Natural Language Processing CSCI 4152/6509 — Lecture 15 P0 Discussion (3): Sum-Product Algorithm

Instructors: Vlado Keselj
Time and date: 14:35 – 15:55, 20-Nov-2025
Location: Studley LSC-Psychology P5260

Previous Lecture

- P0 discussion (2): P-01, P-04, P-06, P-13, P-14, P-15, P-16, P-17
- **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)

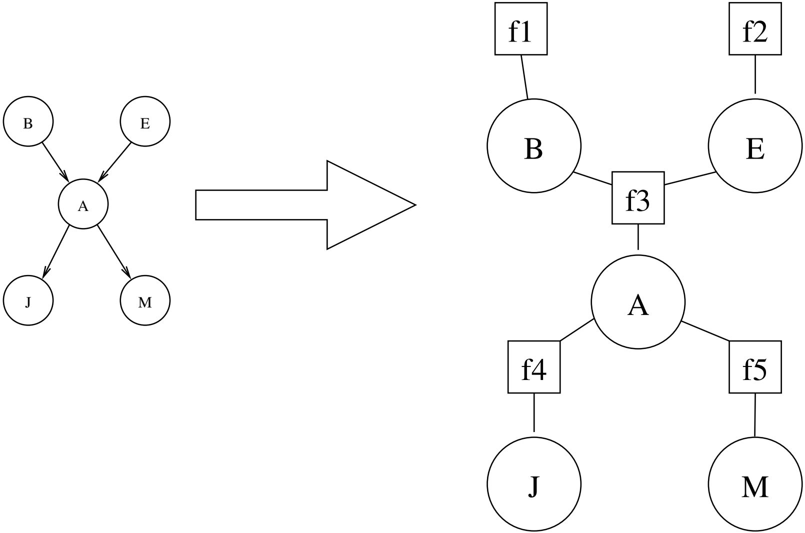
P0 Topics Discussion (3)

- Continued discussion of individual projects as proposed in P0 submissions (part 3)
- Projects discussed: P-05

Principles of Message Passing

- A message summarizes computation in the corresponding part of graph
- Messages are vectors of real numbers
- Each node passes to each neighbour node a message exactly once
- To pass a message to a neighbour node, a node needs to receive messages from all other neighbour nodes
- Important property: a tree-structured Bayesian Network leads to a tree factor graph

Message Passing Ex.: Order of Computation



Computation Problems Solved by Message Passing

- Applicable to all inference problems
- Two main types of computation:
  - **Summation** of resulting overall products where variables take different domain values
  - **Maximization:** Finding variable values for which the resulting overall product is maximized
- Two main situations:
  - Factor node passing a message to variable node
  - Variable node passing a message to factor node

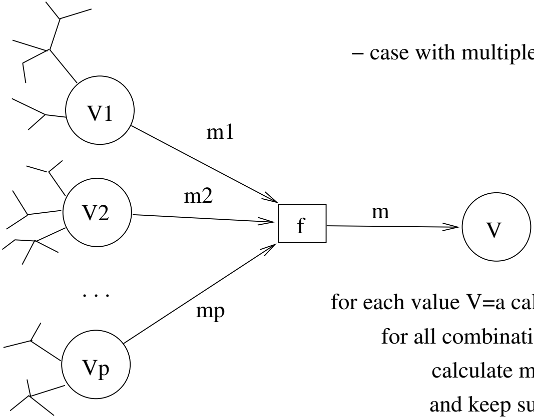
Four Cases of Message Computation

- 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):

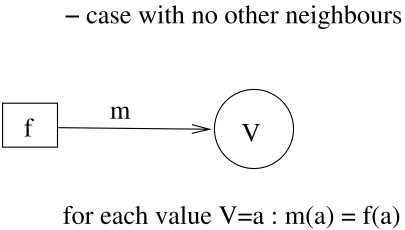
for all combinations of V1 .. Vp

calculate m1*m2*..mp*f

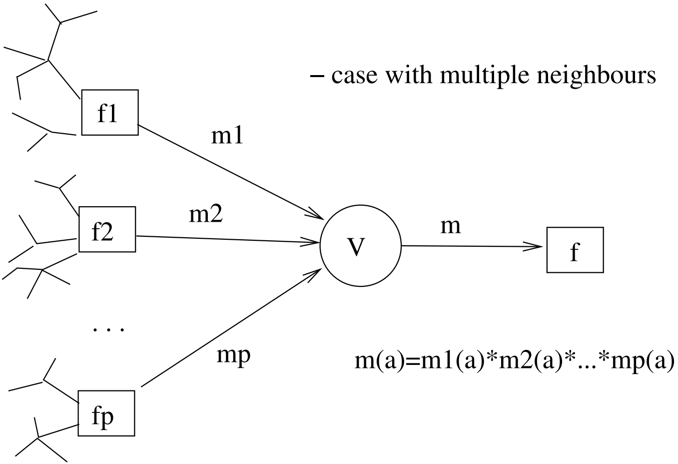
and keep sum or max

m(a) is resulting sum or max

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

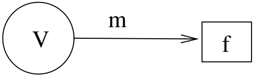


Variable Node with Multiple Neighbours Passing a Message to Factor Node



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

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

Marginalization with One Variable

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

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

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

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:

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

Conditioning with One Variable

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

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

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

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

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

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.

Completion

Completion with one variable: use conditioning on one variable; otherwise

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

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:

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

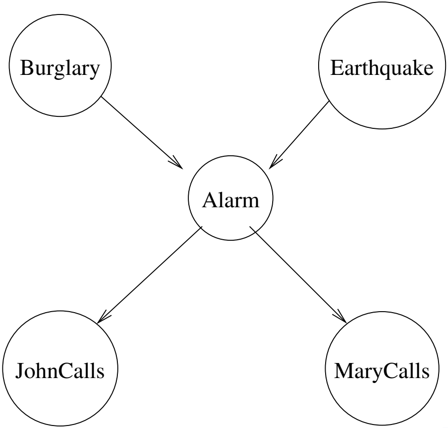
At the end

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

Variables must be assigned consistently (check by "hard-wiring")

Message Passing Algorithm: Burglar-Earthquake Example

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



CP Tables

B	$P(B)$
T	0.001
F	0.999

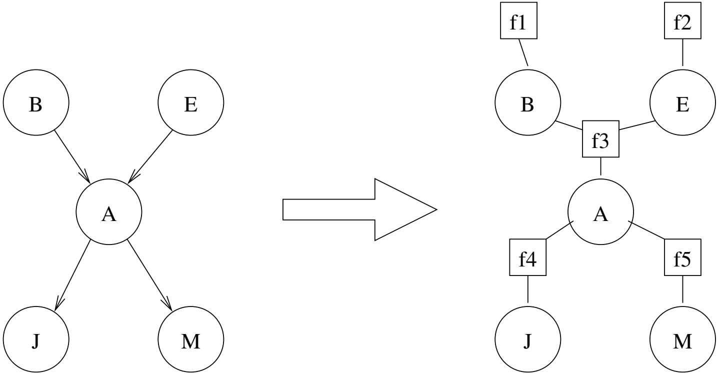
E	$P(E)$
T	0.002
F	0.998

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

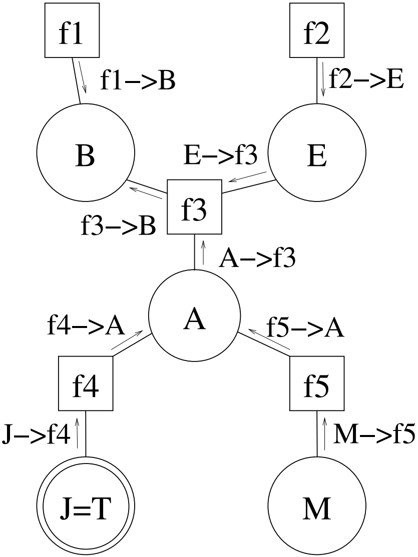
Factor Graph



Burglar-Earthquake Example Problem

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

Message Passing



Message Calculation

B	$f_1 \to B$
T	0.001
F	0.999

E	$f_2 \to E$
T	0.002
F	0.998

E	$E \to f_3$
T	0.002
F	0.998

J	$J \to f_4$
T	1
F	0

M	$M \to f_5$
T	1
F	1

$f_4 \to A$

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

A	$f_4 \to A$
T	0.9
F	0.05

and

A	$f_5 \to A$
T	1
F	1

. Then:

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	$\cdot 0.9$	$\cdot 0.95$	$= 0.00171$
	T	F	0.002	$\cdot 0.05$	$\cdot 0.05$	$= 0.000005$
	F	T	0.998	$\cdot 0.9$	$\cdot 0.94$	$= 0.844308$
	F	F	0.998	$\cdot 0.05$	$\cdot 0.06$	$= 0.002994$
					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	$\cdot 0.9$	$\cdot 0.29$	$= 0.000522$
	T	F	0.002	$\cdot 0.05$	$\cdot 0.71$	$= 0.000071$
	F	T	0.998	$\cdot 0.9$	$\cdot 0.001$	$= 0.0008982$
	F	F	0.998	$\cdot 0.05$	$\cdot 0.999$	$= 0.0498501$
					Σ	$= 0.0513413$

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

B	$f_3 \to B$
T	0.849017
F	0.0513413

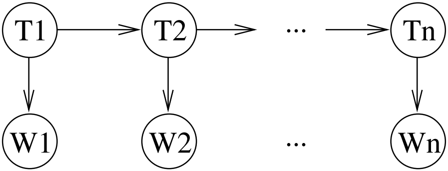
Final Calculation $P(B = T|J = T)$

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

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

Message Passing Algorithm: POS Tagging Example

- HMM Example, revisited



- HMM can be seen as a tree-structured Bayesian Network

Generated Tables

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

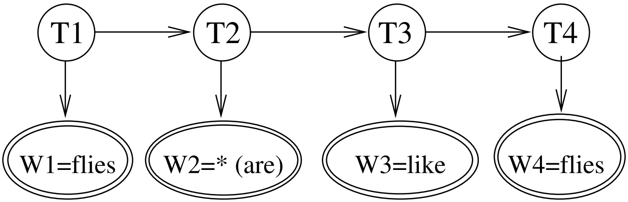
T_1	$\mathrm{P}(T_1)$
N	0.5
V	0.5

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

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

Tagging Example

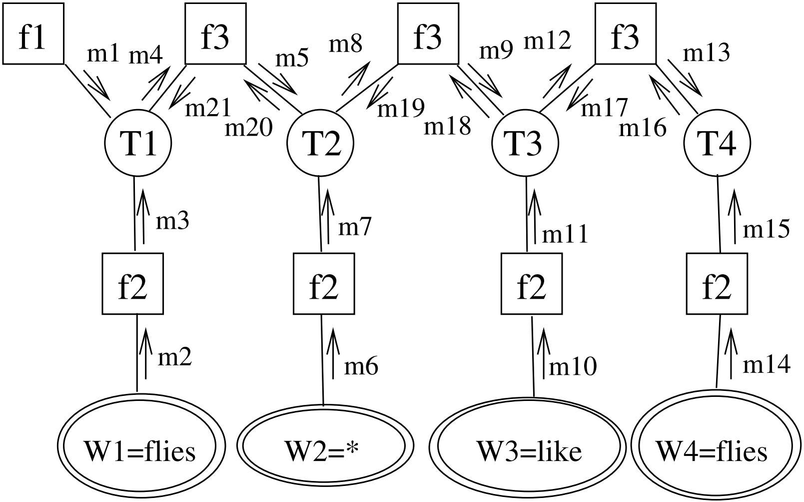
- Example: "flies are like flies"
- Represent HMM as the following Bayesian Network:



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

Factor Graph (with messages)



T_1	m_1
D	0
N	0.5
P	0
V	0.5

W_1	m_2
flies	1
an	0
*	0
⋮	0

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

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

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

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	$\frac{1}{18} \cdot 0.2 = 1/90$

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:$1/180$

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

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	$\frac{1}{180} \cdot 0 = 0$

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

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

W_4	m_{14}
flies	1
⋮	0

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

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	$0 \cdot 0.4$	$= 0$

$\boxed{T_4^* = N}$ "hard-wire" T_4

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

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

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

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	$\frac{1}{180} \cdot 0 \cdot \frac{1}{9}$	$= 0$

$\boxed{T_3^* = P}$

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	$\frac{4}{45} \cdot \frac{1}{2} = 2/45$

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	$\frac{2}{45} \cdot \frac{1}{18} \cdot 0.2$	$= 1/2025$

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:

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