

1 マルコフモデル

$p(c_i)$ c_i が現れる確率 $p(c_i) = \sum_j p(c_i, c_j)$

$p(c_i, c_j)$ c_i に続いて c_j が現れる確率

$p(c_j|c_i)$ c_i が現れるとわかった後に c_j が現れる確率 $p(c_j|c_i) = p(c_i, c_j)/p(c_i)$

2 文脈自由型文法

チョムスキーの標準形

$$\left. \begin{array}{l} A \rightarrow BC \\ A \rightarrow a \end{array} \right\} \quad A, B, C \in V_N, a \in V_T$$

グライバッハの標準形

$$A \rightarrow aB_1B_2 \cdots B_n \quad a \in V_T, A, B_1, B_2, \dots, B_n \in V_N, n \leq 0$$

3 形態素解析

形態素解析アルゴリズム

- [1] 入力文中の位置 (文字と文字の間) を示すポインターを用意する。初期状態としてポインターを位置 0(先頭の文字の左側) におく。また文頭という仮想的なノードをおく。
- [2] ポインター位置から始まる語を辞書で検索する。
- [3] ポインター位置で終わっている語 (位置 0 の場合は文頭ノード) とポインター位置から始まる語の各ペアについて連接可能なものがあればその間にリンクをはる。ポインター位置から始まる語の中でいずれの語とも連接可能でない語は排除する。
- [4] ポインター位置から右側を順番に調べ、そこで終る語が存在する位置までポインターを移動する。ポインターが文末にくるまでステップ [2],[3],[4] を繰り返す。文末の場合には、文末という仮想的なノードをおき、文末位置で終わっている語と文末との連接可能性を調べ、可能なものだけを文末ノードにリンクして処理を終了する。
- [5] 最終的に、文頭ノードから文末ノードまでのパス (ノードとリンクの並び) が入力文に対する形態素解析結果となっている。

ヴィテルビ・アルゴリズム

- [1] 前アルゴリズムとおなじ
- [2] 前アルゴリズムとおなじ
- [3] あるポインター位置で終る語を W_e1, W_e2, \dots , そのポインター位置から始まる語を W_s1, W_s2, \dots , とする。各 W_si について左に連接可能な W_ej を求め、 W_si との間にリンクをはる。このとき
 - W_ej までの部分最小コスト

- W_{ej}, W_{si} 間の接続コスト
- W_{si} の単語コスト

の和が最小であるような W_{ej} を求め、その最小値を W_{si} までの部分最小コストとし、 W_{ej} と W_{si} とのリンクに特別のマークを付ける。

[4] 前アルゴリズムとおなじ

[5] 前アルゴリズムとおなじ

[6] 最終的にマークの付いたリンクによる文頭から文末までのパスがコスト最小の解となる。

英語文の形態素解析 入力単語列を W_1, W_2, \dots, W_n としたとき、求めたい品詞列 C_1, C_2, \dots, C_n は次の条件付き確率を最大にする品詞列

$$P(C_1, C_2, \dots, C_n | w_1, w_2, \dots, w_n)$$

ベイズの定理によって次のように変形できる。

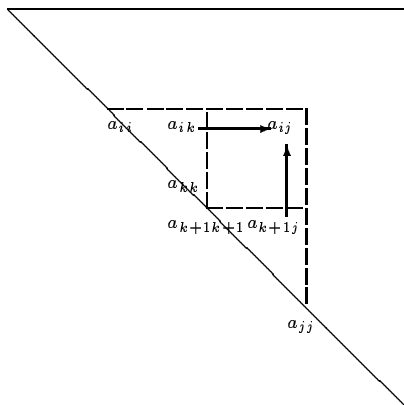
$$\frac{P(C_1, C_2, \dots, C_n) \times P(w_1, w_2, \dots, w_n | C_1, C_2, \dots, C_n)}{P(w_1, w_2, \dots, w_n)}$$

分子を最大とする品詞列を求める。分子を近似するとつぎのようになる。

$$\prod_{i=1}^n P(C_i | C_{i-1}) \times P(w_i | C_i)$$

4 構文解析

CKY法 三角行列 $a_{ij} (1 \leq i \leq j \leq n, n$ は入力文の単語数) を用いて解析をおこなう。まず、辞書規則を参照することによって、入力文中の i 番目の単語を生成する前終端記号を対角線要素 a_{ij} として与える。次に、要素 a_{ij} に i 番目から j 番目までの隣接する品詞列を生成することができるような非終端記号を与える。これは $a_{ik} (i \leq k < j)$ 中の記号と a_{k+1j} 中の記号をまとめる句構造規則を探し、その左辺の非終端記号を a_{ij} に与えることによって実現する。



チャート法

上昇型

- [1] e_i を文法カテゴリ α の不活性弧とし、節点 v と v' を結んでいるとする。このとき、 $\gamma_1 = \alpha$ となるすべての文法規則 $\beta \rightarrow \gamma_1 \cdots \gamma_n$ に対して、 $[a'[\gamma_1]_{\gamma_2} \cdots [\gamma_n]_{\beta}]$ という項をもつ活性弧を v から v' の間に追加せよ。ただし、すでに同じ弧が存在するときは追加する必要がない。ここに、 a' は e_i の構造を表す項である。
- [2] e_a と e_i を隣り合う活性弧および不活性弧とせよ。つまり、 e_a が節点 v および v' を張り、 e_i が v' および w の間を張るとする。このとき、 e_a 内で最初に現れる空所が $[\gamma]_{\alpha}$ であり、 e_i の文法カテゴリが α であるなら、 e_a の項の最初の空所を e_i の項の値で置き換えた項をもち、 v と w の間を張る新しい弧をつくれ。

#	locus	length	term	(proc,items)
1	0	1	[failing] _A	
2	0	1	[failing] _{GRD}	
3	1	1	[student] _N	
4	2	1	[looked] _V	
5	3	1	[hard] _A	
6	3	1	[hard] _{AV}	
7	0	1	[[failing] _A [?] _N] _{NP}	(1,2)
8	0	1	[[failing] _{GRD} [?] _N] _{NP}	(1,3)
9	0	2	[[failing] _A [student] _N] _{NP}	(2,#3)
10	0	2	[[failing] _{GRD} [student] _N] _{NP}	(2,#3)
11	0	2	[[[failing] _A [student] _N] _{NP} [?] _{VP}] _S	(1,1)
12	0	2	[[[failing] _{GRD} [student] _N] _{NP} [?] _{VP}] _S	(1,1
)
13	2	1	[[looked] _V [?] _A] _{VP}	(1,4)
14	2	1	[[looked] _V [?] _{AV}] _{VP}	(1,5)
15	2	2	[[looked] _V [hard] _A] _{VP}	(2,#5)
16	2	2	[[looked] _V [hard] _{AV}] _{VP}	(2,#6)
17	0	4	[[[failing] _A [student] _N] _{NP} [[looked] _V [hard] _A] _{VP}] _S	(2,#11)
18	0	4	[[[failing] _{GRD} [student] _N] _{NP} [[looked] _V [hard] _A] _{VP}] _S	(2,#12)
19	0	4	[[[failing] _A [student] _N] _{NP} [[looked] _V [hard] _{AV}] _{VP}] _S	(2,#11)
20	0	4	[[[failing] _{GRD} [student] _N] _{NP} [[looked] _V [hard] _{AV}] _{VP}] _S	(2,#12)

下降型

- [1] e_a を活性弧とし、節点 v と w を結んでいるとする。 e_a のなかで最初に現れる空所が $[\gamma]_{\alpha}$ であるとき、 $\alpha \rightarrow \gamma_1, \cdots, \gamma_n$ というすべての文法規則に対して、 $[[\gamma]_{\gamma_1} [\gamma_2], \cdots, [\gamma_n]_{\alpha}]$ という項をもつ活性弧を節点 w から w の間に追加せよ。

#	locus	length	term	(proc,items)
1	0	0	[?]S	
2	0	1	[failing] _A	
3	0	1	[failing] _{GRD}	
4	1	1	[student] _N	
5	2	1	[looked] _V	
6	3	1	[hard] _A	
7	3	1	[hard] _{AV}	
8	0	0	[[?]NP [?]VP]S	(1,1)
9	0	0	[[?]A [?]N]NP	(1,2)
10	0	0	[[?]GRD [?]N]NP	(1,3)
11	0	1	[[failing] _A [?]N]NP	(2,#2)
12	0	1	[[failing] _{GRD} [?]N]NP	(2,#3
)
13	0	2	[[failing] _A [student] _N]NP	(2,#4)

5 単一化文法

素性 非/前終端記号に付与する属性

- 単数・複数
- 原形・現在分詞・過去形

素性構造 構文カテゴリー自体も一つの素性の値と考え、素性構造 (=素性-値の対) の関係を規定するものとして文法をとらえる。

単一化文法の特徴

- 文脈自由文法に対する効率のよい構文解析法をそのまま利用できる。
- 構文解析と意味解析を統合して行うことも可能。
- 素性構造に選言が多く含まれる場合、計算時間が非常にかかる。

6 意味解析

表層格 主格・目的格

深層格 文中の動詞に対して他の単語が果たす役割

格フレーム・格構造表現・必須格・任意格