

## 2. Übung zur Vorlesung “NLP – Analyse des Wissensrohstoffes Text” im Sommersemester 2008 – mit Musterlösungen –

Dr. Andreas Hotho, Dipl.-Inform. Dominik Benz, Wi.-Inf. Beate Krause

28. Mai 2008

### 1 Part-of-Speech Tagging

#### 1.1 Grundlagen

Part-of-Speech Tagging stellt eine wichtige Zwischenstufe auf dem Weg dar, natürliche Sprache automatisch zu verarbeiten.

1. Erklären Sie mit Ihren eigenen Worten, worum es sich bei einem “Part-of-Speech” handelt und worum es beim Part-of-Speech Tagging geht.
2. Welchen Mehrwert kann ein korrekt getaggtter Text im Vergleich zu einem nicht-getaggtten Text haben? Zählen Sie einige Anwendungen auf, die diesen ausnutzen.
3. Welche Informationen können für das POS-Tagging verwendet werden? Welche Informationsquelle ist besonders nützlich und weshalb?

LÖSUNGSVORSCHLAG:

1. *Parts of Speech* entsprechen Wortarten. Beim POS Tagging geht es darum in einem Text jedem Wort seine richtige Wortart (im Kontext eines Satzes) zuzuweisen.
2. Auf der Tagging-Information können weitere NLP-Methoden aufbauen (Parsing, ...). Des Weiteren kann die Zusatzinformation hilfreich sein für Methoden der Information Extraction oder des Question Answering.
3. Grob unterscheidet man zwischen *syntagmatischer* Information (z.B. durch Untersuchung typischer Folgen von Wortarten wie AT JJ NN) und *lexikalischer* Information. Letztere beinhaltet das Auszählen von Häufigkeiten bestimmter Wortarten für bestimmte Worte. Da diese Verteilung typischerweise sehr ungleichmässig ist, ist diese Methode recht erfolgreich (90%), während rein syntagmatisch basierte Ansätze auf ca. 70% kommen.

Tabelle 1: Häufigkeiten der Abfolge einiger POS Tags aus dem Brown Corpus (idealisiert)

Erstes Tag	zweites Tag					
	AT	BEZ	IN	NN	VB	PERIOD
AT	0	0	0	48636	0	19
BEZ	1973	0	426	187	0	38
IN	43322	0	1325	17314	0	185
NN	1067	3720	42470	11773	614	21392
VB	6072	42	4758	1476	129	1522
PERIOD	8016	75	4656	1329	954	0

## 1.2 Visible Markov Model Taggers

Ein Ansatz zum POS-Tagging sind Visible Markov Models (VMM). Erklären Sie kurz, was in diesen Modellen den Parts-of-Speech entspricht.

1. Welche vereinfachenden Annahmen über die Struktur von Sprache liegen der Anwendung von Markov-Modellen zum POS-Taggen zugrunde?
2. Welche Phänomene natürlicher Sprache widersprechen diesen Annahmen?

Folgende Daten zur Abfolge bestimmter POS Tags und deren Häufigkeit für bestimmte Worte stammen aus dem Brown Corpus (siehe Manning/Schütze, S. 348f):

1. Beschreiben Sie in Ihren eigenen Worten, was passiert, wenn ein Visible Markov Model zum Taggen auf einen bestimmten Text trainiert wird. Was genau wird dabei “gelernt”?
2. Berechnen Sie die Wahrscheinlichkeit der beiden Taggings AT NN BEZ IN AT NN sowie AT NN BEZ IN AT VB für den Satz *The bear is on the move*. Machen Sie dabei klar, an welcher Stelle das Markov Model trainiert wird.
3. Nachdem ein VMM trainiert wurde, kann es zum Taggen von unbekanntem Corpora eingesetzt werden. Diskutieren Sie, inwiefern es sich dann noch um ein “visible” Markov Model handelt. Welche Verfahren kennen Sie, um mit einem fertig trainierten Modell zu taggen? Welches Verfahren ist effizient?
4. Verwenden Sie die Daten aus den Tabellen 1.1 und 1.2, um mittels des Viterbi-Algorithmus den Satz *The bear is on the move* zu taggen.
5. Ein grosses Problem beim POS-Tagging allgemein sind unbekannte Wörter. Skizzieren Sie kurz Lösungsansätze.

LÖSUNGSVORSCHLAG:

Tabelle 2: Häufigkeiten bestimmter Tags für einige Wörter aus dem Brown Corpus (idealisiert)

	AT	BEZ	IN	NN	VB	PERIOD
<i>bear</i>	0	0	0	10	43	0
<i>is</i>	0	10065	0	0	0	0
<i>move</i>	0	0	0	36	133	0
<i>on</i>	0	0	5484	0	0	0
<i>president</i>	0	0	0	382	0	0
<i>progress</i>	0	0	0	108	4	0
<i>the</i>	69016	0	0	0	0	0
.	0	0	0	0	0	48809
total (all words)	120991	10065	130534	134171	20976	49267

1. Vereinfachende Annahmen sind zunächst alle Eigenschaften eine Markov-Kette (limited horizon, time invariant). Zusätzlich:
  - Unabhängigkeit der Wörter untereinander
  - Wort hängt nur von seiner eigenen Wortart ab, nicht von der von Vorgänger-Worten
2. Gegenbeispiele zur Unabhängigkeit der Wörter untereinander:
  - Kollokationen (*grünes Licht, New York*)
  - Eigennamen (*Pfarrer Braun, ...*)
3. Gegenbeispiele zu "Wort hängt nur von eigener Wortart ab":
  - *Fruit flies like a banana.*
  - wieder Kollokationen, Idiome, ...

Zweiter Fragenblock:

1. Aus dem (getaggten) Trainingsdatensatz ergeben sich die Menge der Zustände (Tags) und das Ausgabealphabet (Wörter). "Gelernt" werden nun die Übergangswahrscheinlichkeiten zwischen den Zuständen, sowie die Ausgabewahrscheinlichkeiten der Symbole, also Wörter. Beides geschieht mittels der Maximum Likelihood Abschätzung.
2. Berechnung s.u.
3. De facto handelt es sich bei der Anwendung eines VMM zum Taggen eines unbe-

kannten Textes um ein Hidden Markov Model, da im ZIELTEXT ja nur die Wörter, nicht jedoch die Zustände vorhanden sind. Die Grundfrage ist für jeden Satz, welche Abfolge von Zuständen (also Tags) die wahrscheinlichste ist. Der naive Ansatz wäre das Durchprobieren aller möglichen Zustandsfolgen, was jedoch exponentiell in der Anzahl der Zustände ist. Der Viterbi-Algorithmus bietet eine effizientere Lösung (s.u.).

4. Berechnung s.u.
5. Markov-Modelle per se können mit unbekanntem Wörtern nicht umgehen, da diese noch nicht beobachtet wurden und somit keine Ausgabewahrscheinlichkeiten im Modell vorhanden sind. Um dem entgegenzuwirken, werden folgende Heuristiken eingesetzt:
  - Annahme: unbekannte Wörter können von jeder Wortart sein, mit der Wahrscheinlichkeit dieser Wortart über das gesamte Lexikon
  - Ausnutzung weiterer lexikalischer Hinweise, z.B. Gross-/Kleinschreibung, Endung, ...
  - siehe Tabelle 10.5, S. 352, Manning/Schütze

**Berechnung der Wahrscheinlichkeiten für Taggings** In unserem Fall besteht das VMM aus 6 Zuständen (AT NN BEZ IN AT VB PERIOD), das Ausgabealphabet aus 8 Wörtern (*bear, is, move, on, president, progress, the, .*). Der Einfachheit halber berechnen wir nicht alle Wahrscheinlichkeiten des Modells, sondern nur die notwendigen:

*Zustandsübergänge:*

$$P(AT|PERIOD) = \frac{8016}{49267} = 0.16$$

$$P(NN|AT) = \frac{48636}{120991} = 0.40$$

$$P(BEZ|NN) = \frac{3720}{134171} = 0.03$$

$$P(IN|BEZ) = \frac{426}{10065} = 0.04$$

$$P(AT|IN) = \frac{43322}{130534} = 0.33$$

$$P(VB|AT) = \frac{0}{120991} = 0$$

*Ausgabe:*

$$P(the|AT) = \frac{69016}{120991} = 0.57$$

$$P(bear|NN) = \frac{10}{134171} = 0.00007$$

$$P(is|BEZ) = \frac{10065}{10065} = 1$$

$$P(on|IN) = \frac{5484}{130534} = 0.04$$

$$P(move|NN) = \frac{36}{134171} = 0.0002$$

$$P(move|VB) = \frac{133}{20976} = 0.006$$

Mit der Berechnung dieser Wahrscheinlichkeiten ist das Training abgeschlossen.

Mit  $w_{1,6} = \textit{The bear is on the move.}$  und  $t_{1,6} = \textit{AT NN BEZ IN AT NN}$  gilt nun

$$\begin{aligned}
 P(t_{1,6}|w_{1,6}) &= \prod_{i=1}^6 P(w_i|t_i)P(t_i|t_{i-1}) \\
 &= P(\textit{AT|PERIOD})P(\textit{the|AT}) \times P(\textit{NN|AT})P(\textit{bear|NN}) \times \\
 &\quad P(\textit{NN|BEZ})P(\textit{is|BEZ}) \times P(\textit{IN|BEZ})P(\textit{on|IN}) \times \\
 &\quad P(\textit{AT|IN})P(\textit{the|AT}) \times P(\textit{NN|AT})P(\textit{move|NN}) \\
 &= 0.57 * 0.16 \times 0.0007 * 0.40 \times 1 * 0.03 \times 0.04 * 0.04 \times 0.57 * 0.33 \times 0.0002 * 0.40 \\
 &= 0.000000000000018444754944
 \end{aligned}$$

Mit  $t_{1,6} = \textit{AT NN BEZ IN AT VB}$  gilt

$$\begin{aligned}
 P(t_{1,6}|w_{1,6}) &= \prod_{i=1}^6 P(w_i|t_i)P(t_i|t_{i-1}) \\
 &= P(\textit{AT|PERIOD})P(\textit{the|AT}) \times P(\textit{NN|AT})P(\textit{bear|NN}) \times \\
 &\quad P(\textit{NN|BEZ})P(\textit{is|BEZ}) \times P(\textit{IN|BEZ})P(\textit{on|IN}) \times \\
 &\quad P(\textit{AT|IN})P(\textit{the|AT}) \times P(\textit{VB|AT})P(\textit{move|VB}) \\
 &= 0.57 * 0.16 \times 0.0007 * 0.40 \times 1 * 0.03 \times 0.04 * 0.04 \times 0.57 * 0.33 \times 0 * 0.40 \\
 &= 0
 \end{aligned}$$

### Tagging mittels des Viterbi-Algorithmus

Der erste Schritt besteht darin, alle Übergangs-/Emissionswahrscheinlichkeiten mittels Maximum-Likelihood-Abschätzung auszurechnen. Diese finden sich in den Tabellen 1.2 und 1.2.

#### Initialization

$$\phi_0(\textit{PERIOD}) = 1$$

#### Induktion:

#### Viterbi-Wahrscheinlichkeiten für das Wort 'the'

Abbildung 1 zeigt den ersten Induktionsschritt, in dem die Viterbi-Wahrscheinlichkeiten aller Tags für das Wort 'the' berechnet werden. Die übrigen Berechnungsschritte befinden sich der Vollständigkeit halber im Anhang.

Nach allen Durchläufen lassen sich die Ergebnisse am übersichtlichsten in einer Tabelle (siehe 1.2) zusammenfassen

#### Terminierung / Auslesen des Pfades

Tabelle 3: Übergangswahrscheinlichkeiten einiger Tags im Brown-Corpus (idealisiert)

Erstes Tag	zweites Tag					
	AT	BEZ	IN	NN	VB	PERIOD
AT	0	0	0	0.40	0	0.0002
BEZ	0.20	0	0.04	0.02	0	0.004
IN	0.33	0	0.01	0.13	0	0.001
NN	0.008	0.03	0.32	0.09	0.005	0.16
VB	0.29	0.002	0.23	0.07	0.006	0.07
PERIOD	0.16	0.002	0.09	0.03	0.02	0

Tabelle 4: Emissionswahrscheinlichkeiten einiger Wörter aus dem Brown-Corpus (idealisiert)

	AT	BEZ	IN	NN	VB	PERIOD
<i>bear</i>	0	0	0	0.00007	0.002	0
<i>is</i>	0	1	0	0	0	0
<i>move</i>	0	0	0	0.002	0.006	0
<i>on</i>	0	0	0.04	0	0	0
<i>president</i>	0	0	0	0.003	0	0
<i>progress</i>	0	0	0	0.0008	0.0002	0
<i>the</i>	0.57	0	0	0	0	0
.	0	0	0	0	0	0.99

Tabelle 5: Übersicht über den Viterbi-Algorithmus

t	1		2		3		4		5		6	
	<i>the</i>	<i>bear</i>	<i>is</i>	<i>on</i>	<i>the</i>	<i>move</i>						
	$\delta_1$	$\delta_2$	$\psi_2$	$\delta_3$	$\psi_3$	$\delta_4$	$\psi_4$	$\delta_5$	$\psi_5$	$\delta_6$	$\psi_6$	
AT	0.0912	0	/	0	/	0	/	$2.3 * 10^{-11}$	<b>IN</b>	0	/	
BEZ	0	0	/	$7.7 * 10^{-8}$	<b>NN</b>	0	/	0	/	0	/	
IN	0	0	/	0	/	$1.2 * 10^{-10}$	<b>BEZ</b>	0	/	0	/	
NN	0	$2.5 * 10^{-6}$	<b>AT</b>	0	/	0	/	0	/	$1.84 * 10^{-14}$	<b>AT</b>	
VB	0	0	/	0	/	0	/	0	/	0	/	
PERIOD	0	0	/	0	/	0	/	0	/	0	/	

$$\begin{aligned}
\delta_1(AT) &= \max( \\
\delta_0(AT) * P(AT|AT) * P(the|AT) &= 0 * 0 * 0.57 = 0.0 \\
\delta_0(BEZ) * P(AT|BEZ) * P(the|AT) &= 0 * 0.2 * 0.57 = 0.0 \\
\delta_0(IN) * P(AT|IN) * P(the|AT) &= 0 * 0.33 * 0.57 = 0.0 \\
\delta_0(NN) * P(AT|NN) * P(the|AT) &= 0 * 0.008 * 0.57 = 0.0 \\
\delta_0(VB) * P(AT|VB) * P(the|AT) &= 0 * 0.29 * 0.57 = 0.0 \\
\delta_0(PERIOD) * P(AT|PERIOD) * P(the|AT) &= 1 * 0.16 * 0.57 = 0.0912 \\
) & \\
\psi_1(AT) &= PERIOD
\end{aligned}$$

$$\begin{aligned}
\delta_1(BEZ) &= \max( \\
\delta_0(AT) * P(BEZ|AT) * P(the|BEZ) &= 0 * 0 * 0 = 0 \\
\delta_0(BEZ) * P(BEZ|BEZ) * P(the|BEZ) &= 0 * 0 * 0 = 0 \\
\delta_0(IN) * P(BEZ|IN) * P(the|BEZ) &= 0 * 0 * 0 = 0 \\
\delta_0(NN) * P(BEZ|NN) * P(the|BEZ) &= 0 * 0.03 * 0 = 0.0 \\
\delta_0(VB) * P(BEZ|VB) * P(the|BEZ) &= 0 * 0.002 * 0 = 0.0 \\
\delta_0(PERIOD) * P(BEZ|PERIOD) * P(the|BEZ) &= 1 * 0.002 * 0 = 0.0 \\
) & \\
\psi_1(BEZ) &= PERIOD
\end{aligned}$$

$$\begin{aligned}
\delta_1(IN) &= \max( \\
\delta_0(AT) * P(IN|AT) * P(the|IN) &= 0 * 0 * 0 = 0 \\
\delta_0(BEZ) * P(IN|BEZ) * P(the|IN) &= 0 * 0.04 * 0 = 0.0 \\
\delta_0(IN) * P(IN|IN) * P(the|IN) &= 0 * 0.01 * 0 = 0.0 \\
\delta_0(NN) * P(IN|NN) * P(the|IN) &= 0 * 0.32 * 0 = 0.0 \\
\delta_0(VB) * P(IN|VB) * P(the|IN) &= 0 * 0.23 * 0 = 0.0 \\
\delta_0(PERIOD) * P(IN|PERIOD) * P(the|IN) &= 1 * 0.09 * 0 = 0.0 \\
) & \\
\psi_1(IN) &= PERIOD
\end{aligned}$$

$$\begin{aligned}
\delta_1(NN) &= \max( \\
\delta_0(AT) * P(NN|AT) * P(the|NN) &= 0 * 0.4 * 0 = 0.0 \\
\delta_0(BEZ) * P(NN|BEZ) * P(the|NN) &= 0 * 0.02 * 0 = 0.0 \\
\delta_0(IN) * P(NN|IN) * P(the|NN) &= 0 * 0.13 * 0 = 0.0 \\
\delta_0(NN) * P(NN|NN) * P(the|NN) &= 0 * 0.09 * 0 = 0.0 \\
\delta_0(VB) * P(NN|VB) * P(the|NN) &= 0 * 0.07 * 0 = 0.0 \\
\delta_0(PERIOD) * P(NN|PERIOD) * P(the|NN) &= 1 * 0.03 * 0 = 0.0 \\
) & \\
\psi_1(NN) &= PERIOD
\end{aligned}$$

$$\begin{aligned}
\delta_1(VB) &= \max( \\
\delta_0(AT) * P(VB|AT) * P(the|VB) &= 0 * 0 * 0 = 0 \\
\delta_0(BEZ) * P(VB|BEZ) * P(the|VB) &= 0 * 0 * 0 = 0 \\
\delta_0(IN) * P(VB|IN) * P(the|VB) &= 0 * 0 * 0 = 0 \\
\delta_0(NN) * P(VB|NN) * P(the|VB) &= 0 * 0.005 * 0 = 0.0 \\
\delta_0(VB) * P(VB|VB) * P(the|VB) &= 0 * 0.006 * 0 = 0.0 \\
\delta_0(PERIOD) * P(VB|PERIOD) * P(the|VB) &= 1 * 0.02 * 0 = 0.0 \\
) & \\
\psi_1(VB) &= PERIOD
\end{aligned}$$

$$\begin{aligned}
\delta_1(PERIOD) &= \max( \\
\delta_0(AT) * P(PERIOD|AT) * P(the|PERIOD) &= 0 * 0.0002 * 0 = 0.0 \\
\delta_0(BEZ) * P(PERIOD|BEZ) * P(the|PERIOD) &= 0 * 0.04 * 0 = 0.0 \\
\delta_0(IN) * P(PERIOD|IN) * P(the|PERIOD) &= 0 * 0.001 * 0 = 0.0 \\
\delta_0(NN) * P(PERIOD|NN) * P(the|PERIOD) &= 0 * 0.16 * 0 = 0.0 \\
\delta_0(VB) * P(PERIOD|VB) * P(the|PERIOD) &= 0 * 0.07 * 0 = 0.0 \\
\delta_0(PERIOD) * P(PERIOD|PERIOD) * P(the|PERIOD) &= 1 * 0 * 0 = 0 \\
) & \\
\psi_1(PERIOD) &= PERIOD
\end{aligned}$$

Abbildung 1: 1. Induktionsschritt der Berechnung des Viterbi-Algorithmus (Viterbi-Wahrscheinlichkeiten für das Wort 'the')

Probability of the most likely path:  $\max_t \phi_6(t) = 1.8444754944e - 014$

6'th tag of the most likely path:  $\arg \max_t \phi_6(t) = NN$

5'th tag of the most likely path:  $\psi_6(NN) = AT$

4'th tag of the most likely path:  $\psi_5(AT) = IN$

3'rd tag of the most likely path:  $\psi_4(IN) = BEZ$

2'nd tag of the most likely path:  $\psi_3(BEZ) = NN$

1'st tag of the most likely path:  $\psi_2(NN) = AT$

### **Ergebnis**

AT NN BEZ IN AT NN

## **1.3 Hidden Markov Model Taggers**

Auch Hidden Markov Models (HMM) werden für das POS-Tagging eingesetzt.

1. Unter welchen Voraussetzungen wird man ein Hidden Markov Model (HMM) zum POS-Tagging verwenden?
2. Das POS-Tagging mittels HMM teilt sich in drei Abschnitte auf: Initialisierung, Training, Tagging. Beschreiben Sie kurz, was in den drei Abschnitten passiert. Wo liegen die Unterschiede / Parallelen zu den Visible Markov Models?

### LÖSUNGSVORSCHLAG:

1. Ein HMM wird dann verwendet, wenn kein getaggtter Trainingskorpus verfügbar ist.
2. Abschnitte:
  - Initialisierung: Initialisieren der Ausgabewahrscheinlichkeiten anhand eines Lexikons (Jelinek / Kupiec). Mittels des initialisierten Modells wird nun der Trainingskorpus getaggt.
  - Training: Mittels des Forward-Backward Algorithmus können nun die Parameter angepasst werden. Das erhaltene angepasste Modell wird dann verwendet, um den Trainingskorpus erneut zu taggen. Beides ist dann der Input für eine neue Runde des Forward-Backward-Algorithmus. Dies wird so lange wiederholt, bis sich beim Schätzen der Parameter die Parameter nur noch geringfügig verändern.
  - Tagging: Nach der Trainingsphase wird das HMM genau wie ein VMM zum Taggen eingesetzt, hier gibt es keine Unterschiede.



## 1.4 Transformation-Based Taggers

Markov Models liefern gute Ergebnisse, sind aber für manche Eigenschaften natürlicher Sprache zu starr. Eine andere Methode, die ein breiteres Spektrum von lexikalischen und syntaktischen Regularitäten erfassen kann, sind transformations-basierte Tagging-Ansätze.

1. Welche Komponenten hat ein transformations-basierter Ansatz? Welche Art von Input ist notwendig?
2. Bei der Initialisierung wird beim transformations-basierten Tagging im Trainingsdatensatz zunächst jedem Wort ein initiales Tag zugewiesen. Im Buch ist die Strategie beschrieben, hierfür das häufigste Tag (laut einem Wörterbuch) zu verwenden. Eine andere Möglichkeit wäre, jedem Wort das gleiche Tag (z.B. *NN*) zuzuweisen, oder die Trainingsdaten von einem externen Tagger initial taggen zu lassen. Welche Vor- und Nachteile hat jede dieser Alternativen?

LÖSUNGSVORSCHLAG:

1. Komponenten:
  - eine Liste mit erlaubten Transformationen
  - ein Lernalgorithmus
2. Input:
  - ein getagger Korpus
  - ein Lexikon
  - eine Liste mit erlaubten Transformationen
3. Unterschiedliche Initialisierungen müssen je nach Art der vorhandenen Transformationsregeln ausgesucht werden. Wenn diese schon ein halbwegs sinnvolles Tagging voraussetzen (z.B. bei vielen tag-triggered environments), kann der Input eines anderen Taggers angebracht sein. Gibt es viele word-triggered environments, kann ein einfaches Tagging mit einem einzigen Tag auch zum Erfolg führen - dessen Vorteil ist, dass es keine weiteren Daten braucht.

## 1.5 Vergleich

Vervollständigen Sie die Tabelle 1.5, die zum Vergleich der vorgestellten POS-Tagging Methoden dient. Fallen Ihnen noch andere Vergleichs-Dimensionen ein?

LÖSUNGSVORSCHLAG:

Tabelle 6: Vergleich verschiedener POS-Tagging Ansätze

	VMM	HMM	transformationsbasiert
nötiger Input			
Annahmen			
Gefahr des Overtrainings?			
...			

Tabelle 7: Vergleich verschiedener POS-Tagging Ansätze

	VMM	HMM	transformationsbasiert
nötiger Input	getaggtter Korpus	beliebiger Korpus	Liste mit erlaubten Transformationen; getaggtter Korpus; Wörterbuch
Annahmen	alle Markov-Eigenschaften; Unabhängigkeit der Wörter untereinander; ein Wort hängt nur von seinem eigenen Tag ab	s. VMM	keine
Gefahr des Overtrainings?	ja, zu stark an getaggtten Korpus angepasst	ja, je nach Initialisierung / training patterns (classical; early / initial maximum)	eher robust
...			

## 2 Praxisübung - POS Tagging

Das NLTK-Toolkit sind einige POS-Tagger enthalten, unter anderem ein HiddenMarkov-ModelTagger. Dieser kann auch trainiert werden.

- Implementieren und trainieren Sie einen HMM Tagger für eine Teilmenge des Brown-Corpus. Bestimmen Sie die Grösse der Teilmenge so, dass das Training in vertretbarer Zeit abläuft. Das Training soll hierbei unüberwacht mittels des Baum-Welch-Algorithmus stattfinden (dieser ist ebenfalls im NLTK-Toolkit verfügbar). Verwenden Sie dazu mindestens folgende Initialisierungen:
  - Ein Markov-Model, das mittels einem überwachten Lernverfahren gelernt wurde
  - Eine Initialisierung mittels Jelinek's Methode
- Wenden Sie Ihren Tagger auf folgende Testdaten an und vergleichen Sie die Accuracy des Taggings:
  - dieselbe Teilmenge, auf der der Tagger trainiert wurde
  - eine andere Teilmenge, ggf. den ganzen Korpus

Wie verändert sich die Qualität Ihrer Ergebnisse in Abhängigkeit der Initialisierung? Wie verändert sich die Qualität Ihrer Ergebnisse in zwischen mehreren Testläufen mit derselben Initialisierung?

- *Zusatzaufgabe* (freiwillig): Vergleichen Sie die Ergebnisse des gelernten HMM-Taggers mit in NLTK verfügbaren Standard-Taggern (Brill, Bigramm, . . .) oder eine Kombination aus diesen.

**Hinweis:** Im NLTK-Quellcode `nltk/tag/hmm.py` finden Sie bereits einige Beispiele, die Sie als Einstieg verwenden können.

### 3 Anhang: vollständige Weiterführung des Viterbi-Algorithmus (1.2, Frage 4)

Viterbi-Wahrscheinlichkeiten für das Wort 'bear'

$$\begin{aligned} \delta_2(AT) &= \max( \\ &\delta_1(AT) * P(AT|AT) * P(bear|AT) = 0.0912 * 0 * 0 = 0.0 \\ &\delta_1(BEZ) * P(AT|BEZ) * P(bear|AT) = 0 * 0.2 * 0 = 0.0 \\ &\delta_1(IN) * P(AT|IN) * P(bear|AT) = 0 * 0.33 * 0 = 0.0 \\ &\delta_1(NN) * P(AT|NN) * P(bear|AT) = 0 * 0.008 * 0 = 0.0 \\ &\delta_1(VB) * P(AT|VB) * P(bear|AT) = 0 * 0.29 * 0 = 0.0 \\ &\delta_1(PERIOD) * P(AT|PERIOD) * P(bear|AT) = 0 * 0.16 * 0 = 0.0 \\ & ) \\ \psi_2(AT) &= PERIOD \end{aligned}$$

$$\begin{aligned} \delta_2(BEZ) &= \max( \\ &\delta_1(AT) * P(BEZ|AT) * P(bear|BEZ) = 0.0912 * 0 * 0 = 0.0 \\ &\delta_1(BEZ) * P(BEZ|BEZ) * P(bear|BEZ) = 0 * 0 * 0 = 0 \\ &\delta_1(IN) * P(BEZ|IN) * P(bear|BEZ) = 0 * 0 * 0 = 0 \\ &\delta_1(NN) * P(BEZ|NN) * P(bear|BEZ) = 0 * 0.03 * 0 = 0.0 \\ &\delta_1(VB) * P(BEZ|VB) * P(bear|BEZ) = 0 * 0.002 * 0 = 0.0 \\ &\delta_1(PERIOD) * P(BEZ|PERIOD) * P(bear|BEZ) = 0 * 0.002 * 0 = 0.0 \\ & ) \\ \psi_2(BEZ) &= PERIOD \end{aligned}$$

$$\begin{aligned} \delta_2(IN) &= \max( \\ &\delta_1(AT) * P(IN|AT) * P(bear|IN) = 0.0912 * 0 * 0 = 0.0 \\ &\delta_1(BEZ) * P(IN|BEZ) * P(bear|IN) = 0 * 0.04 * 0 = 0.0 \\ &\delta_1(IN) * P(IN|IN) * P(bear|IN) = 0 * 0.01 * 0 = 0.0 \\ &\delta_1(NN) * P(IN|NN) * P(bear|IN) = 0 * 0.32 * 0 = 0.0 \\ &\delta_1(VB) * P(IN|VB) * P(bear|IN) = 0 * 0.23 * 0 = 0.0 \\ &\delta_1(PERIOD) * P(IN|PERIOD) * P(bear|IN) = 0 * 0.09 * 0 = 0.0 \\ & ) \\ \psi_2(IN) &= PERIOD \end{aligned}$$

$$\begin{aligned} \delta_2(NN) &= \max( \\ &\delta_1(AT) * P(NN|AT) * P(bear|NN) = 0.0912 * 0.4 * 7e - 005 = 2.5536e - 006 \\ &\delta_1(BEZ) * P(NN|BEZ) * P(bear|NN) = 0 * 0.02 * 7e - 005 = 0.0 \\ &\delta_1(IN) * P(NN|IN) * P(bear|NN) = 0 * 0.13 * 7e - 005 = 0.0 \\ &\delta_1(NN) * P(NN|NN) * P(bear|NN) = 0 * 0.09 * 7e - 005 = 0.0 \\ &\delta_1(VB) * P(NN|VB) * P(bear|NN) = 0 * 0.07 * 7e - 005 = 0.0 \\ &\delta_1(PERIOD) * P(NN|PERIOD) * P(bear|NN) = 0 * 0.03 * 7e - 005 = 0.0 \\ & ) \end{aligned}$$

$$\psi_2(NN) = AT$$

$$\begin{aligned} \delta_2(VB) = \max( \\ & \delta_1(AT) * P(VB|AT) * P(bear|VB) = 0.0912 * 0 * 0.002 = 0.0 \\ & \delta_1(BEZ) * P(VB|BEZ) * P(bear|VB) = 0 * 0 * 0.002 = 0.0 \\ & \delta_1(IN) * P(VB|IN) * P(bear|VB) = 0 * 0 * 0.002 = 0.0 \\ & \delta_1(NN) * P(VB|NN) * P(bear|VB) = 0 * 0.005 * 0.002 = 0.0 \\ & \delta_1(VB) * P(VB|VB) * P(bear|VB) = 0 * 0.006 * 0.002 = 0.0 \\ & \delta_1(PERIOD) * P(VB|PERIOD) * P(bear|VB) = 0 * 0.02 * 0.002 = 0.0 \\ & ) \\ \psi_2(VB) = AT \end{aligned}$$

$$\begin{aligned} \delta_2(PERIOD) = \max( \\ & \delta_1(AT) * P(PERIOD|AT) * P(bear|PERIOD) = 0.0912 * 0.0002 * 0 = 0.0 \\ & \delta_1(BEZ) * P(PERIOD|BEZ) * P(bear|PERIOD) = 0 * 0.04 * 0 = 0.0 \\ & \delta_1(IN) * P(PERIOD|IN) * P(bear|PERIOD) = 0 * 0.001 * 0 = 0.0 \\ & \delta_1(NN) * P(PERIOD|NN) * P(bear|PERIOD) = 0 * 0.16 * 0 = 0.0 \\ & \delta_1(VB) * P(PERIOD|VB) * P(bear|PERIOD) = 0 * 0.07 * 0 = 0.0 \\ & \delta_1(PERIOD) * P(PERIOD|PERIOD) * P(bear|PERIOD) = 0 * 0 * 0 = 0.0 \\ & ) \\ \psi_2(PERIOD) = AT \end{aligned}$$

### Viterbi-Wahrscheinlichkeiten für das Wort 'is'

$$\begin{aligned} \delta_3(AT) = \max( \\ & \delta_2(AT) * P(AT|AT) * P(is|AT) = 0 * 0 * 0 = 0 \\ & \delta_2(BEZ) * P(AT|BEZ) * P(is|AT) = 0 * 0.2 * 0 = 0.0 \\ & \delta_2(IN) * P(AT|IN) * P(is|AT) = 0 * 0.33 * 0 = 0.0 \\ & \delta_2(NN) * P(AT|NN) * P(is|AT) = 2.5536e - 006 * 0.008 * 0 = 0.0 \\ & \delta_2(VB) * P(AT|VB) * P(is|AT) = 0 * 0.29 * 0 = 0.0 \\ & \delta_2(PERIOD) * P(AT|PERIOD) * P(is|AT) = 0 * 0.16 * 0 = 0.0 \\ & ) \\ \psi_3(AT) = AT \end{aligned}$$

$$\begin{aligned} \delta_3(BEZ) = \max( \\ & \delta_2(AT) * P(BEZ|AT) * P(is|BEZ) = 0 * 0 * 1 = 0 \\ & \delta_2(BEZ) * P(BEZ|BEZ) * P(is|BEZ) = 0 * 0 * 1 = 0 \\ & \delta_2(IN) * P(BEZ|IN) * P(is|BEZ) = 0 * 0 * 1 = 0 \\ & \delta_2(NN) * P(BEZ|NN) * P(is|BEZ) = 2.5536e - 006 * 0.03 * 1 = 7.6608e - 008 \\ & \delta_2(VB) * P(BEZ|VB) * P(is|BEZ) = 0 * 0.002 * 1 = 0.0 \\ & \delta_2(PERIOD) * P(BEZ|PERIOD) * P(is|BEZ) = 0 * 0.002 * 1 = 0.0 \\ & ) \\ \psi_3(BEZ) = NN \end{aligned}$$

$$\begin{aligned}
\delta_3(IN) &= \max( \\
&\delta_2(AT) * P(IN|AT) * P(is|IN) = 0 * 0 * 0 = 0 \\
&\delta_2(BEZ) * P(IN|BEZ) * P(is|IN) = 0 * 0.04 * 0 = 0.0 \\
&\delta_2(IN) * P(IN|IN) * P(is|IN) = 0 * 0.01 * 0 = 0.0 \\
&\delta_2(NN) * P(IN|NN) * P(is|IN) = 2.5536e - 006 * 0.32 * 0 = 0.0 \\
&\delta_2(VB) * P(IN|VB) * P(is|IN) = 0 * 0.23 * 0 = 0.0 \\
&\delta_2(PERIOD) * P(IN|PERIOD) * P(is|IN) = 0 * 0.09 * 0 = 0.0 \\
& ) \\
\psi_3(IN) &= NN
\end{aligned}$$

$$\begin{aligned}
\delta_3(NN) &= \max( \\
&\delta_2(AT) * P(NN|AT) * P(is|NN) = 0 * 0.4 * 0 = 0.0 \\
&\delta_2(BEZ) * P(NN|BEZ) * P(is|NN) = 0 * 0.02 * 0 = 0.0 \\
&\delta_2(IN) * P(NN|IN) * P(is|NN) = 0 * 0.13 * 0 = 0.0 \\
&\delta_2(NN) * P(NN|NN) * P(is|NN) = 2.5536e - 006 * 0.09 * 0 = 0.0 \\
&\delta_2(VB) * P(NN|VB) * P(is|NN) = 0 * 0.07 * 0 = 0.0 \\
&\delta_2(PERIOD) * P(NN|PERIOD) * P(is|NN) = 0 * 0.03 * 0 = 0.0 \\
& ) \\
\psi_3(NN) &= NN
\end{aligned}$$

$$\begin{aligned}
\delta_3(VB) &= \max( \\
&\delta_2(AT) * P(VB|AT) * P(is|VB) = 0 * 0 * 0 = 0 \\
&\delta_2(BEZ) * P(VB|BEZ) * P(is|VB) = 0 * 0 * 0 = 0 \\
&\delta_2(IN) * P(VB|IN) * P(is|VB) = 0 * 0 * 0 = 0 \\
&\delta_2(NN) * P(VB|NN) * P(is|VB) = 2.5536e - 006 * 0.005 * 0 = 0.0 \\
&\delta_2(VB) * P(VB|VB) * P(is|VB) = 0 * 0.006 * 0 = 0.0 \\
&\delta_2(PERIOD) * P(VB|PERIOD) * P(is|VB) = 0 * 0.02 * 0 = 0.0 \\
& ) \\
\psi_3(VB) &= NN
\end{aligned}$$

$$\begin{aligned}
\delta_3(PERIOD) &= \max( \\
&\delta_2(AT) * P(PERIOD|AT) * P(is|PERIOD) = 0 * 0.0002 * 0 = 0.0 \\
&\delta_2(BEZ) * P(PERIOD|BEZ) * P(is|PERIOD) = 0 * 0.04 * 0 = 0.0 \\
&\delta_2(IN) * P(PERIOD|IN) * P(is|PERIOD) = 0 * 0.001 * 0 = 0.0 \\
&\delta_2(NN) * P(PERIOD|NN) * P(is|PERIOD) = 2.5536e - 006 * 0.16 * 0 = 0.0 \\
&\delta_2(VB) * P(PERIOD|VB) * P(is|PERIOD) = 0 * 0.07 * 0 = 0.0 \\
&\delta_2(PERIOD) * P(PERIOD|PERIOD) * P(is|PERIOD) = 0 * 0 * 0 = 0 \\
& ) \\
\psi_3(PERIOD) &= NN
\end{aligned}$$

**Viterbi-Wahrscheinlichkeiten für das Wort 'on'**

$$\begin{aligned}
&\delta_4(AT) = \max( \\
&\delta_3(AT) * P(AT|AT) * P(on|AT) = 0 * 0 * 0 = 0 \\
&\delta_3(BEZ) * P(AT|BEZ) * P(on|AT) = 7.6608e - 008 * 0.2 * 0 = 0.0 \\
&\delta_3(IN) * P(AT|IN) * P(on|AT) = 0 * 0.33 * 0 = 0.0 \\
&\delta_3(NN) * P(AT|NN) * P(on|AT) = 0 * 0.008 * 0 = 0.0 \\
&\delta_3(VB) * P(AT|VB) * P(on|AT) = 0 * 0.29 * 0 = 0.0 \\
&\delta_3(PERIOD) * P(AT|PERIOD) * P(on|AT) = 0 * 0.16 * 0 = 0.0 \\
& ) \\
&\psi_4(AT) = NN
\end{aligned}$$

$$\begin{aligned}
&\delta_4(BEZ) = \max( \\
&\delta_3(AT) * P(BEZ|AT) * P(on|BEZ) = 0 * 0 * 0 = 0 \\
&\delta_3(BEZ) * P(BEZ|BEZ) * P(on|BEZ) = 7.6608e - 008 * 0 * 0 = 0.0 \\
&\delta_3(IN) * P(BEZ|IN) * P(on|BEZ) = 0 * 0 * 0 = 0 \\
&\delta_3(NN) * P(BEZ|NN) * P(on|BEZ) = 0 * 0.03 * 0 = 0.0 \\
&\delta_3(VB) * P(BEZ|VB) * P(on|BEZ) = 0 * 0.002 * 0 = 0.0 \\
&\delta_3(PERIOD) * P(BEZ|PERIOD) * P(on|BEZ) = 0 * 0.002 * 0 = 0.0 \\
& ) \\
&\psi_4(BEZ) = NN
\end{aligned}$$

$$\begin{aligned}
&\delta_4(IN) = \max( \\
&\delta_3(AT) * P(IN|AT) * P(on|IN) = 0 * 0 * 0.04 = 0.0 \\
&\delta_3(BEZ) * P(IN|BEZ) * P(on|IN) = 7.6608e - 008 * 0.04 * 0.04 = 1.225728e - 010 \\
&\delta_3(IN) * P(IN|IN) * P(on|IN) = 0 * 0.01 * 0.04 = 0.0 \\
&\delta_3(NN) * P(IN|NN) * P(on|IN) = 0 * 0.32 * 0.04 = 0.0 \\
&\delta_3(VB) * P(IN|VB) * P(on|IN) = 0 * 0.23 * 0.04 = 0.0 \\
&\delta_3(PERIOD) * P(IN|PERIOD) * P(on|IN) = 0 * 0.09 * 0.04 = 0.0 \\
& ) \\
&\psi_4(IN) = BEZ
\end{aligned}$$

$$\begin{aligned}
&\delta_4(NN) = \max( \\
&\delta_3(AT) * P(NN|AT) * P(on|NN) = 0 * 0.4 * 0 = 0.0 \\
&\delta_3(BEZ) * P(NN|BEZ) * P(on|NN) = 7.6608e - 008 * 0.02 * 0 = 0.0 \\
&\delta_3(IN) * P(NN|IN) * P(on|NN) = 0 * 0.13 * 0 = 0.0 \\
&\delta_3(NN) * P(NN|NN) * P(on|NN) = 0 * 0.09 * 0 = 0.0 \\
&\delta_3(VB) * P(NN|VB) * P(on|NN) = 0 * 0.07 * 0 = 0.0 \\
&\delta_3(PERIOD) * P(NN|PERIOD) * P(on|NN) = 0 * 0.03 * 0 = 0.0 \\
& ) \\
&\psi_4(NN) = BEZ
\end{aligned}$$

$$\begin{aligned}
&\delta_4(VB) = \max( \\
&\delta_3(AT) * P(VB|AT) * P(on|VB) = 0 * 0 * 0 = 0 \\
&\delta_3(BEZ) * P(VB|BEZ) * P(on|VB) = 7.6608e - 008 * 0 * 0 = 0.0 \\
&\delta_3(IN) * P(VB|IN) * P(on|VB) = 0 * 0 * 0 = 0 \\
& )
\end{aligned}$$

$$\begin{aligned}
&\delta_3(NN) * P(VB|NN) * P(on|VB) = 0 * 0.005 * 0 = 0.0 \\
&\delta_3(VB) * P(VB|VB) * P(on|VB) = 0 * 0.006 * 0 = 0.0 \\
&\delta_3(PERIOD) * P(VB|PERIOD) * P(on|VB) = 0 * 0.02 * 0 = 0.0 \\
& ) \\
&\psi_4(VB) = BEZ
\end{aligned}$$

$$\begin{aligned}
&\delta_4(PERIOD) = max( \\
&\delta_3(AT) * P(PERIOD|AT) * P(on|PERIOD) = 0 * 0.0002 * 0 = 0.0 \\
&\delta_3(BEZ) * P(PERIOD|BEZ) * P(on|PERIOD) = 7.6608e - 008 * 0.04 * 0 = 0.0 \\
&\delta_3(IN) * P(PERIOD|IN) * P(on|PERIOD) = 0 * 0.001 * 0 = 0.0 \\
&\delta_3(NN) * P(PERIOD|NN) * P(on|PERIOD) = 0 * 0.16 * 0 = 0.0 \\
&\delta_3(VB) * P(PERIOD|VB) * P(on|PERIOD) = 0 * 0.07 * 0 = 0.0 \\
&\delta_3(PERIOD) * P(PERIOD|PERIOD) * P(on|PERIOD) = 0 * 0 * 0 = 0.0 \\
& ) \\
&\psi_4(PERIOD) = BEZ
\end{aligned}$$

#### Viterbi-Wahrscheinlichkeiten für das Wort 'the'

$$\begin{aligned}
&\delta_5(AT) = max( \\
&\delta_4(AT) * P(AT|AT) * P(the|AT) = 0 * 0 * 0.57 = 0.0 \\
&\delta_4(BEZ) * P(AT|BEZ) * P(the|AT) = 0 * 0.2 * 0.57 = 0.0 \\
&\delta_4(IN) * P(AT|IN) * P(the|AT) = 1.225728e - 010 * 0.33 * 0.57 = 2.305594368e - 011 \\
&\delta_4(NN) * P(AT|NN) * P(the|AT) = 0 * 0.008 * 0.57 = 0.0 \\
&\delta_4(VB) * P(AT|VB) * P(the|AT) = 0 * 0.29 * 0.57 = 0.0 \\
&\delta_4(PERIOD) * P(AT|PERIOD) * P(the|AT) = 0 * 0.16 * 0.57 = 0.0 \\
& ) \\
&\psi_5(AT) = IN
\end{aligned}$$

$$\begin{aligned}
&\delta_5(BEZ) = max( \\
&\delta_4(AT) * P(BEZ|AT) * P(the|BEZ) = 0 * 0 * 0 = 0 \\
&\delta_4(BEZ) * P(BEZ|BEZ) * P(the|BEZ) = 0 * 0 * 0 = 0 \\
&\delta_4(IN) * P(BEZ|IN) * P(the|BEZ) = 1.225728e - 010 * 0 * 0 = 0.0 \\
&\delta_4(NN) * P(BEZ|NN) * P(the|BEZ) = 0 * 0.03 * 0 = 0.0 \\
&\delta_4(VB) * P(BEZ|VB) * P(the|BEZ) = 0 * 0.002 * 0 = 0.0 \\
&\delta_4(PERIOD) * P(BEZ|PERIOD) * P(the|BEZ) = 0 * 0.002 * 0 = 0.0 \\
& ) \\
&\psi_5(BEZ) = IN
\end{aligned}$$

$$\begin{aligned}
&\delta_5(IN) = max( \\
&\delta_4(AT) * P(IN|AT) * P(the|IN) = 0 * 0 * 0 = 0 \\
&\delta_4(BEZ) * P(IN|BEZ) * P(the|IN) = 0 * 0.04 * 0 = 0.0 \\
&\delta_4(IN) * P(IN|IN) * P(the|IN) = 1.225728e - 010 * 0.01 * 0 = 0.0 \\
&\delta_4(NN) * P(IN|NN) * P(the|IN) = 0 * 0.32 * 0 = 0.0
\end{aligned}$$



$$\begin{aligned} \delta_4(VB) * P(IN|VB) * P(the|IN) &= 0 * 0.23 * 0 = 0.0 \\ \delta_4(PERIOD) * P(IN|PERIOD) * P(the|IN) &= 0 * 0.09 * 0 = 0.0 \\ ) \\ \psi_5(IN) &= IN \end{aligned}$$

$$\begin{aligned} \delta_5(NN) &= \max( \\ \delta_4(AT) * P(NN|AT) * P(the|NN) &= 0 * 0.4 * 0 = 0.0 \\ \delta_4(BEZ) * P(NN|BEZ) * P(the|NN) &= 0 * 0.02 * 0 = 0.0 \\ \delta_4(IN) * P(NN|IN) * P(the|NN) &= 1.225728e - 010 * 0.13 * 0 = 0.0 \\ \delta_4(NN) * P(NN|NN) * P(the|NN) &= 0 * 0.09 * 0 = 0.0 \\ \delta_4(VB) * P(NN|VB) * P(the|NN) &= 0 * 0.07 * 0 = 0.0 \\ \delta_4(PERIOD) * P(NN|PERIOD) * P(the|NN) &= 0 * 0.03 * 0 = 0.0 \\ ) \\ \psi_5(NN) &= IN \end{aligned}$$

$$\begin{aligned} \delta_5(VB) &= \max( \\ \delta_4(AT) * P(VB|AT) * P(the|VB) &= 0 * 0 * 0 = 0 \\ \delta_4(BEZ) * P(VB|BEZ) * P(the|VB) &= 0 * 0 * 0 = 0 \\ \delta_4(IN) * P(VB|IN) * P(the|VB) &= 1.225728e - 010 * 0 * 0 = 0.0 \\ \delta_4(NN) * P(VB|NN) * P(the|VB) &= 0 * 0.005 * 0 = 0.0 \\ \delta_4(VB) * P(VB|VB) * P(the|VB) &= 0 * 0.006 * 0 = 0.0 \\ \delta_4(PERIOD) * P(VB|PERIOD) * P(the|VB) &= 0 * 0.02 * 0 = 0.0 \\ ) \\ \psi_5(VB) &= IN \end{aligned}$$

$$\begin{aligned} \delta_5(PERIOD) &= \max( \\ \delta_4(AT) * P(PERIOD|AT) * P(the|PERIOD) &= 0 * 0.0002 * 0 = 0.0 \\ \delta_4(BEZ) * P(PERIOD|BEZ) * P(the|PERIOD) &= 0 * 0.04 * 0 = 0.0 \\ \delta_4(IN) * P(PERIOD|IN) * P(the|PERIOD) &= 1.225728e - 010 * 0.001 * 0 = 0.0 \\ \delta_4(NN) * P(PERIOD|NN) * P(the|PERIOD) &= 0 * 0.16 * 0 = 0.0 \\ \delta_4(VB) * P(PERIOD|VB) * P(the|PERIOD) &= 0 * 0.07 * 0 = 0.0 \\ \delta_4(PERIOD) * P(PERIOD|PERIOD) * P(the|PERIOD) &= 0 * 0 * 0 = 0 \\ ) \\ \psi_5(PERIOD) &= IN \end{aligned}$$

#### Viterbi-Wahrscheinlichkeiten für das Wort 'move'

$$\begin{aligned} \delta_6(AT) &= \max( \\ \delta_5(AT) * P(AT|AT) * P(move|AT) &= 2.305594368e - 011 * 0 * 0 = 0.0 \\ \delta_5(BEZ) * P(AT|BEZ) * P(move|AT) &= 0 * 0.2 * 0 = 0.0 \\ \delta_5(IN) * P(AT|IN) * P(move|AT) &= 0 * 0.33 * 0 = 0.0 \\ \delta_5(NN) * P(AT|NN) * P(move|AT) &= 0 * 0.008 * 0 = 0.0 \\ \delta_5(VB) * P(AT|VB) * P(move|AT) &= 0 * 0.29 * 0 = 0.0 \end{aligned}$$

$$\delta_5(PERIOD) * P(AT|PERIOD) * P(move|AT) = 0 * 0.16 * 0 = 0.0$$

$$)$$

$$\psi_6(AT) = IN$$

$$\delta_6(BEZ) = max($$

$$\delta_5(AT) * P(BEZ|AT) * P(move|BEZ) = 2.305594368e - 011 * 0 * 0 = 0.0$$

$$\delta_5(BEZ) * P(BEZ|BEZ) * P(move|BEZ) = 0 * 0 * 0 = 0$$

$$\delta_5(IN) * P(BEZ|IN) * P(move|BEZ) = 0 * 0 * 0 = 0$$

$$\delta_5(NN) * P(BEZ|NN) * P(move|BEZ) = 0 * 0.03 * 0 = 0.0$$

$$\delta_5(VB) * P(BEZ|VB) * P(move|BEZ) = 0 * 0.002 * 0 = 0.0$$

$$\delta_5(PERIOD) * P(BEZ|PERIOD) * P(move|BEZ) = 0 * 0.002 * 0 = 0.0$$

$$)$$

$$\psi_6(BEZ) = IN$$

$$\delta_6(IN) = max($$

$$\delta_5(AT) * P(IN|AT) * P(move|IN) = 2.305594368e - 011 * 0 * 0 = 0.0$$

$$\delta_5(BEZ) * P(IN|BEZ) * P(move|IN) = 0 * 0.04 * 0 = 0.0$$

$$\delta_5(IN) * P(IN|IN) * P(move|IN) = 0 * 0.01 * 0 = 0.0$$

$$\delta_5(NN) * P(IN|NN) * P(move|IN) = 0 * 0.32 * 0 = 0.0$$

$$\delta_5(VB) * P(IN|VB) * P(move|IN) = 0 * 0.23 * 0 = 0.0$$

$$\delta_5(PERIOD) * P(IN|PERIOD) * P(move|IN) = 0 * 0.09 * 0 = 0.0$$

$$)$$

$$\psi_6(IN) = IN$$

$$\delta_6(NN) = max($$

$$\delta_5(AT) * P(NN|AT) * P(move|NN) = 2.305594368e - 011 * 0.4 * 0.002 = 1.8444754944e - 014$$

$$\delta_5(BEZ) * P(NN|BEZ) * P(move|NN) = 0 * 0.02 * 0.002 = 0.0$$

$$\delta_5(IN) * P(NN|IN) * P(move|NN) = 0 * 0.13 * 0.002 = 0.0$$

$$\delta_5(NN) * P(NN|NN) * P(move|NN) = 0 * 0.09 * 0.002 = 0.0$$

$$\delta_5(VB) * P(NN|VB) * P(move|NN) = 0 * 0.07 * 0.002 = 0.0$$

$$\delta_5(PERIOD) * P(NN|PERIOD) * P(move|NN) = 0 * 0.03 * 0.002 = 0.0$$

$$)$$

$$\psi_6(NN) = AT$$

$$\delta_6(VB) = max($$

$$\delta_5(AT) * P(VB|AT) * P(move|VB) = 2.305594368e - 011 * 0 * 0.006 = 0.0$$

$$\delta_5(BEZ) * P(VB|BEZ) * P(move|VB) = 0 * 0 * 0.006 = 0.0$$

$$\delta_5(IN) * P(VB|IN) * P(move|VB) = 0 * 0 * 0.006 = 0.0$$

$$\delta_5(NN) * P(VB|NN) * P(move|VB) = 0 * 0.005 * 0.006 = 0.0$$

$$\delta_5(VB) * P(VB|VB) * P(move|VB) = 0 * 0.006 * 0.006 = 0.0$$

$$\delta_5(PERIOD) * P(VB|PERIOD) * P(move|VB) = 0 * 0.02 * 0.006 = 0.0$$

$$)$$

$$\psi_6(VB) = AT$$

$$\begin{aligned}
\delta_6(PERIOD) &= \max( \\
&\delta_5(AT) * P(PERIOD|AT) * P(move|PERIOD) = 2.305594368e - 011 * 0.0002 * 0 = 0.0 \\
&\delta_5(BEZ) * P(PERIOD|BEZ) * P(move|PERIOD) = 0 * 0.04 * 0 = 0.0 \\
&\delta_5(IN) * P(PERIOD|IN) * P(move|PERIOD) = 0 * 0.001 * 0 = 0.0 \\
&\delta_5(NN) * P(PERIOD|NN) * P(move|PERIOD) = 0 * 0.16 * 0 = 0.0 \\
&\delta_5(VB) * P(PERIOD|VB) * P(move|PERIOD) = 0 * 0.07 * 0 = 0.0 \\
&\delta_5(PERIOD) * P(PERIOD|PERIOD) * P(move|PERIOD) = 0 * 0 * 0 = 0 \\
& ) \\
\psi_6(PERIOD) &= AT
\end{aligned}$$