Faculty of Computer Science, Dalhousie University *4-Nov-2024*

CSCI 4152/6509 — Natural Language Processing

Lecture 16: Efficient Inference for Bayesian Networks and HMMs

Location: Carleton Tupper Building Theatre C Instructor: Vlado Keselj
Time: 16:05 – 17:25

Previous Lecture

- HMM POS example (continued)
- HMM Brute-force approach
- HMM Inference: Viterbi algorithm
- **HMM as Bayesian Network**
- Bayesian Network definition
- Burglar-earthquake example
- BN inference using brute force
- Complexity of general inference in BNs
- Sum-product algorithms (started)

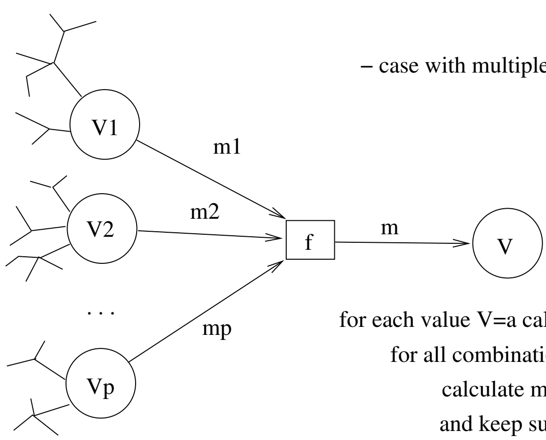
Slide notes:

Four Cases of Message Computation (repeated)

- Actually, we can distinguish 4 cases of message computation:

1. Factor node with multiple neighbours to variable node
2. Factor leaf node to variable node
3. Variable node with multiple neighbours to factor node
4. Variable leaf node to factor node

Factor Node with Multiple Neighbours Passing a Message to Variable Node



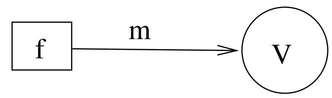
– case with multiple neighbours

for each value V=a calculate m(a):
for all combinations of V1 .. Vp
calculate m1*m2*..mp*f
and keep sum or max

m(a) is resulting sum or max

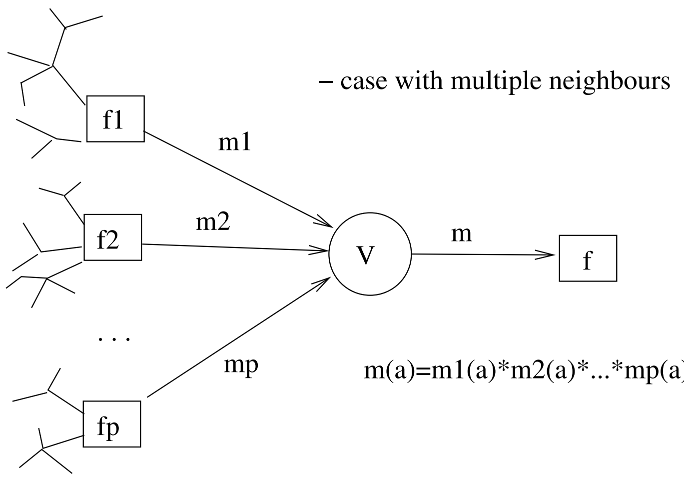
Factor Node with No Other Neighbours Passing a Message to Variable Node

– case with no other neighbours



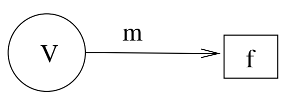
for each value V=a : m(a) = f(a)

Variable Node with Multiple Neighbours Passing a Message to Factor Node



Variable Node with No Other Neighbours Passing a Message to Factor Node

– case with no other neighbours



for each value a of V: m(a) = 1

16.3.1 Solving Inference Tasks with Message-Passing Algorithms

Slide notes:

Solving Inference Tasks

– Distinguish the following cases of inference tasks:

1. Marginalization with one variable
2. Marginalization in general
3. Conditioning with one variable
4. Conditioning in general
5. Completion

Message-Passing Algorithm for Marginalization with One Variable

We first consider the algorithm when we marginalize only on one variable.

Slide notes:

Marginalization with One Variable

- $\mathrm{P}(V_i = x_i) = ?$
- Apply general message passing rules with summation
- At the end

$$\mathrm{P}(V_i = x_i) \quad = \quad M_{f_1 \to V_i}(x_i) \cdots M_{f_p \to V_i}(x_i)$$

- Running time: $O(nm^{p+1})$

- Messages M are vectors of real numbers $(u_1, ..., u_m)$, where u_k is a number that summarizes the computation for the case $V = d_k$, where the domain for all variables is $\{d_1, d_2, \ldots, d_m\}$.
- Messages are passed from variable nodes to function nodes, and from function nodes to variable nodes.
- A node can send a message to its neighbor only when it has received all of the messages from its other neighbors.
- Given a tree, the algorithm can start by sending messages from each of the leaves, and stops once every node has passed a message to every neighbor. At the end, two messages will pass each edge in the graph: one for each of the two directions.
- Function to variable messages $M_{f \to V}(x)$ are computed by

$$M_{f \to V}(x) = \sum_{x_1, ..., x_p} f(x, x_1, ..., x_p) M_{V_1 \to f}(x_1) \cdots M_{V_p \to f}(x_p)$$

  over all other variables $V_1, ..., V_p$ (beside V) connected to f. If f is connected only to V, then $M_{f \to V}(x) = f(x)$.
- Variable to function messages $M_{V \to f}(x)$ are computed by

$$M_{V \to f}(x) \quad = \quad \begin{cases} 1 & \text{if only } f \text{ is connected to } V \\ M_{f_1 \to V}(x) \cdots M_{f_p \to V}(x) & \text{otherwise} \end{cases}$$

  over all other functions $f_1, ..., f_p$ (beside f) adjacent to V.
- Once all of the messages have been passed, then the final marginal for any variable V_i can be calculated by

$$\mathrm{P}(V_i = x_i) \quad = \quad M_{f_1 \to V_i}(x_i) \cdots M_{f_p \to V_i}(x_i)$$

  for all $f_1, ..., f_p$ adjacent to V_j.

This algorithm is efficient: There are $2n - 1$ edges in an undirected tree containing $2n$ nodes (n variables and n function nodes). $2(2n - 1)$ messages get sent (one in each direction along each edge). Each function to variable message can be computed in time $O(m^p)$ where p is the number of function neighbors, each variable to function message can be computed in time $O(mp)$ where p is the number of variable neighbors, and the final marginal can be computed in time $O(mp)$. Thus, the total running time is bounded by $O(nm^p)$ where p is the maximum number of neighbors of any node in the graph. This is linear in n and polynomial in m (but exponential in p, so the maximum number of neighbors has to be bounded).

Message-passing Algorithm for Marginalization in General

Slide notes:

Marginalization in General

- Consider calculating $\mathrm{P}(V_1 = x_1, ..., V_k = x_k)$.
- The variables $V_1, ..., V_k$ are called evidence variables and the instantiated values $x_1, \ldots, x_k$ are called observed evidence.
- An evidence-variable to function message is computed in the same way as before if $x = x_j$ (i.e., it is equal to observed evidence), otherwise it is 0.
- Final computation is done in any evidence node V_j:

$$\mathrm{P}(V_1 = x_1, ..., V_k = x_k) \quad = \quad M_{f_1 \to V_j}(x_j) \cdots M_{f_p \to V_j}(x_j)$$

Consider computing the marginal of one particular partial configuration $\mathrm{P}(V_1 = x_1, ..., V_k = x_k)$. The variables $V_1, ..., V_k$ are called evidence variables and the instantiated values $x_1, \ldots, x_k$ are called observed evidence. Then we can compute the desired probability by using the same message passing algorithm as above, except:

- An evidence-variable to function message is computed in the same way as before if $x = x_j$ (i.e., it is equal to observed evidence), otherwise it is 0. I.e.,

$$M_{V \to f}(x) \quad = \quad \begin{cases} 0 & \text{if } x \neq x_j \\ 1 & \text{if } x = x_j \text{ and only } f \text{ is adjacent to } V \\ M_{f_1 \to V}(x) \cdots M_{f_p \to V}(x) & \text{otherwise } (x = x_j) \end{cases}$$

  over all other functions $f_1, ..., f_p$ (besides f) adjacent to V.
- Once all of the messages have been passed, then the final marginal can be determined by taking any evidence variable $V_j \in \{V_1, ..., V_k\}$ and computing

$$\mathrm{P}(V_1 = x_1, ..., V_k = x_k) \quad = \quad M_{f_1 \to V_j}(x_j) \cdots M_{f_p \to V_j}(x_j)$$

  over all $f_1, ..., f_p$ adjacent to V_j.

Message-passing Algorithm for Conditioning with One Variable

Let us assume that we need to calculate the following conditional probability: $\mathrm{P}(V_{k+1} = y_{k+1} | V_1 = x_1, ..., V_k = x_k)$. We can use the same message passing algorithm as above, treating $V_1, ..., V_k$ as evidence variables, except that

- once all of the messages have been passed, then the final conditional probability can be determined by

$$\begin{aligned} &\mathrm{P}(V_{k+1} = y_{k+1} | V_1 = x_1, ..., V_k = x_k) \\ = \; &\frac{M_{f_1 \to V_{k+1}}(y_{k+1}) \cdots M_{f_p \to V_{k+1}}(y_{k+1})}{Z} \end{aligned}$$

  where Z is a normalization constant over choices of V_{k+1}; that is,

$$Z \quad = \quad \sum_y M_{f_1 \to V_{k+1}}(y) \cdots M_{f_p \to V_{k+1}}(y)$$

Message-passing Algorithm for Conditioning in General

To compute arbitrary conditional probability $\mathrm{P}(\mathbf{V}_\alpha = \mathbf{y}_\alpha | \mathbf{V}_\beta = \mathbf{x}_\beta)$, where α and β are two disjoint sets of indices from $\{1, \ldots, n\}$, we can use formula:

$$\mathrm{P}(\mathbf{V}_\alpha = \mathbf{y}_\alpha | \mathbf{V}_\beta = \mathbf{x}_\beta) = \frac{\mathrm{P}(\mathbf{V}_\alpha = \mathbf{y}_\alpha, \mathbf{V}_\beta = \mathbf{x}_\beta)}{\mathrm{P}(\mathbf{V}_\beta = \mathbf{x}_\beta)},$$

where we know how to calculate marginal probabilities $\mathrm{P}(\mathbf{V}_\alpha = \mathbf{y}_\alpha, \mathbf{V}_\beta = \mathbf{x}_\beta)$ and $\mathrm{P}(\mathbf{V}_\beta = \mathbf{x}_\beta)$ using the message-passing algorithm.

Message-passing Algorithm for Completion

If we are computing completion with one variable, it is easy to use the algorithm for conditioning on one variable to obtain the result.

However, for completion in general, we apply a new message passing algorithm.

To compute

$$y^*_{k+1}, ..., y^*_n = \operatorname*{arg\,max}_{y_{k+1},\ldots,y_n} \mathrm{P}(V_{k+1}=y_{k+1}, ..., V_n=y_n | V_1=x_1, ..., V_k=x_k)$$

we can use the same message passing algorithm as the algorithm for calculating marginal probability $\mathrm{P}(V_1 = x_1, \ldots, V_k = x_k)$, except:

- Function to variable messages $M_{f\to V}(x)$ are computed by

$$M_{f\to V}(x) \quad = \quad \max_{x_1,\ldots,x_p} f(x, x_1, ..., x_p) M_{V_1\to f}(x_1) \cdots M_{V_p\to f}(x_p)$$

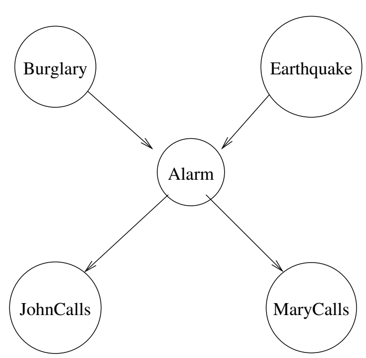
  over all other variables $V_1, ..., V_p$ (besides V) adjacent to f.
- Once all of the messages have been passed, then the maximum probability completion for any free variable V_{k+j} can be calculated by

$$y^*_{k+j} \quad = \quad \operatorname*{arg\,max}_{y_{k+j}} \quad M_{f_1\to V_{k+j}}(y_{k+j}) \cdots M_{f_p\to V_{k+j}}(y_{k+j})$$

  over all $f_1, ..., f_p$ containing V_{k+j}.
- If there are two or more values for a variable for which the maximal conditional probability is reached, we need to make sure that all variables are assigned consistently by hard-wiring the chosen variable value.

16.4 Message-Passing Inference Algorithm: Burglar-Earthquake Example

In this example we use the previously given Burglar-Earthquake Bayesian Network:



The given tables are:

B	P(B)
T	0.001
F	0.999

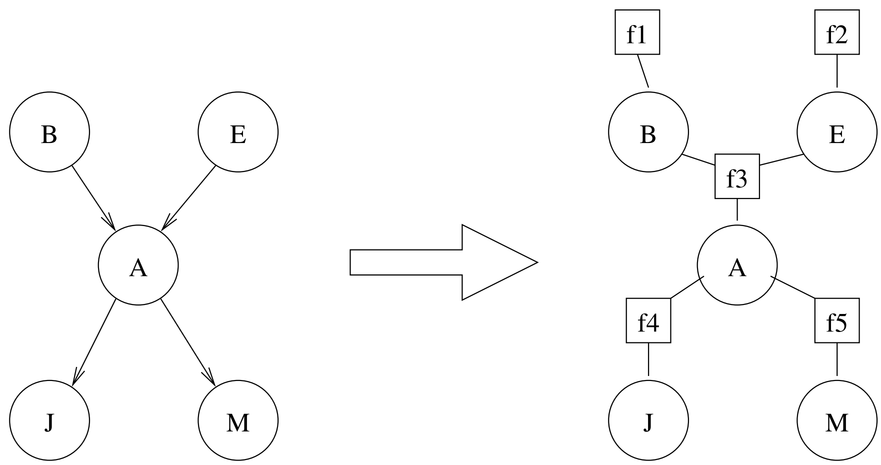
E	P(E)
T	0.002
F	0.998

B	E	A	P($A\|B,E$)
T	T	T	0.95
T	T	F	0.05
T	F	T	0.94
T	F	F	0.06
F	T	T	0.29
F	T	F	0.71
F	F	T	0.001
F	F	F	0.999

A	J	P($J\|A$)
T	T	0.90
T	F	0.10
F	T	0.05
F	F	0.95

A	M	P($M\|A$)
T	T	0.70
T	F	0.30
F	T	0.01
F	F	0.99

Our first step is to translate this network into a factor graph:



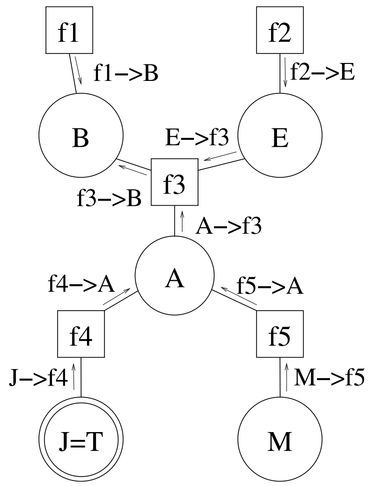
The function nodes correspond to conditional probabilities in the following way: $f_1 \sim \mathrm{P}(B)$, $f_2 \sim \mathrm{P}(E)$, $f_3 \sim \mathrm{P}(A|B,E)$, $f_4 \sim \mathrm{P}(J|A)$, and $f_5 \sim \mathrm{P}(M|A)$.

Burglar-Earthquake Example Problem

- John called, probability that Burglar is in the house
- $P(B=T|J=T)=?$
- Conditioning with one variable

Problem: Calculate the probability that a burglar is in the house, if we know that John has called.

We "hard-wire" the variable J to the value T, and analyze which messages we need to compute:



The messages are calculated in the following way:

$f_1 \to B$ is simple table copy of $\mathrm{P}(B)$:

B	$f_1 \to B$
T	0.001
F	0.999

Similarly:

E	$f_2 \to E$
T	0.002
F	0.998

E	$E \to f_3$
T	0.002
F	0.998

J is "hardwired" to T (observed evidence) so we get:

J	$J \to f_4$
T	1
F	0

M is not "hardwired":

M	$M \to f_5$
T	1
F	1

Calculation of the remaining messages requires a bit more calculations:

$f_4 \to A$

A	J	$J \to f_4$	f_4	
$A = T$	T	1	$\cdot 0.90$	$= 0.9$
	F	0	$\cdot 0.10$	$= 0$
			Σ	$= 0.9$
$A = F$	T	1	$\cdot 0.05$	$= 0.05$
	F	0	$\cdot 0.95$	$= 0$
			Σ	$= 0.05$

$f_5 \to A$

A	M	$M \to f_5$	f_5	
$A = T$	T	1	$\cdot 0.70$	$= 0.7$
	F	1	$\cdot 0.30$	$= 0.3$
			Σ	$= 1$
$A = F$	T	1	$\cdot 0.01$	$= 0.01$
	F	1	$\cdot 0.99$	$= 0.99$
			Σ	$= 1$

Hence the messages are:

A	$f_4 \to A$
T	0.9
F	0.05

and

A	$f_5 \to A$
T	1
F	1

. The message $A \to f_3$ is obtained by component-wise multiplication of messages coming into A:

A	$A \to f_3$
T	0.9
F	0.05

Finally, we compute the message $f_3 \to B$:

$f_3 \to B$						
B	E	A	$E \to f_3$	$A \to f_3$	f_3	
$B=T$	T	T	0.002	·0.9	·0.95	= 0.00171
	T	F	0.002	·0.05	·0.05	= 0.000005
	F	T	0.998	·0.9	·0.94	= 0.844308
	F	F	0.998	·0.05	·0.06	= 0.002994
					Σ	= 0.849017

$f_3 \to B$						
B	E	A	$E \to f_3$	$A \to f_3$	f_3	
$B=F$	T	T	0.002	·0.9	·0.29	= 0.000522
	T	F	0.002	·0.05	·0.71	= 0.000071
	F	T	0.998	·0.9	·0.001	= 0.0008982
	F	F	0.998	·0.05	·0.999	= 0.0498501
					Σ	= 0.0513413

Hence, the message $f_3 \to B$ is:

B	$f_3 \to B$
T	0.849017
F	0.0513413

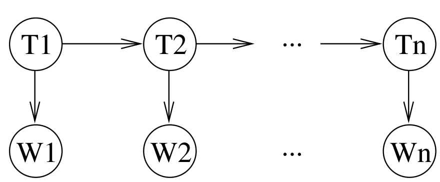
Final Calculation $P(B=T|J=T)$

Now, we can compute P$(B = T|J = T)$ by multiplying component-wise the messages arriving at B, and by normalizing the result:

$$
\begin{aligned}
P(B=T|J=T) &= \frac{f_1 \to B(T) \cdot f_3 \to B(T)}{f_1 \to B(T) \cdot f_3 \to B(T) + f_1 \to B(F) \cdot f_3 \to B(F)} \\
&= \frac{0.001 \cdot 0.849017}{0.001 \cdot 0.849017 + 0.999 \cdot 0.513413} = 0.01628373
\end{aligned}
$$

16.5 Message Passing Algorithm: POS Tagging Example

The HMM tagging using message passing would work as follows:



Training data:

```
swat V flies N like P ants N
time N flies V like P an D arrow N
```

Trained HMM Model:

T_1	$\mathrm{P}(T_1)$
N	0.5
V	0.5

,

T_{i-1}	T_i	$\mathrm{P}(T_i \mid T_{i-1})$
D	N	1
N	P	0.5
N	V	0.5
P	D	0.5
P	N	0.5
V	N	0.5
V	P	0.5

and

T_i	W_i	$\mathrm{P}(W_i \mid T_i)$
D	an	$2/3 \approx 0.666666667$
D	$\star$	$1/3 \approx 0.333333333$
N	ants	$2/9 \approx 0.222222222$
N	arrow	$2/9 \approx 0.222222222$
N	flies	$2/9 \approx 0.222222222$
N	time	$2/9 \approx 0.222222222$
N	$\star$	$1/9 \approx 0.111111111$
P	like	0.8
P	$\star$	0.2
V	flies	0.4
V	swat	0.4
V	$\star$	0.2

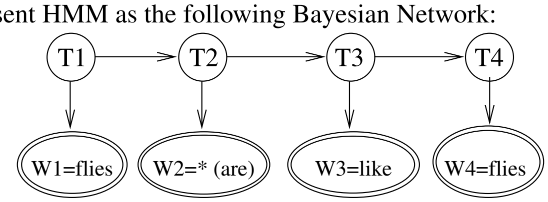
.

Tagging Example

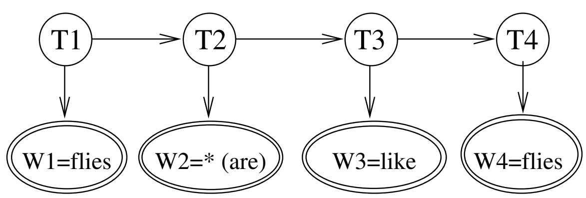
Slide notes:

Tagging Example

- Example: “flies are like flies”
- Represent HMM as the following Bayesian Network:



Let us again use the example sentence “flies are like flies”, which we used in a previous example with HMM. First, we will represent HMM configuration as a Bayesian Network with observable variables “hard-wired” to their values, as follows:

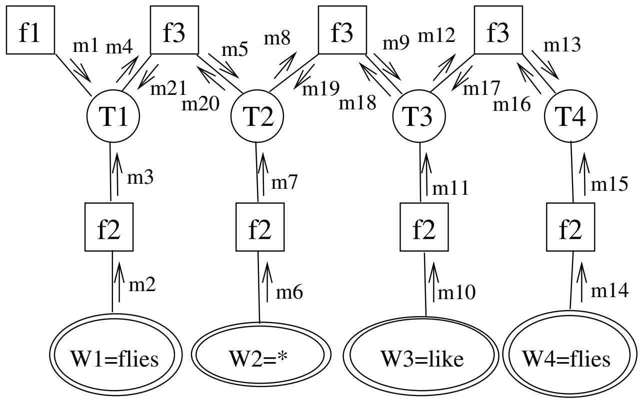


Slide notes:

POS Tagging as Message Passing

- Solving a completion problem
- Algorithm steps:
  - Create a factor graph
  - Hard-wire output variables
  - Use message passing with maximization
  - Find maximum-likely completion
- We will calculate only necessary messages

The corresponding factor graph is:



The messages are calculated as follows:

T_1	m_1
D	0
N	0.5
P	0
V	0.5

, and

W_1	m_2
flies	1
an	0
*	0
⋮	0

.

Calculation of m_3 is done as follows:

m_3			
$T_1 = D$	$W_1 =$	flies: $1 \cdot 0$	$= 0$
	$W_1 =$	an: $0 \cdot \frac{2}{3}$	$= 0$
	$W_1 =$	⋮ ⋮	$= 0$
			max:0
$T_1 = N$	$W_1 =$	flies $: 1 \cdot \frac{2}{9}$	$= \frac{2}{9}$
	$W_1 =$	an $: 0 \cdot \frac{1}{9}$	$= 0$
			max:2/9
⋮			

and we obtain

T_1	m_3
D	0
N	$2/9$
P	0
V	0.4

.

The other messages are:

T_1	$m_4 (= m_1 \cdot m_3)$
D	$0 \cdot 0 = 0$
N	$0.5 \cdot 2/9 = 1/9$
P	$0 \cdot 0 = 0$
V	$0.5 \cdot 0.4 = 0.2$

T_2	m_5
D	0
N	0.1
P	0.1
V	$1/18$

m_5 is calculated as follows:

m_5	$m_4 \cdot f_3$	
$T_2 = D$	$T_1 = D : 0 \cdot 0$	$= 0$
	$T_1 = N : \frac{1}{9} \cdot 0$	$= 0$
	$T_1 = P : 0 \cdot 0.5$	$= 0$
	$T_1 = V : 0.2 \cdot 0$	$= 0$
		max:0

m_5	$m_4 \cdot f_3$	
$T_2 = N$	$T_1 = D : 0 \cdot 1$	$= 0$
	$T_1 = N : \frac{1}{9} \cdot 0$	$= 0$
	$T_1 = P : 0 \cdot 0.5$	$= 0$
	$T_1 = V : 0.2 \cdot 0.5$	$= 0.1$
		max:0.1

m_5	$m_4 \cdot f_3$	
$T_2 = P$	$T_1 = D : 0 \cdot 0$	$= 0$
	$T_1 = N : \frac{1}{9} \cdot 0.5$	$= 1/18$
	$T_1 = P : 0 \cdot 0$	$= 0$
	$T_1 = V : 0.2 \cdot 0.5$	$= 0.1$
		max:0.1

m_5	$m_4 \cdot f_3$	
$T_2 = V$	$T_1 = D : 0 \cdot 0$	$= 0$
	$T_1 = N : \frac{1}{9} \cdot 0.5$	$= 1/18$
	$T_1 = P : 0 \cdot 0$	$= 0$
	$T_1 = V : 0.2 \cdot 0$	$= 0$
		max:$1/18$

We continue calculating:

W_2	m_6
flies	0
an	0
*	1
⋮	0

,

T_2	m_7
D	$1/3$
N	$1/9$
P	0.2
V	0.2

,

T_2	$m_8 (= m_5 \cdot m_7)$
D	$0 \cdot \frac{1}{3} = 0$
N	$0.1 \cdot \frac{1}{9} = 1/90$
P	$0.1 \cdot 0.2 = 0.02$
V	$\frac{1}{18} \cdot 0.2 = 1/90$

.

To calculate m_9, we have the following intermediate calculations:

m_9	$m_8 \cdot f_3$	
$T_3 = D$	$T_2 = D : 0 \cdot 0$	$= 0$
	$T_2 = N : \frac{1}{90} \cdot 0$	$= 0$
	$T_2 = P : \frac{1}{50} \cdot 0.5$	$= 0.01$
	$T_2 = V : \frac{1}{90} \cdot 0$	$= 0$
		max:0.01

m_9	$m_8 \cdot f_3$	
$T_3 = N$	$T_2 = D : 0 \cdot 1$	$= 0$
	$T_2 = N : \frac{1}{90} \cdot 0$	$= 0$
	$T_2 = P : \frac{1}{50} \cdot 0.5$	$= 0.01$
	$T_2 = V : \frac{1}{90} \cdot 0.5$	$= 1/180$
		max:0.01

m_9	$m_8 \cdot f_3$	
$T_3 = P$	$T_2 = D : 0 \cdot 0$	$= 0$
	$T_2 = N : \frac{1}{90} \cdot 0.5$	$= 1/180$
	$T_2 = P : \frac{1}{50} \cdot 0$	$= 0$
	$T_2 = V : \frac{1}{90} \cdot 0.5$	$= 1/180$
		max:$1/180$

m_9	$m_8 \cdot f_3$	
$T_3 = V$	$T_2 = D : 0 \cdot 0$	$= 0$
	$T_2 = N : \frac{1}{90} \cdot 0.5$	$= 1/180$
	$T_2 = P : \frac{1}{50} \cdot 0$	$= 0$
	$T_2 = V : \frac{1}{90} \cdot 0$	$= 0$
		max:$1/180$

and we obtain:

T_3	m_9
D	0.01
N	0.01
P	$1/180$
V	$1/180$

. Then,

W_3	m_{10}
like	1
$\vdots$	0

,

T_3	m_{11}
D	0
N	0
P	0.8
V	0

,

T_3	$m_{12} (= m_9 \cdot m_{11})$
D	$0.01 \cdot 0 = 0$
N	$0.01 \cdot 0 = 0$
P	$\frac{1}{180} \cdot 0.8 = 1/225$
V	$\frac{1}{180} \cdot 0 = 0$

.

To calculate m_{13}, we have the following intermediate calculations:

m_{13}	$m_{12} \cdot f_3$	
$T_4 = D$	$T_3 = D : 0 \cdot 0$	$= 0$
	$T_3 = N : 0 \cdot 0$	$= 0$
	$T_3 = P : \frac{1}{225} \cdot 0.5$	$= 1/450$
	$T_3 = V : 0 \cdot 0$	$= 0$
		max:$1/450$

m_{13}	$m_{12} \cdot f_3$	
$T_4 = N$	$T_3 = D : 0 \cdot 1$	$= 0$
	$T_3 = N : 0 \cdot 0$	$= 0$
	$T_3 = P : \frac{1}{225} \cdot 0.5$	$= 1/450$
	$T_3 = V : 0 \cdot 0.5$	$= 0$
		max:$1/450$

m_{13}	$m_{12} \cdot f_3$	
$T_4 = P$	$T_3 = D : 0 \cdot 0$	$= 0$
	$T_3 = N : 0 \cdot 0.5$	$= 0$
	$T_3 = P : \frac{1}{225} \cdot 0$	$= 0$
	$T_3 = V : 0 \cdot 0.5$	$= 0$
		max:0

m_{13}	$m_{12} \cdot f_3$	
$T_4 = V$	$T_3 = D : 0 \cdot 0$	$= 0$
	$T_3 = N : 0 \cdot 0.5$	$= 0$
	$T_3 = P : \frac{1}{225} \cdot 0$	$= 0$
	$T_3 = V : 0 \cdot 0$	$= 0$
		max:0

and we obtain:

T_4	m_{13}
D	$1/450$
N	$1/450$
P	0
V	0

. Then,

W_4	m_{14}
flies	1
$\vdots$	0

, and

T_4	m_{15}
D	0
N	$2/9$
P	0
V	0.4

.

To maximize the product of probabilities of T_4 we calculate:

T_4	$m_{13} \cdot m_{15}$	
D	$\frac{1}{450}\cdot$	$= 0$
N	$\frac{1}{450} \cdot \frac{2}{9}$	$= 1/2025$
P	$0 \cdot 0$	$= 0$
V	$0 \cdot 0.4$	$= 0$

and we obtain $\boxed{T_4^* = N}$, which we use in further messages, as a “hard-wired” value. We calculate

T_4	m_{16}
D	0
N	$2/9$
P	0
V	0

, and for m_{17} use only $T_4 = N$ in $m_{16} \cdot f_3$:

$m_{16} \cdot f_3$	
$\frac{2}{9} \cdot 1$	$= 2/9$
$\frac{2}{9} \cdot 0$	$= 0$
$\frac{2}{9} \cdot 0.5$	$= 1/9$
$\frac{2}{9} \cdot 0.5$	$= 1/9$

, and we obtain:

T_3	m_{17}
D	$2/9$
N	0
P	$1/9$
V	$1/9$

.

To find optimal T_3 we calculate:

T_3	$m_9 \cdot m_{11} \cdot m_{17}$	
D	$0.01 \cdot 0 \cdot \frac{2}{9}$	$= 0$
N	$0.01 \cdot 0 \cdot 0$	$= 0$
P	$\frac{1}{180} \cdot 0.8 \cdot \frac{1}{9}$	$= 1/2025$
V	$\frac{1}{180} \cdot 0 \cdot \frac{1}{9}$	$= 0$

and we obtain: $\boxed{T_3^* = P}$

Then,

T_3	$m_{18} = m_{17} \cdot m_{11}$
D	0
N	0
P	$\frac{1}{9} \cdot 0.8 = 4/45$
V	0

,

T_2	$m_{19} = m_{18} \cdot f_3$ for $T_3 = P$
D	$\frac{4}{45} \cdot 0 = 0$
N	$\frac{4}{45} \cdot \frac{1}{2} = 2/45$
P	$\frac{4}{45} \cdot 0 = 0$
V	$\frac{4}{45} \cdot \frac{1}{2} = 2/45$

.

To find optimal T_2 we calculate:

T_2	$m_{19} \cdot m_5 \cdot m_7$	
D	$0 \cdot 0 \cdot \frac{1}{3}$	$= 0$
N	$\frac{2}{45} \cdot 0.1 \cdot \frac{1}{9}$	$= 1/2025$
P	$0 \cdot 0.1 \cdot 0.2$	$= 0$
V	$\frac{2}{45} \cdot \frac{1}{18} \cdot 0.2$	$= 1/2025$

and we can choose either N or V. Let us choose $\boxed{T_2^* = V}$.

T_2	$m_{20} = m_7 \cdot m_{19}$
D	0
N	0
P	0
V	$0.2 \cdot \frac{2}{45} = 2/225$

,

T_1	$m_{21} = m_{20} \cdot f_3$ for $T_2 = V$
D	$\frac{2}{225} \cdot 0 = 0$
N	$\frac{2}{225} \cdot \frac{1}{2} = 1/225$
P	$\frac{2}{225} \cdot 0 = 0$
V	$\frac{2}{225} \cdot 0 = 0$

.

To find optimal T_1 we calculate:

T_1	$m_1 \cdot m_3 \cdot m_{21}$	
D	$0 \cdot 0 \cdot 0$	$= 0$
N	$0.5 \cdot \frac{2}{9} \cdot \frac{1}{225}$	$= 1/2025$
P	$0 \cdot 0 \cdot 0$	$= 0$
V	$0.5 \cdot 0.4 \cdot 0$	$= 0$

and we obtain $\boxed{T_1^* = N}$.

To summarize, the most probable values of unknown variables T_1, T_2, T_3, and T_4 are:

$$\boxed{T_1^* = N \quad T_2^* = V \quad T_3^* = P \quad T_4^* = N}$$