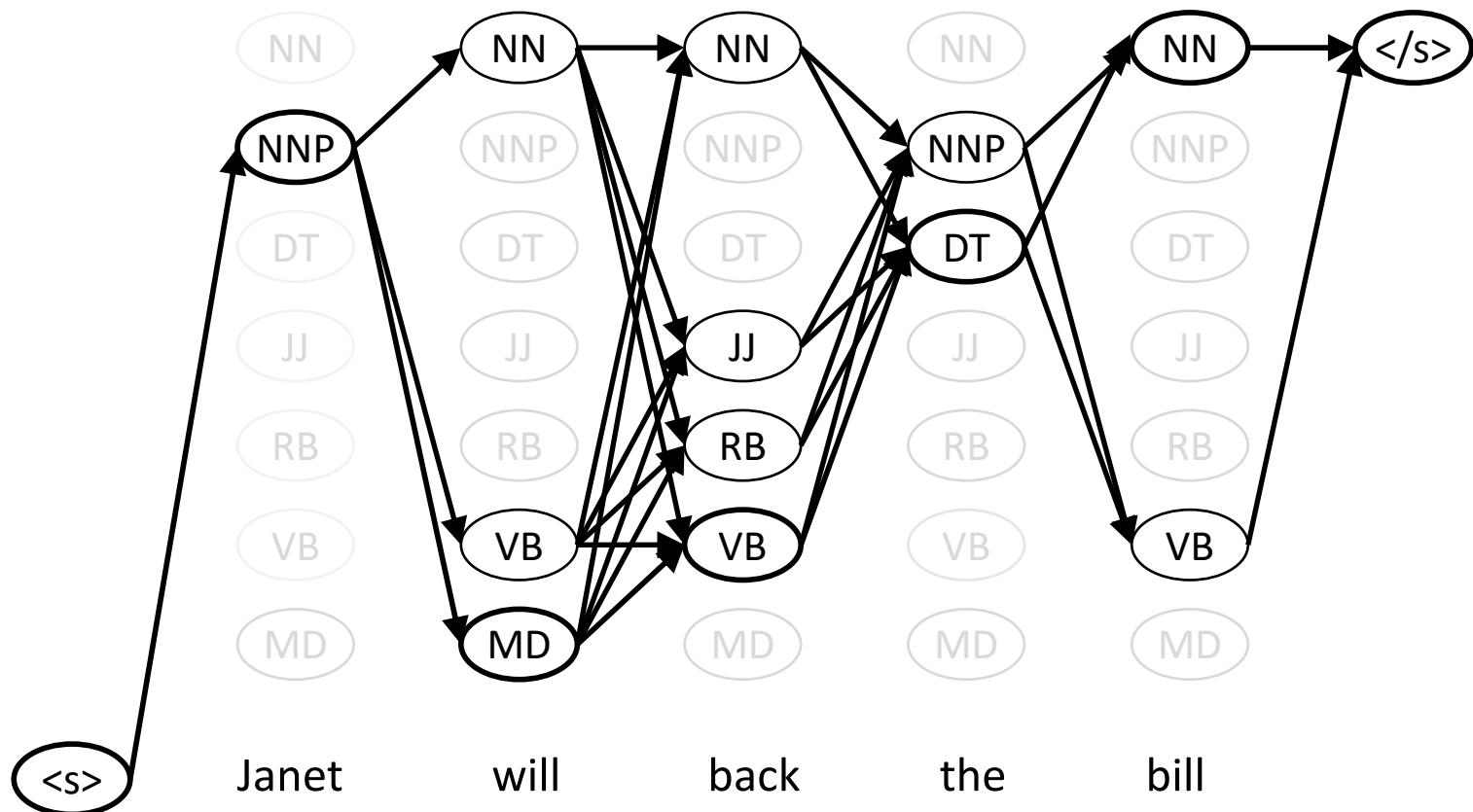
- 出力確率  $P(x_t|y_t)$  は簡単に 0 になる
  - 1.  $x_t$  が学習データにない未知語 (固有名詞, 略語 etc.)
  - 2.  $y_t$  が低頻度のタグ  $((x_t, y_t))$  が学習データにない
- 対策: 表層に基づく未知語クラスの出現確率への  
バックオフ
  - アイデア: 表層の特徴 (接尾辞など) で語の品詞は類推可
  - 実装: 低頻度語 (open-class POS) について表層の特徴ごとにバックオフ確率を計算しておいて未知語に使用

$$P(\text{rectification}|\text{NN}) \rightarrow P(\text{-tion}|\text{NN})$$

低頻度語に対する確率の  
総和を未知語クラスに分配

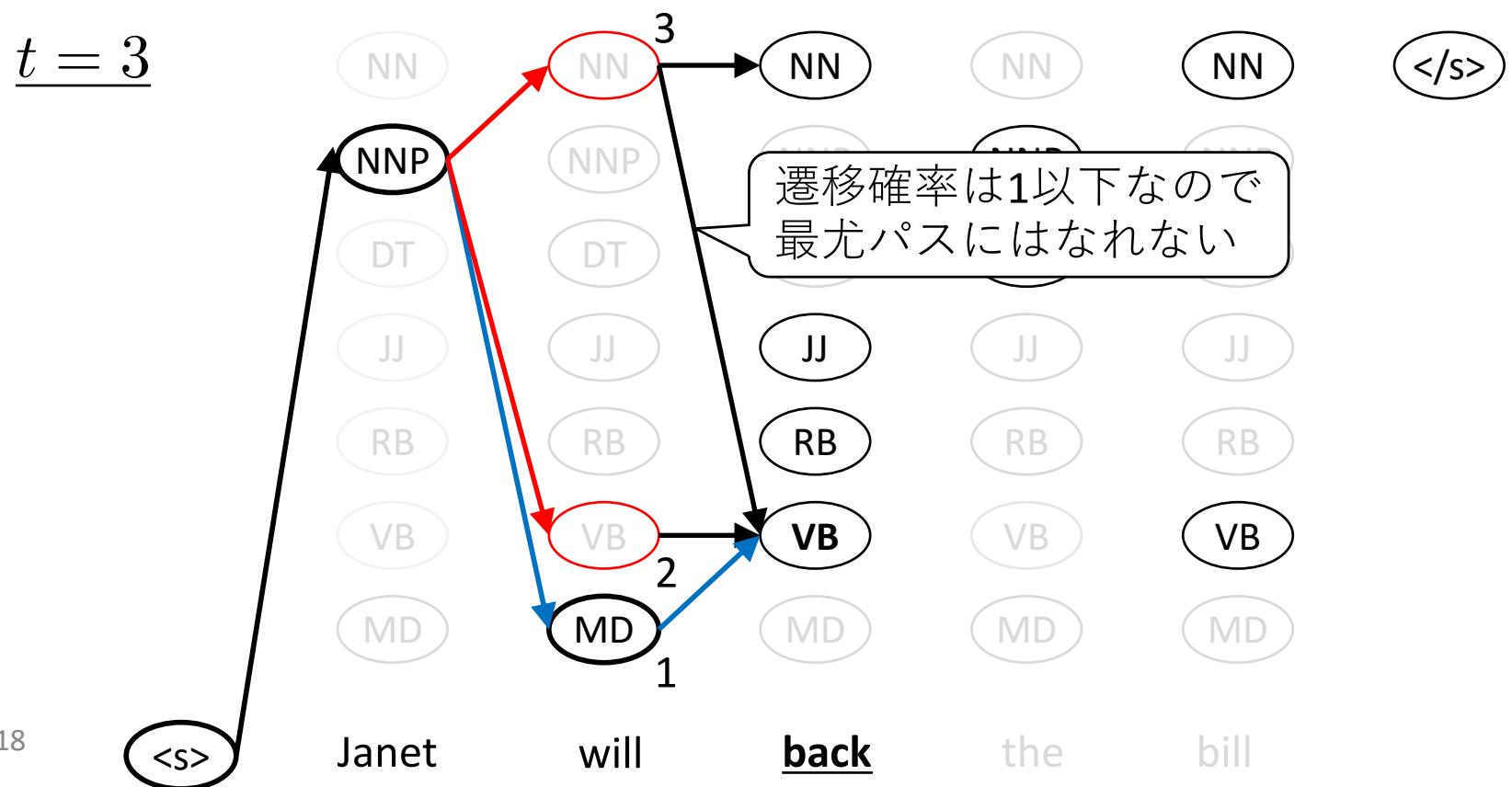
# 高速な復号化 (1/4): 出力確率による枝刈り

- 出力確率が非零の状態のみ確率を計算
  - 既知語の出力確率をスムージングしないなら極めて有効
  - 辞書などを用いて単語が取る品詞を制限する場合も



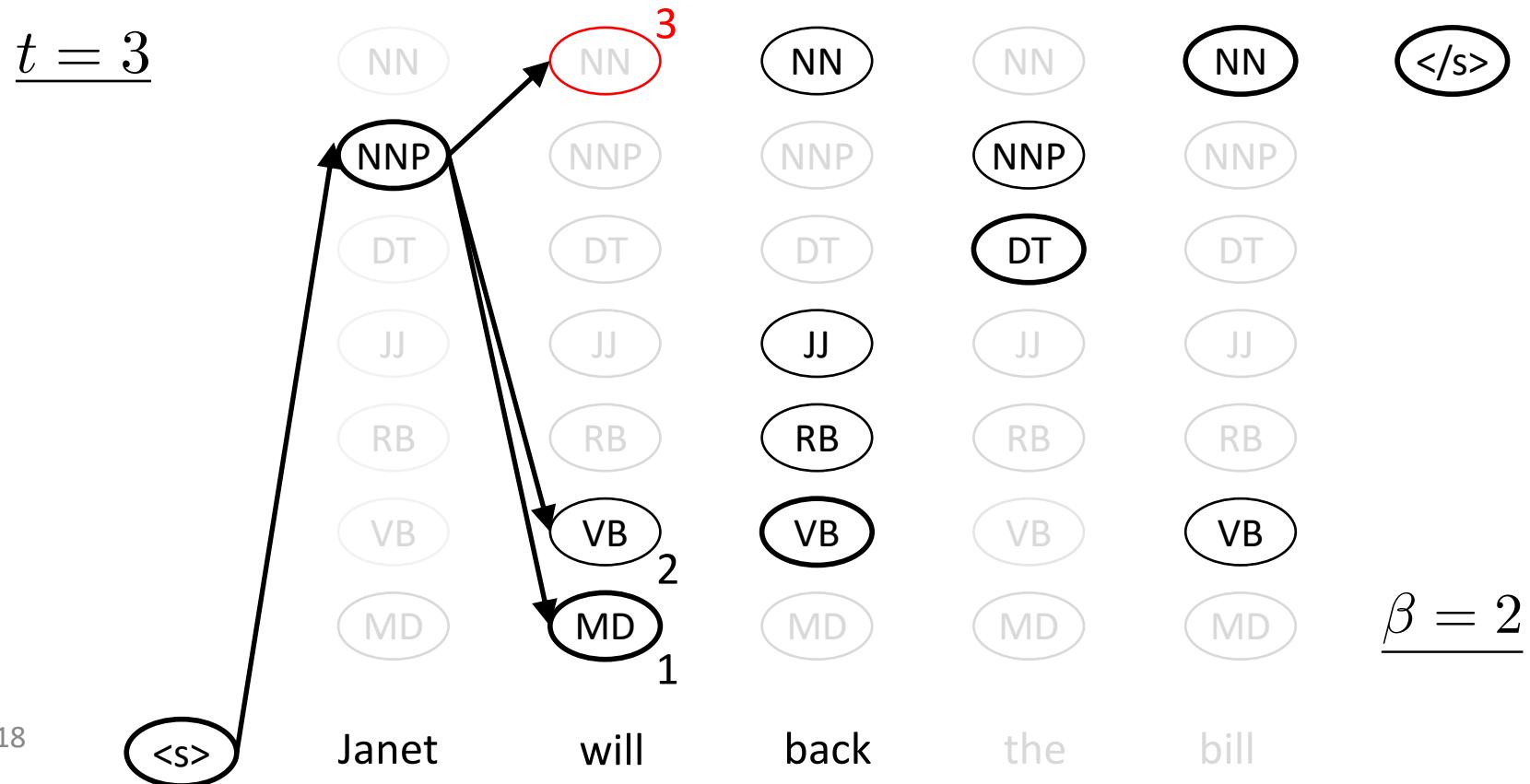
## 高速な復号化 (2/4): 遷移確率による枝刈り

- ・時間  $t$  の確率計算を遷移確率に基づいて枝刈り
    - ・最尤ラベル列の確率で前状態をソート
    - ・前状態までの確率が現状態に対する計算済の最大確率を下回れば残りの前状態に関する計算を枝刈りできる



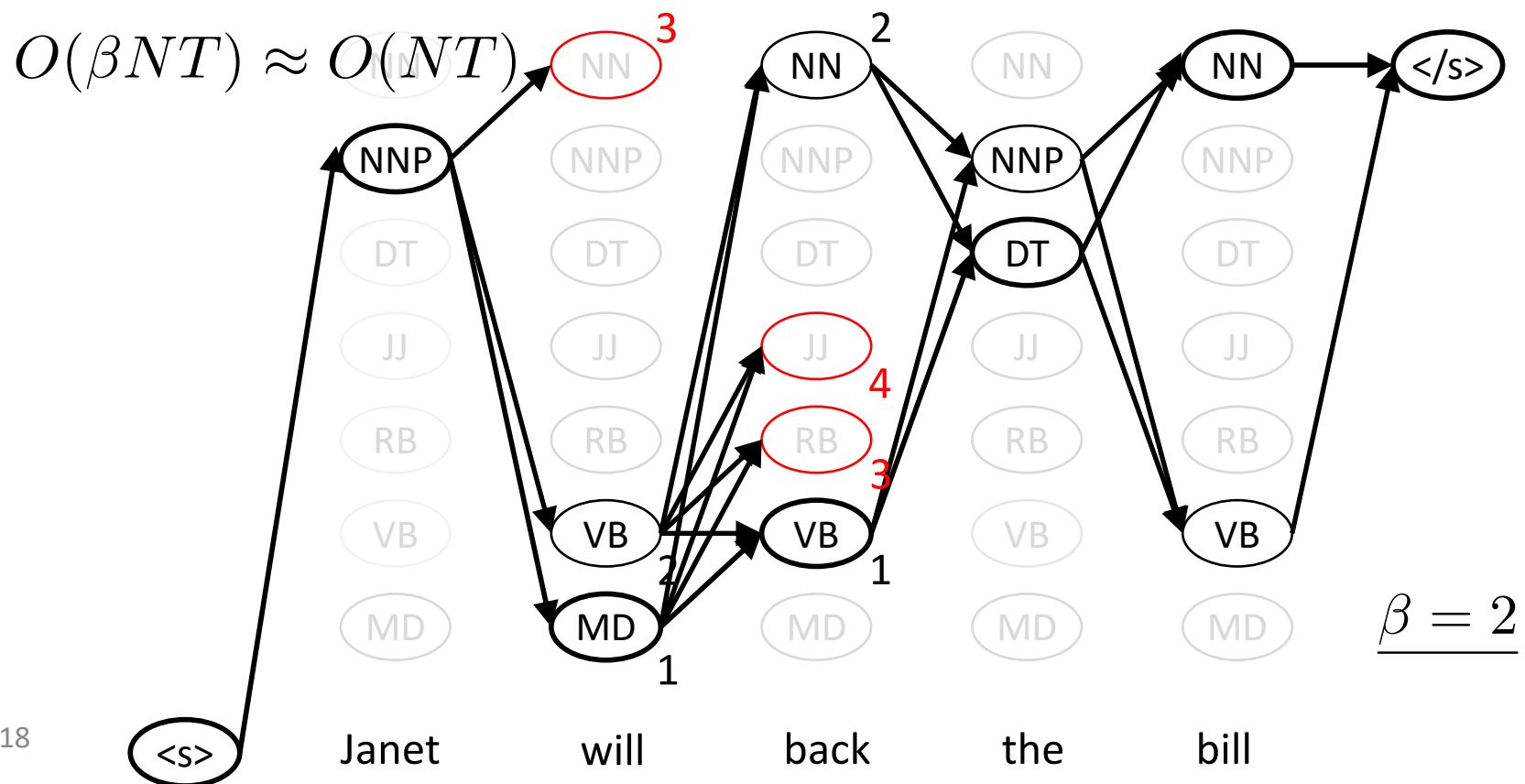
# 高速な復号化 (3/4): ビームサーチ

- Viterbi アルゴリズムでも遅いときのための近似手法  
 $O(N^k T)$ : ラベル数  $N$  が多い or 考慮する文脈長  $k$  が長い
  - 時間  $t$  で考慮する前状態を確率上位  $\beta$  個に制限  
ビーム幅



# 高速な復号化 (3/4): ビームサーチ

- Viterbi アルゴリズムでも遅いときのための近似手法  
 $O(N^k T)$ : ラベル数  $N$  が多い or 考慮する文脈長  $k$  が長い
  - 時間  $t$  で考慮する前状態を確率上位  $\beta$  個に制限  
ビーム幅



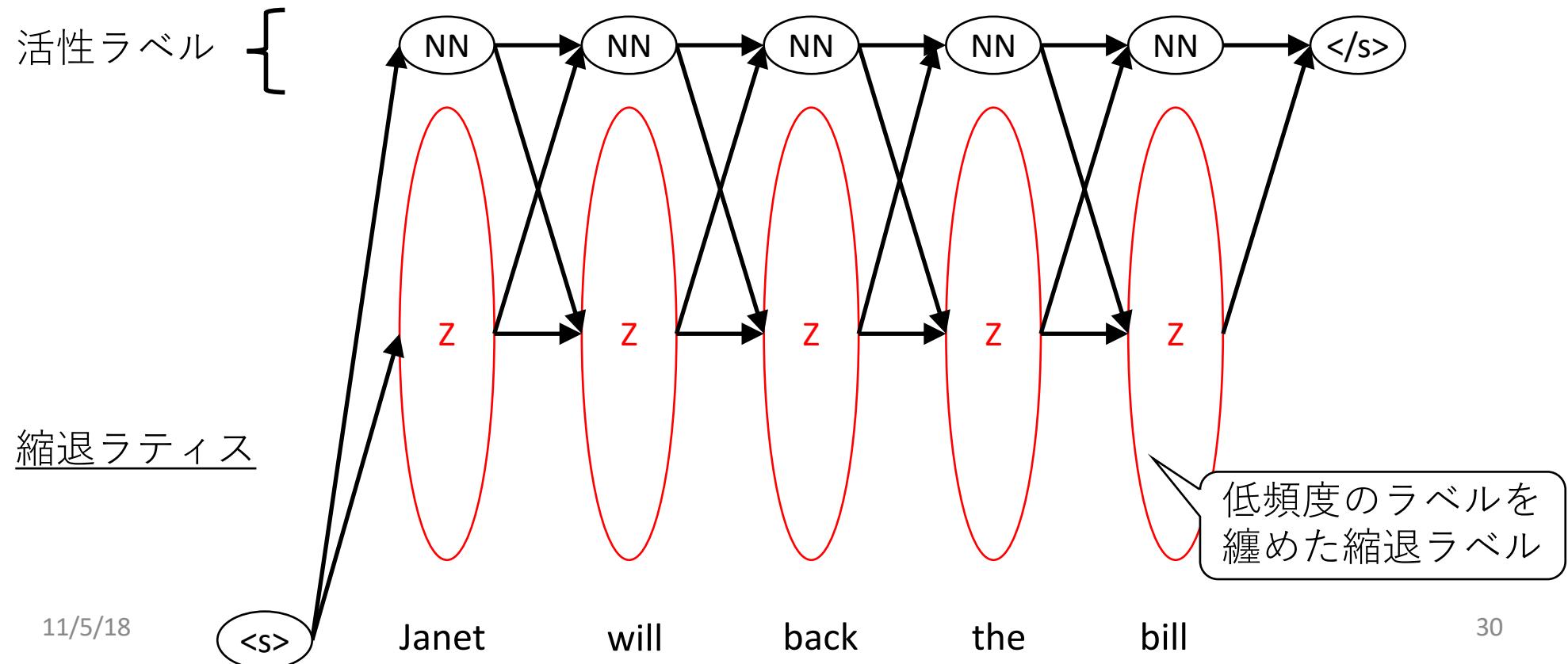
# 高速な復号化 (4/4): Staggered Decoding [Kaji+ 2010]

- ラベルのまとめ上げと枝刈りによる漸次の復号化
  - 複数のラベルをまとめた縮退ラベルに基づくラティス(縮退ラティス)上で最尤系列を探索し枝刈り



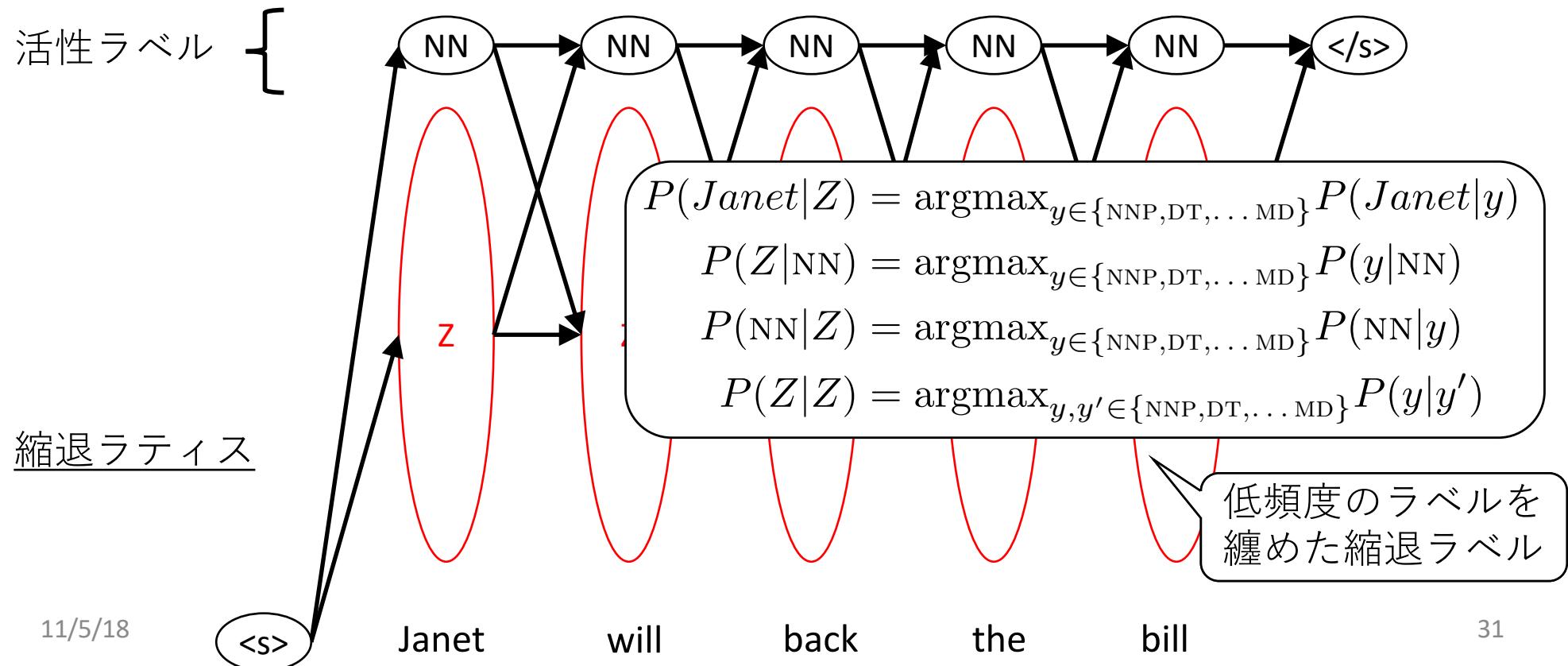
# 高速な復号化 (4/4): Staggered Decoding [Kaji+ 2010]

- ラベルのまとめ上げと枝刈りによる漸次的復号化
  - 複数のラベルをまとめた縮退ラベルに基づくラティス(縮退ラティス)上で最尤系列を探索し枝刈り



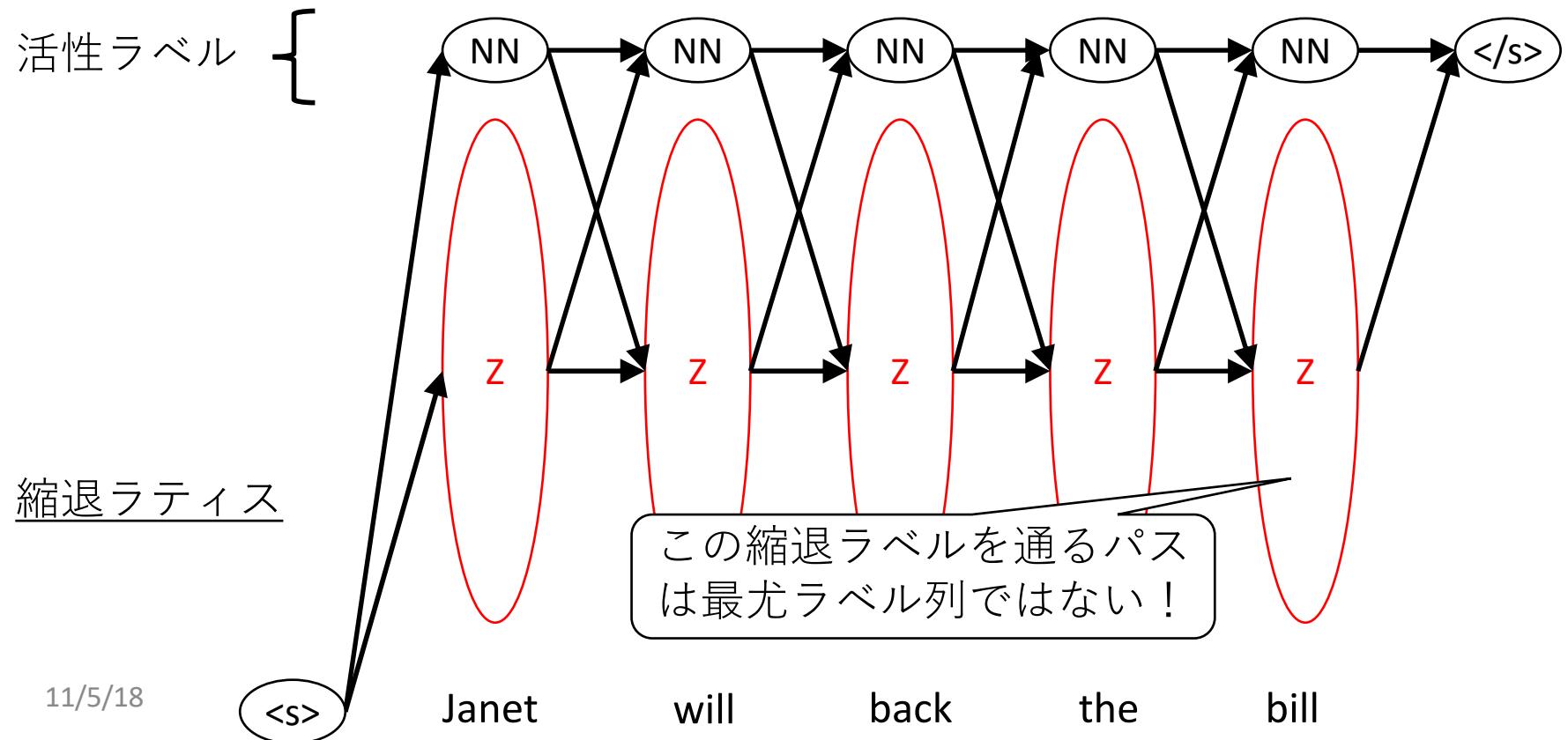
# 高速な復号化 (4/4): Staggered Decoding [Kaji+ 2010]

- ラベルのまとめ上げと枝刈りによる漸次的復号化
  - 複数のラベルをまとめた縮退ラベルに基づくラティス(縮退ラティス)上で最尤系列を探索し枝刈り



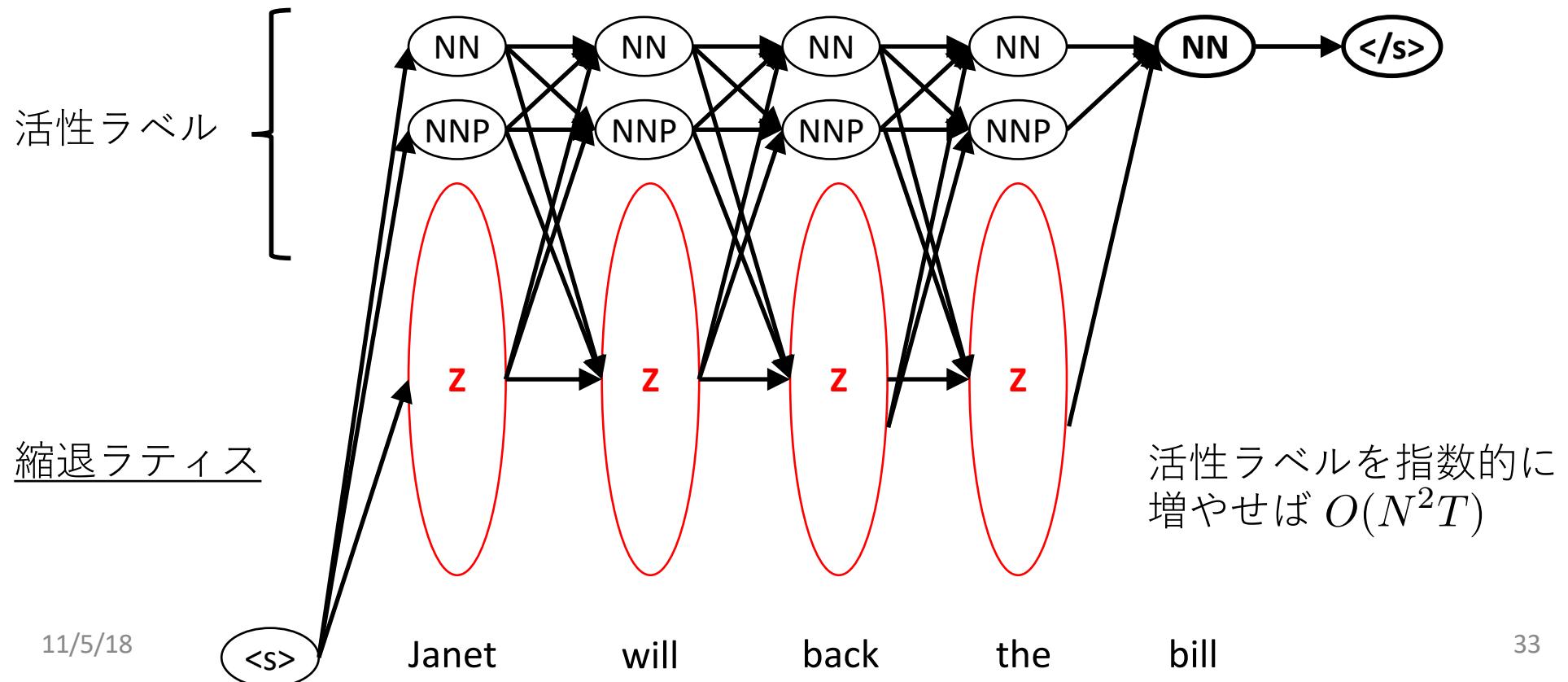
# 高速な復号化 (4/4): Staggered Decoding [Kaji+ 2010]

- ラベルのまとめ上げと枝刈りによる漸次的復号化
  - 複数のラベルをまとめた縮退ラベルに基づくラティス(縮退ラティス)上で最尤系列を探索し枝刈り



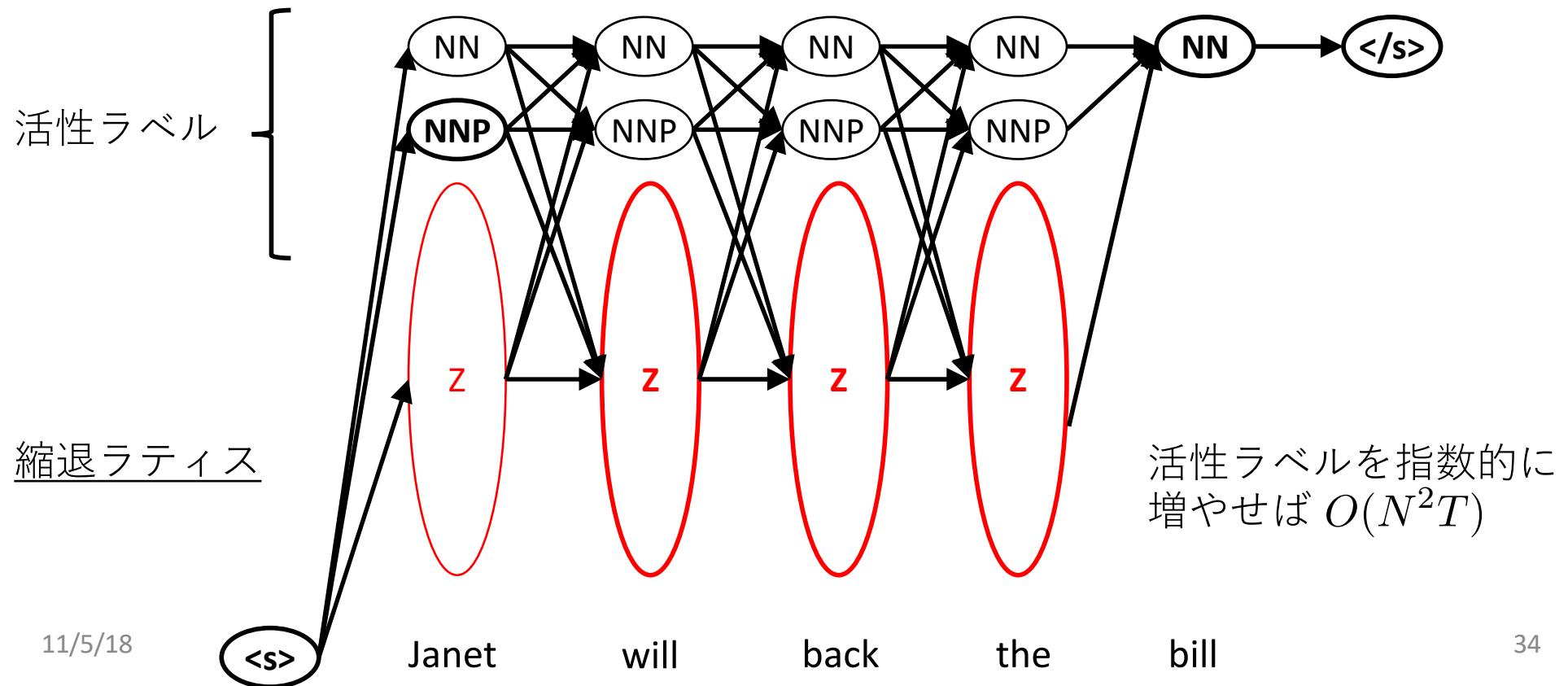
# 高速な復号化 (4/4): Staggered Decoding [Kaji+ 2010]

- ラベルのまとめ上げと枝刈りによる漸次的復号化
  - 複数のラベルをまとめた縮退ラベルに基づくラティス(縮退ラティス)上で最尤系列を探索し枝刈り
  - 最尤系列が縮退ラベルを通る限り分解して再探索



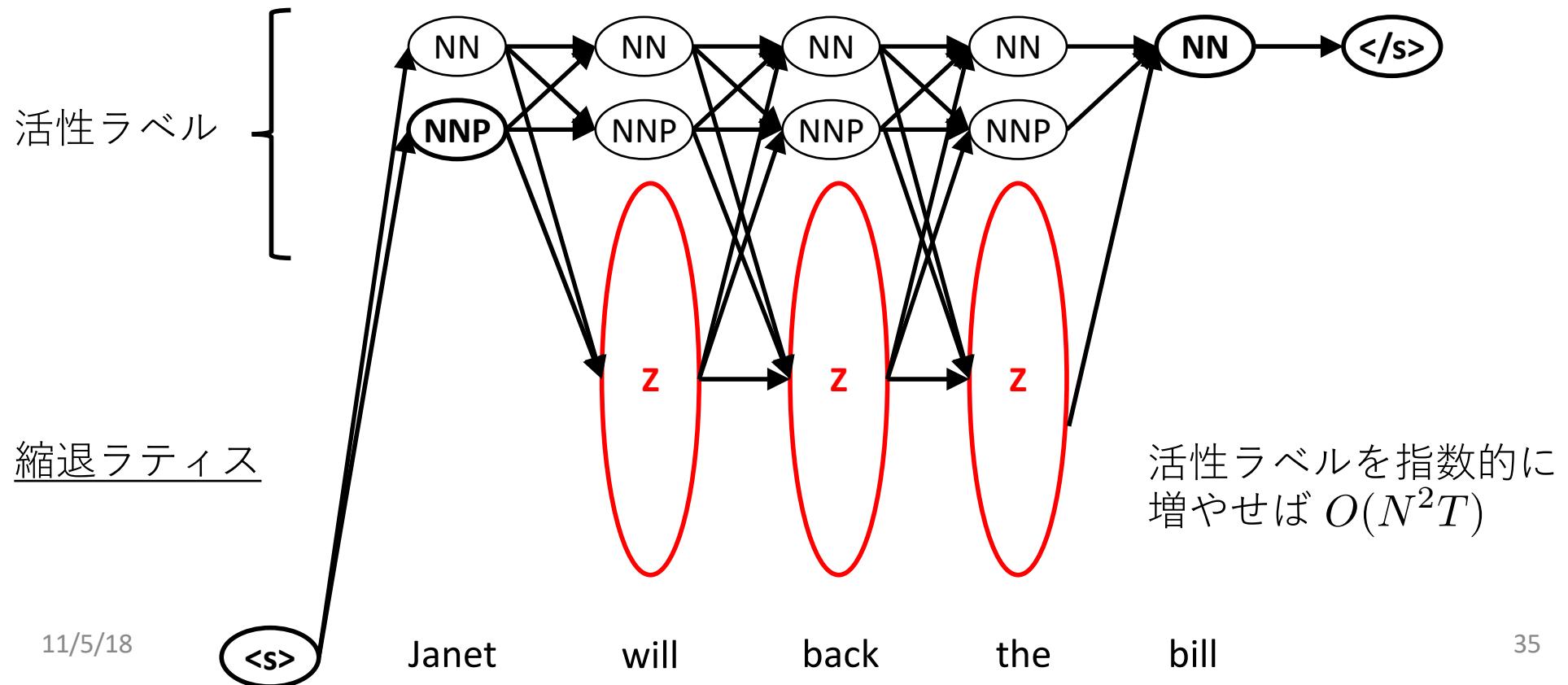
# 高速な復号化 (4/4): Staggered Decoding [Kaji+ 2010]

- ラベルのまとめ上げと枝刈りによる漸次的復号化
  - 複数のラベルをまとめた縮退ラベルに基づくラティス(縮退ラティス)上で最尤系列を探索し枝刈り
  - 最尤系列が縮退ラベルを通る限り分解して再探索



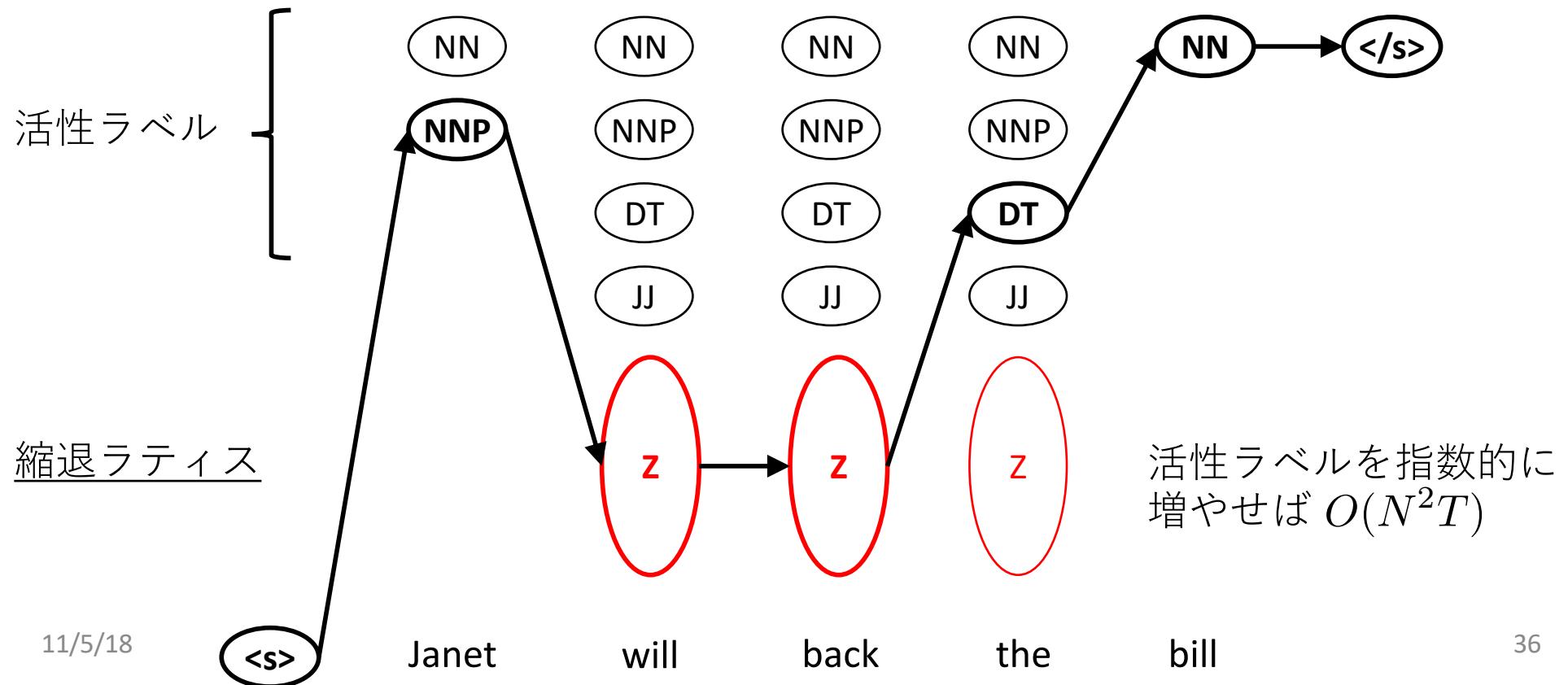
# 高速な復号化 (4/4): Staggered Decoding [Kaji+ 2010]

- ラベルのまとめ上げと枝刈りによる漸次的復号化
  - 複数のラベルをまとめた縮退ラベルに基づくラティス(縮退ラティス)上で最尤系列を探索し枝刈り
  - 最尤系列が縮退ラベルを通る限り分解して再探索



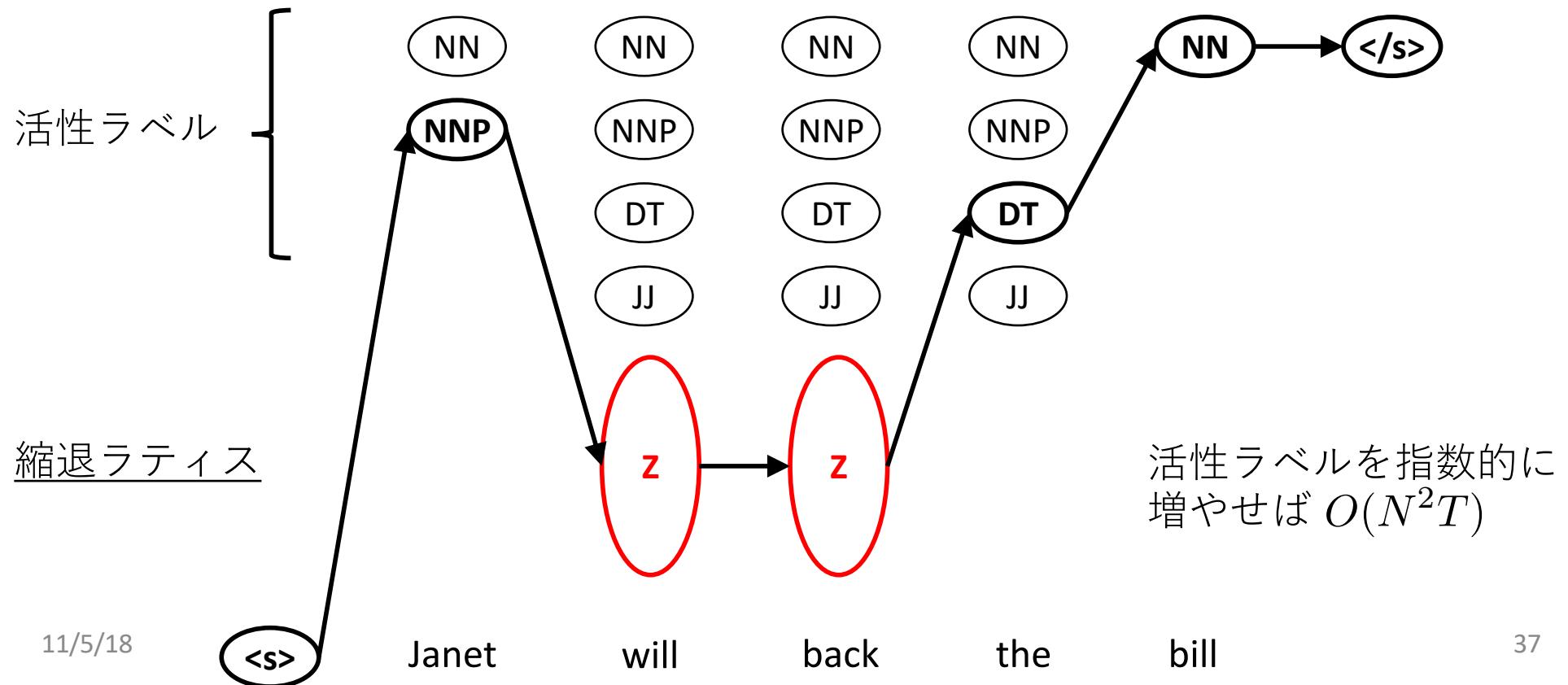
# 高速な復号化 (4/4): Staggered Decoding [Kaji+ 2010]

- ラベルのまとめ上げと枝刈りによる漸次的復号化
  - 複数のラベルをまとめた縮退ラベルに基づくラティス(縮退ラティス)上で最尤系列を探索し枝刈り
  - 最尤系列が縮退ラベルを通る限り分解して再探索



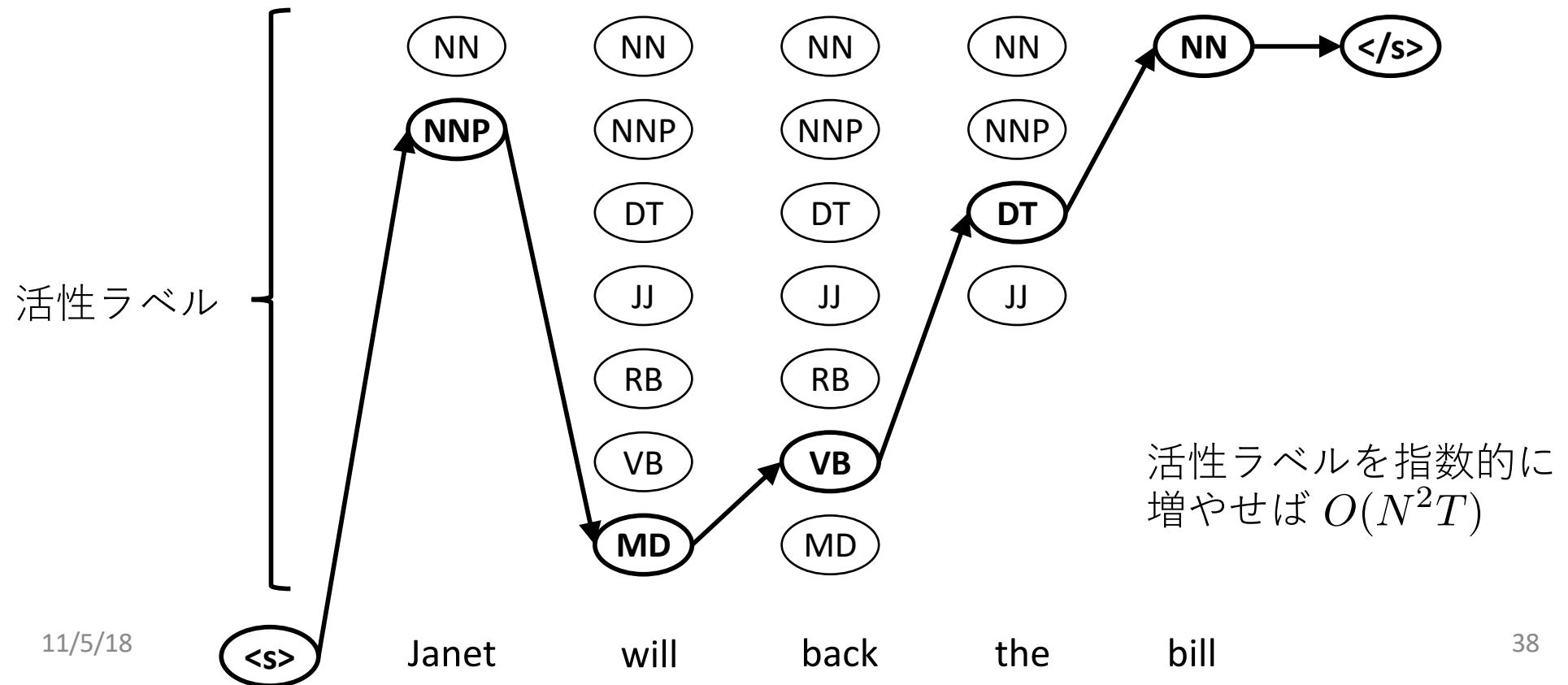
# 高速な復号化 (4/4): Staggered Decoding [Kaji+ 2010]

- ラベルのまとめ上げと枝刈りによる漸次的復号化
  - 複数のラベルをまとめた縮退ラベルに基づくラティス(縮退ラティス)上で最尤系列を探索し枝刈り
  - 最尤系列が縮退ラベルを通る限り分解して再探索



# 高速な復号化 (4/4): Staggered Decoding [Kaji+ 2010]

- ラベルのまとめ上げと枝刈りによる漸次の復号化
  - 複数のラベルをまとめた縮退ラベルに基づくラティス(縮退ラティス)上で最尤系列を探索し枝刈り
  - 最尤系列が縮退ラベルを通る限り分解して再探索



# 高速な復号化 (4/4): Staggered Decoding [Kaji+ 2010]

- ラベルのまとめ上げと枝刈りによる漸次の復号化
  - 複数のラベルをまとめた縮退ラベルに基づくラティス(縮退ラティス)上で最尤系列を探索し枝刈り
  - 最尤系列が縮退ラベルを通る限り分解して再探索

PTBにおける系列ラベリングの復号化速度 [文/秒]

| 手法                               | 品詞タグ付け<br>(N=45) | 品詞タグ付けx<br>NP同定 (N=319) | スーパータギング<br>(N=2602) |
|----------------------------------|------------------|-------------------------|----------------------|
| Viterbi                          | 4000             | 77                      | 1.1                  |
| Beam Search $\beta_{\text{opt}}$ | 18,000           | 2400                    | 160                  |
| Staggered [Kaji+ 2010]           | 14,000           | 1600                    | 300                  |

# 品詞タグ付けの評価

- テストコーパス中の個々の単語の品詞の分類精度で評価することが一般的
- Majority (Most Frequent) Class Baseline  
(各単語ごとに)学習データ中で最頻出のクラスに一律分類  
unigram HMM に相当  
例) 未知語に頻度1の単語の最頻品詞を割り当てた場合
  - English (PTB): 92.34% (学習/テスト: 37,972/5398 文)
  - Japanese (Kyoto): 93.58% (学習/テスト: 24,283/9284 文)

cf. HMM/構造化パーセプトロンによる精度  
(English: 96.03%/97.14%, Japanese: 96.14%/98.04%)

# 発展: HMM に基づく教師なし品詞タグ付け

- Hidden? Markov Model

教師有り品詞タグ付けの学習では品詞は隠れていない

- 教師なし品詞タグ付け

与えられた文(の集合)の確率が最大になるような  
出力確率・遷移確率を求める

- 推定には EM アルゴリズムや変分ベイズ法を使う

$$\theta = \operatorname{argmax}_{\theta} \prod_{(X_i, Y_i) \in \mathcal{D}} P(X, Y | \theta)$$

$$= \operatorname{argmax}_{\substack{P(x|y) \\ P(y|y')}} \prod_{(X_i, Y_i) \in \mathcal{D}} \prod_{t=1}^{T_i} P(x_t | y_t) P(y_t | y_{t-1})$$



# 教師あり学習に基づく構造分類

- タスク: 入力構造  $X$  をクラス構造  $Y$  に分類する
  - 出力構造: ラベル列, 木, グラフ  
例) 品詞分析: 単語列 → 品詞列

- 生成モデル: 同時確率  $P(X, Y)$  をモデル化

$$\begin{aligned}\hat{Y} &= \operatorname{argmax}_{Y \in \mathcal{Y}} P(Y|X) = \operatorname{argmax}_{Y \in \mathcal{Y}} \frac{P(X|Y)P(Y)}{P(Y)} \\ &= \operatorname{argmax}_{Y \in \mathcal{Y}} \underbrace{P(X|Y)}_{\text{尤度}} \underbrace{P(Y)}_{\text{事前確率}}\end{aligned}$$

- 識別モデル: 条件付き確率  $P(Y|X)$  を直接モデル化

$$\hat{Y} = \operatorname{argmax}_{Y \in \mathcal{Y}} P(Y|X)$$

# Maximum Entropy Markov Model [Ratnaparkhi 1996] (最大エントロピーマルコフモデル, MEMM)

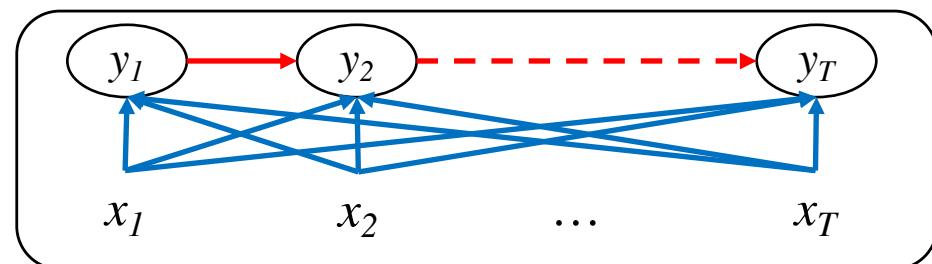
- 最大エントロピーモデル(ロジスティック回帰)に基づく系列ラベリングのための識別モデル
  - マルコフ性を仮定し問題を部分問題  $P(y_t|y_{t-1}, X)$  に分解

$$P(Y = y_1^T | X = x_1^T) = P(y_1, X)P(y_2|y_1, X)\dots P(y_T|y_1^{T-1}, X)$$

$$\approx \prod_{t=1}^T P(y_t|y_{t-1}, X)$$

$$P(y_t = c|y_{t-1}, X) = \frac{1}{1 + e^{-\mathbf{w}^T \phi_c(y_{t-1}, X)}}$$

HMM(生成モデル)と異なり,  
素性間の独立性を仮定しない  
ので重複する素性を利用可!



# MEMMの学習

- 注釈付きコーパスから各語の品詞を当てる多クラス分類問題を切り出しロジスティック回帰で学習

$$P(y_t = c | y_{t-1}, X) = \frac{1}{1 + e^{-\mathbf{w}^T \phi_c(y_{t-1}, X)}}$$

|       |       |      |           |    |
|-------|-------|------|-----------|----|
| party | might | then | have<br>? | to |
| NN    | IN    | RB   |           | TO |

- 素性関数  $\phi_c(y_{t-1}, X)$ 
  - 前後を含む単語の表層 (接尾辞, 接頭辞など)
  - 前の品詞

$$\phi_{c_1}(y_{t-1}, X) = \begin{cases} 1 & \text{if } \text{suffix}(x_t) = \text{-tion} \\ 0 & \text{otherwise} \end{cases} \quad \phi_{c_2}(y_{t-1}, X) = \begin{cases} 1 & \text{if } y_{t-1} = \text{MD} \\ 0 & \text{otherwise} \end{cases}$$

# MEMMの復号化 (1/3): greedy decoding

- Greedy decoding

- 文頭から決定的に復号化
- 計算量:  $O(NT)$

```
 $\hat{Y} = \{\}$ 
for  $t = 1$  to  $length(X)$ 
     $\hat{y}_t = \operatorname{argmax}_c P(y_t = c | \hat{Y}, x_1^T)$ 
     $\hat{Y} = \hat{Y} \cup \{\hat{y}_t\}$ 
return  $\hat{Y}$ 
```

|       |       |      |      |    |
|-------|-------|------|------|----|
| party | might | then | have | to |
| NN    | IN    | RB   | ?    | ?  |

- ☺ 高速 + 決定済のラベルが全てを素性に使える
- ☹ 分類誤りが起きるとエラーが伝播する

# MEMMの復号化 (2/3): Easiest-first decoding

[Tsuruoka+ 2005]

- Easiest-first decoding

- 高い確率で推定できる語から非決定的に復号化
- 計算量:  $O(NT)$

確率の再計算を行う必要  
がある単語が定数の場合

```
 $\hat{Y} = \{\}$ 
while  $length(\hat{Y}) < length(X)$ 
     $\hat{y}_t = \operatorname{argmax}_{t,c} P(y_t = c | \hat{Y}, x_1^T)$ 
     $\hat{Y} = \hat{Y} \cup \{\hat{y}_t\}$ 
return  $\hat{Y}$ 
```

party    might    then    have    to  
?            IN       RB      ?       TO

- ☺ 高速 + 決定済のラベルが全てを素性に使える
- ☹ 分類誤りが起きるとエラーが伝播する

# MEMMの復号化 (3/3): Viterbi decoding

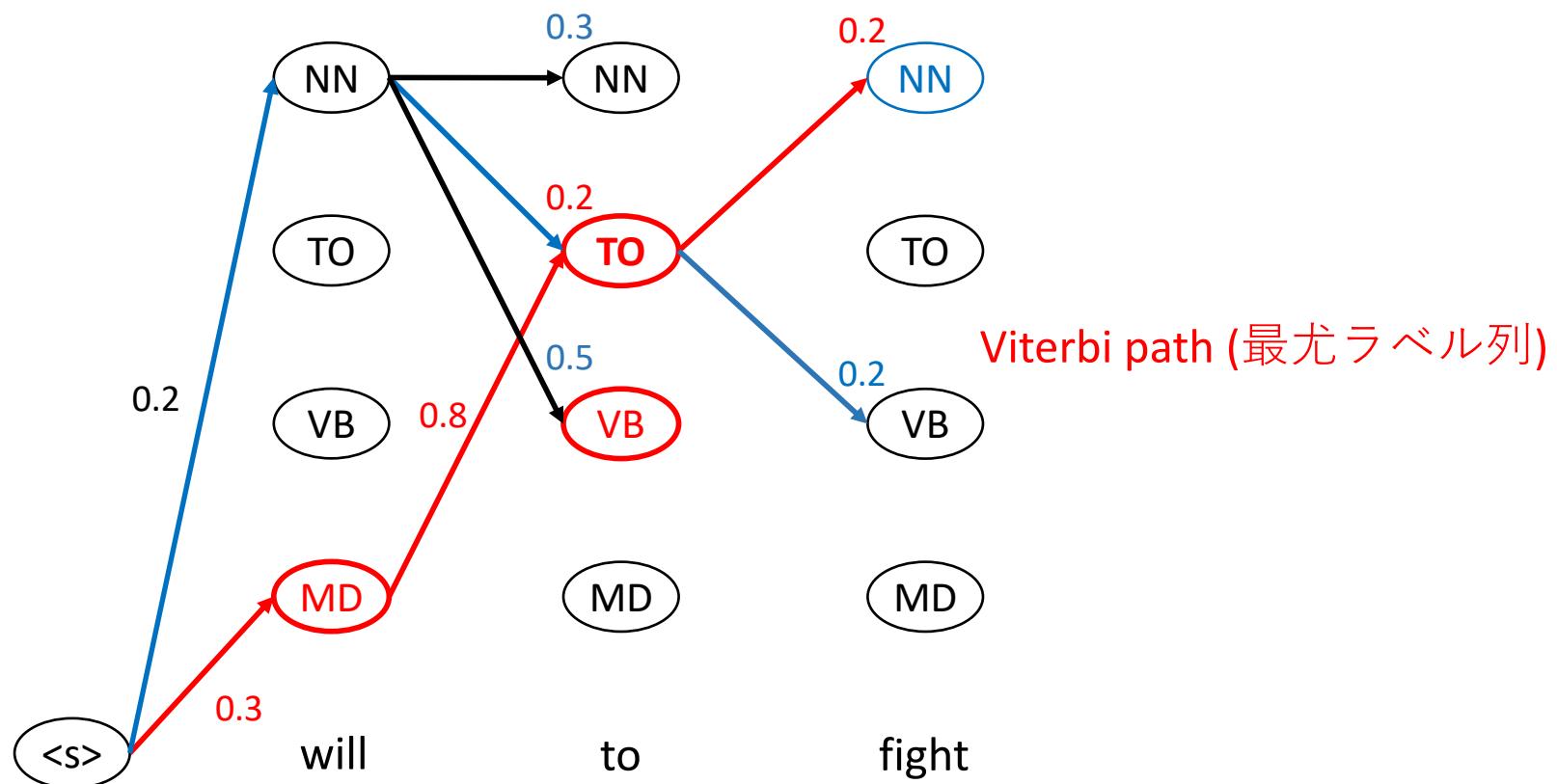
- Viterbi decoding
  - HMM の Viterbi decoding とほぼ同じ計算
$$P(y_t|y_{t-1})P(x_t|y_t) \leftrightarrow P(y_t|y_{t-1}, x_1^T)$$
  - 計算量:  $O(N^k T)$  ( $y_t$  の推定に  $y_{t-k+1}^{t-1}$  を考慮する場合)

party    might    then    have    to  
?           ?           ?           ?           ?

- ☺ 局所的なラベル間の相互依存性を考慮できる
- ☹ 遅い(ビームサーチ, Staggered decoding は有効)

# ラベルバイアス問題 [Lafferty+ 2001]

- HMM/MEMM は曖昧性の低いラベル列を選好
  - 部分問題として単語単位の確率を考えているため



# Conditional Random Field (条件付き確率場, CRF)

[Lafferty+ 2001]

- ラベルバイアス問題に対処した系列ラベリングのためのロジスティック回帰
  - 入力中の個々の単語のラベルを個別分類するのではなく,  
入力から出力ラベル列への写像を直接モデル化

$$P(Y|X) = \frac{1}{Z(x)} \exp(\mathbf{w}^T \phi(x_t, y_{t-1}, y_t))$$

$$\phi(x_t, y_{t-1}, y_t) = \sum_{t=1}^T (\mathbf{w}^T \phi(x_t, y_t) + \mathbf{w}'^T \phi'(y_{t-1} y_t))$$

$$Z(X) = \sum_Y P(Y|X)$$

- 学習: ロジスティック回帰と同じで, 学習データの対数尤度を最大化するよう勾配法で最適化
  - $Z(X)$ の計算は重い

# 構造化パーセプトロン [Collins+ 2002]

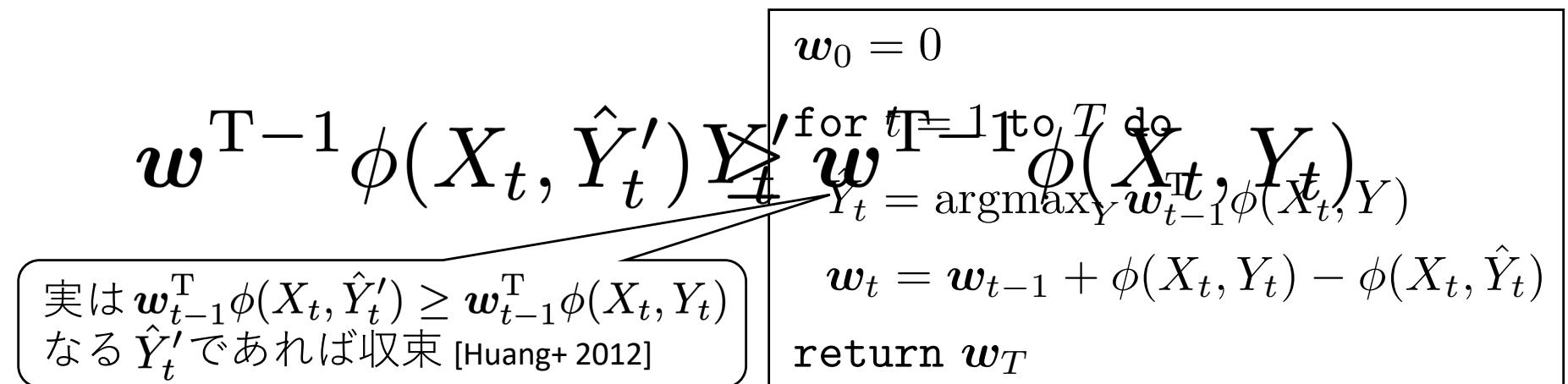
- 構造分類に対応した多クラスパーセプトロン
  - 例ごとに正解ラベル列のスコアが大きくなるよう更新
  - 学習データが線形分離可能なら有限回数で収束(宿題)

```
w0 = 0
for t = 1 to T do
    Y_t-hat = argmax_Y w_{t-1}^T phi(X_t, Y)
    w_t = w_{t-1} + phi(X_t, Y_t) - phi(X_t, Y_t-hat)
return
```

$$\begin{aligned}w_t^T \phi(X_t, Y_t) &= (w_{t-1} + \phi(X_t, Y_t) - \phi(X_t, \hat{Y}_t))^T \phi(X_t, Y_t) \\&= w_{t-1}^T \phi(X_t, Y_t) + (\|\phi(X_t, Y_t)\|^2 - \phi(X_t, \hat{Y}_t)^T \phi(X_t, Y_t)) \\&\geq w_{t-1}^T \phi(X_t, Y_t) \\w_t^T \phi(X_t, \hat{Y}_t) &= (w_{t-1} + \phi(X_t, \hat{Y}_t) - \phi(X_t, \hat{Y}_t))^T \phi(X_t, \hat{Y}_t) \\&= w_{t-1}^T \phi(X_t, \hat{Y}_t) - (\|\phi(X_t, \hat{Y}_t)\|^2 - \phi(X_t, \hat{Y}_t)^T \phi(X_t, \hat{Y}_t)) \\&\leq w_{t-1}^T \phi(X_t, \hat{Y}_t)\end{aligned}$$

# 構造化パーセプトロン [Collins+ 2002]

- 構造分類に対応した多クラスパーセプトロン
  - 例ごとに正解ラベル列のスコアが大きくなるよう更新
  - 学習データが線形分離可能なら有限回数で収束(宿題)



英語の品詞タグ付け

| 手法               | $\beta$  | iter | train time | test acc. |
|------------------|----------|------|------------|-----------|
| Viterbi (exact)  | $\infty$ | 6    | 162m       | 97.28%    |
| Early (beam)     | 6        | 6    | 37m        | 97.35%    |
| Max-viol. (beam) | 2        | 3    | 26m        | 97.33%    |

# 双对分解 [Koo+ 2010]

# まとめ

- 品詞タグ付け: 構造分類(系列ラベリング)問題
  - Universal POSs / Penn TreeBank: open & closed class POSs
- 系列ラベリングの生成モデル: HMM
  - 学習: 最尤推定, 未知語問題のためのスムージング
  - 復号化: Viterbi, ビームサーチ, Staggered decoding
- 系列ラベリングの識別モデル: MEMM / CRF / 構造化パーセプトロン
  - 学習: ロジスティック回帰, 確率的勾配降下法
  - 復号化: Greedy, Easiest-first, Viterbi
- 双対分解