CMPS 561
Fuzzy Set Retrieval

Ryan Benton

September 1, 2010

Agenda

- Fuzzy Sets
- Fuzzy Operations
- Fuzzy Set Retrieval
- Processing Fuzzy Queries
- Set Issues
- λ-level Sets

Fuzzy Sets

Issue

- Problem
  - Boolean assumes documents
    - Exist in 1 of 2 states
    - States are Non-Overlapping
      - CRISP!
  - Other crisp states may exist
    - Grades: A, B, C, D, F
      - Can be in one and not the other.
  - Often, exact boundaries are not known
    - Hot, Warm, Luke-warm, Cool, Chilly, Cold
      - When does Hot become Warm?

Issue

- Fuzzy Sets
  - Informal Definition
    - Fuzzy Set is a class with fuzzy boundaries.
    - Alternatively, fuzzy set is a class with non-crisp boundaries.

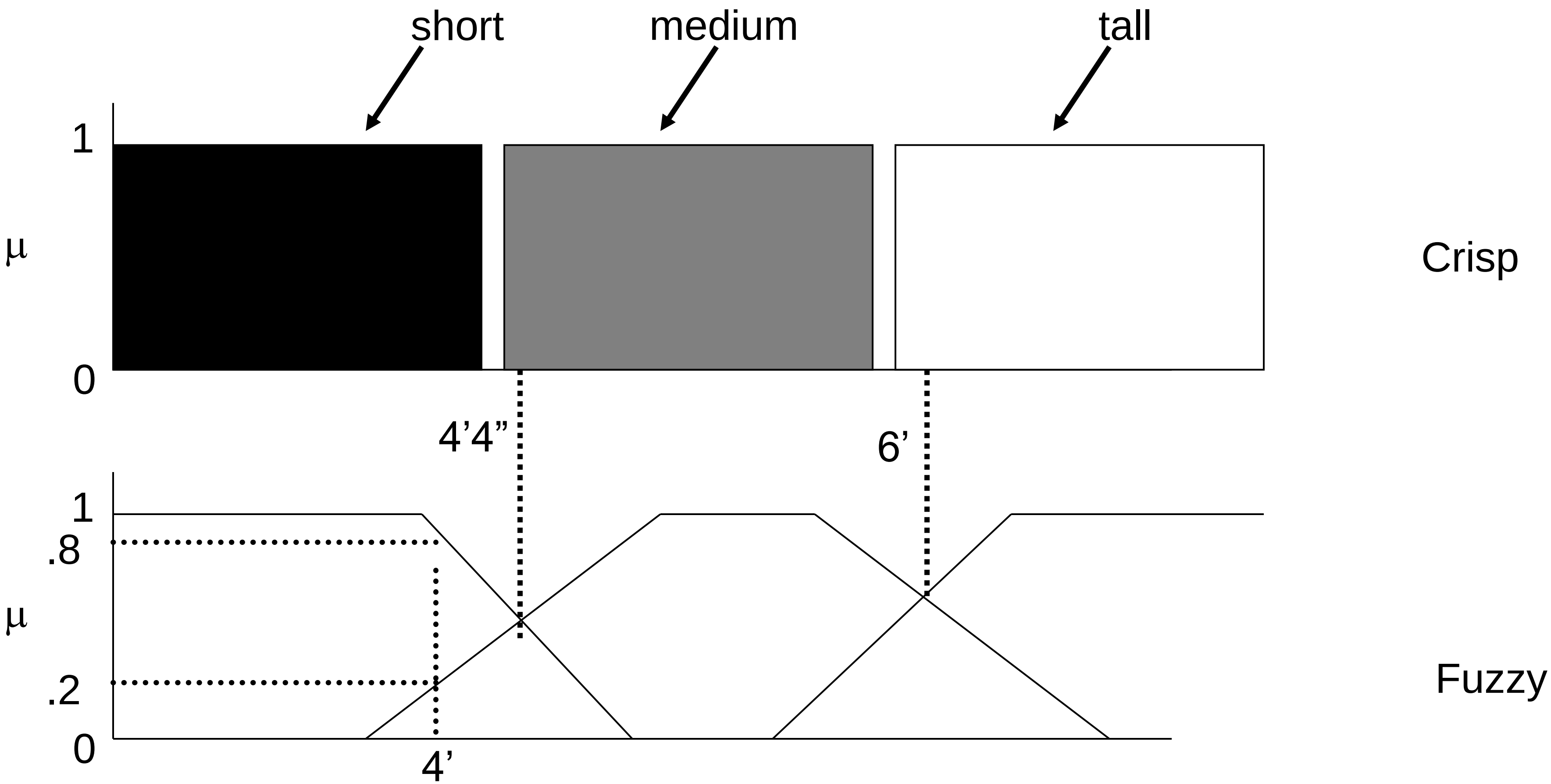
Fuzzy Sets – Slightly More Formal

- Regular (Crisp) set
  - Let $X = \{x_i\}$ be the universe of objects of interest.
- Fuzzy Subset of X → A.
  - A is a set of order pairs
    - $A = \{ < x_i, \mu_A(x) > \}$
  - $\mu_A(x_i)$ → Grade of membership of $x_i \in A$
  - $\mu_A : X \rightarrow [0,1]$

Example 1

- Assume Crisp Set is
  - $X = \{x_1, x_2, x_3, x_4\}$
- Fuzzy Subsets
  - $A = \{<x_1,1.0>,<x_2,0.5>< x_3,0.8>\}$
  - $B = \{<x_2,0.8>\}$

Example 2
Membership Functions



$\mu_{tall}(4') = 0$, $\mu_{medium}(4') = 0.2$, $\mu_{short}(4') = .8$

Fuzzy Operations

Operations 1

- Let
  - $A = \{ < x, \mu_A(x) > \}$
  - $B = \{ < x, \mu_B(x) > \}$

- Equality
  - A = B iff
    - $\mu_A(x) = \mu_B(x)$, for all $x \in X$.

Operations 2

- Containment
  - $A \subseteq B$ iff
    - $\mu_A(x) \le \mu_B(x)$, for all $x \in X$.

- Complement
  - $\bar{A} = \{ < x, \mu_{\bar{A}}(x) > \}$ iff
    - $\mu_{\bar{A}}(x) = 1 - \mu_A(x)$, for all $x \in X$.

Operations 3

- Union
  - $A \cup B = \{< x, \mu_{A \cup B}(x) > \}$ iff
    - $\mu_{A \cup B} = \max(\mu_A(x), \mu_B(x))$, for all $x \in X$.

- Intersection
  - $A \cap B = \{< x, \mu_{A \cap B}(x) > \}$ iff
    - $\mu_{A \cap B} = \min(\mu_A(x), \mu_B(x))$, for all $x \in X$.

Fuzzy Set Retrieval Model

Document Representation Boolean

- Relation
  - $\mathcal{D} = \{ < d_\alpha, t_i, \mu_\mathcal{D}(d_\alpha, t_i)> \}$
  - $\mu_\mathcal{D}$: D x T → {0,1}
  - $\mu_\mathcal{D}(d_\alpha, t_i)$
    - 1, if d_α contains t_i
    - 0, otherwise
- $D_t = \{d_\alpha \in D \mid \mu_\mathcal{D}(d_\alpha, t) = 1\}$
- $d \equiv D_d = \{t_i \in T \mid \mu_\mathcal{D}(d, t_i) = 1\}$

Document Representation Fuzzy

- $\mathcal{D} = \{ < d_\alpha, t_i, \mu_\mathcal{D}(d_\alpha, t_i)> \}$
- $\mu_\mathcal{D}$: D x T → [0,1]
- $\mu_\mathcal{D}(d_\alpha, t_i)$
  - The degree of membership of t_i to d_α
- $\mathcal{D}_t = \{<d_\alpha, \mu_\mathcal{D}(d_\alpha, t)> \mid d_\alpha \in D >\}$
- $\mathcal{D}_d = \{<t_i, \mu_\mathcal{D}(d, t_i)> \mid t_i \in T>\}$

Processing Boolean Query (Method 1)

- $D_{t1\wedge t2} = \{d_\alpha \in D \mid \mu_{\mathcal{D}}(d_\alpha, t_1) \wedge \mu_{\mathcal{D}}(d_\alpha, t_2) = 1\}$
- $D_{t1\vee t2} = \{d_\alpha \in D \mid \mu_{\mathcal{D}}(d_\alpha, t_1) \vee \mu_{\mathcal{D}}(d_\alpha, t_2) = 1\}$

- D_t = set of Documents containing term t
- $T = \{a,b,c,d,e\}$
- D_a, D_b, D_c, D_d, D_e,

Processing Fuzzy Query (Method 1)

- $\mathcal{D}_{t1 \wedge t2} = \{<d_\alpha , \min(\mu_{\mathcal{D}}(d_\alpha, t_1) , \mu_{\mathcal{D}}(d_\alpha, t_2)) > \mid d_\alpha \in D\}$
- $\mathcal{D}_{t1 \vee t2} = \{<d_\alpha , \max(\mu_{\mathcal{D}}(d_\alpha, t_1) , \mu_{\mathcal{D}}(d_\alpha, t_2)) > \mid d_\alpha \in D\}$

- $\mathcal{D}_t$ = set of <Document, membership value> pairs containing term t
- $T = \{a, b, c, d, e\}$
- $\mathcal{D}_a, \mathcal{D}_b, \mathcal{D}_c, \mathcal{D}_d, \mathcal{D}_e$

Retrieval Function
Boolean

- $RSV \equiv F$
- $RSV_t(d_\alpha) = \mu_{\mathcal{D}}(d_\alpha, t)$
- $RSV_{\neg e}(d_\alpha) = 1 - RSV_e(d_\alpha)$
- $RSV_{e1 \wedge e2}(d_\alpha) = RSV_{e1}(d_\alpha) \wedge RSV_{e2}(d_\alpha)$
- $RSV_{e1 \vee e2}(d_\alpha) = RSV_{e1}(d_\alpha) \vee RSV_{e2}(d_\alpha)$

Retrieval Function Fuzzy

- $RSV_t(d_\alpha) = \mu_{\mathcal{D}}(d_\alpha, t)$
- $RSV_{\neg e}(d_\alpha) = 1 - RSV_e(d_\alpha)$
- $RSV_{e1 \wedge e2}(d_\alpha) = \min(RSV_{e1}(d_\alpha), RSV_{e2}(d_\alpha))$
- $RSV_{e1 \vee e2}(d_\alpha) = \max(RSV_{e1}(d_\alpha), RSV_{e2}(d_\alpha))$

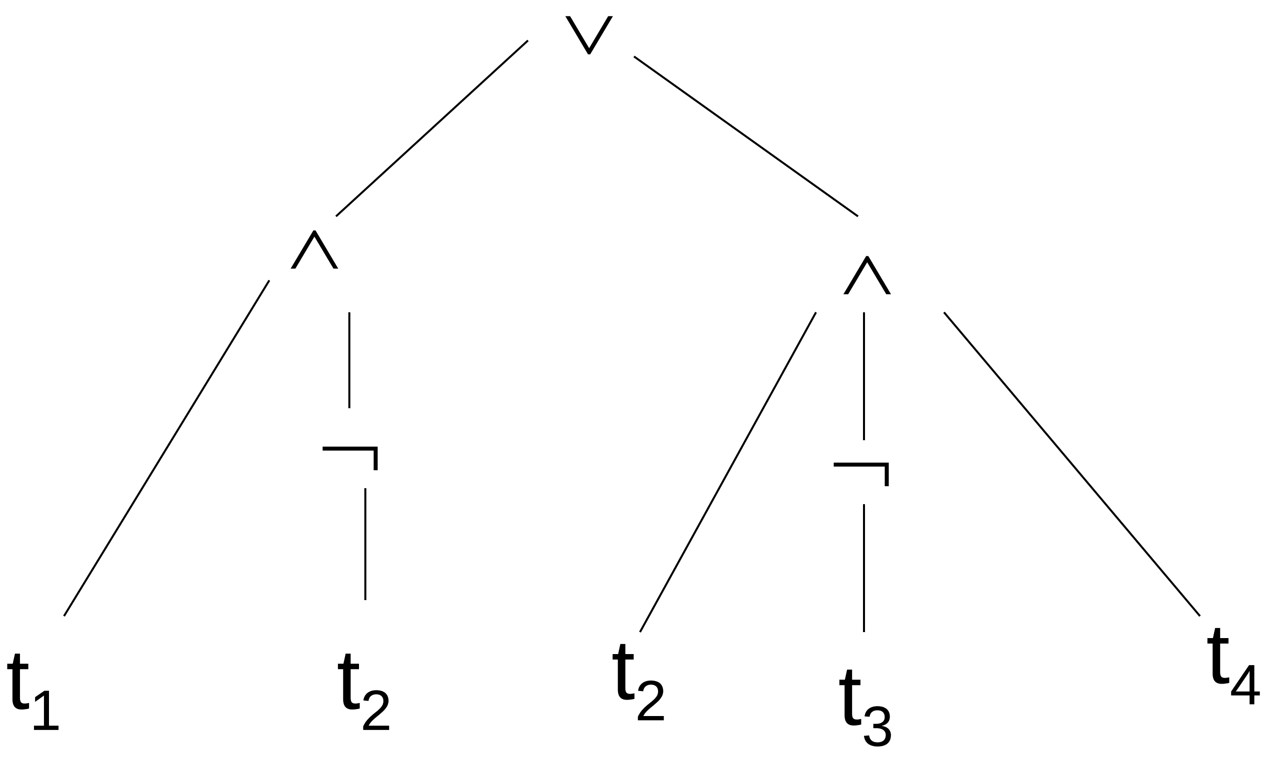
Processing Fuzzy Queries

Term-Document Incidence Matrix

	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8
t_1	0.1	0.7	0.6	0.4	0.8	0.6	0.3	0.6
t_2	0.3	0.8	0.9	0.8	0.6	0.5	0.7	0.2
t_3	0.8	1	0.2	0.6	0.8	0.2	0.8	0.5
t_4	0.4	0.6	0.7	0.5	0.7	0.3	0.1	0.9
t_5	0.7	0.2	0.8	0.4	0.6	0.2	0.8	0.4

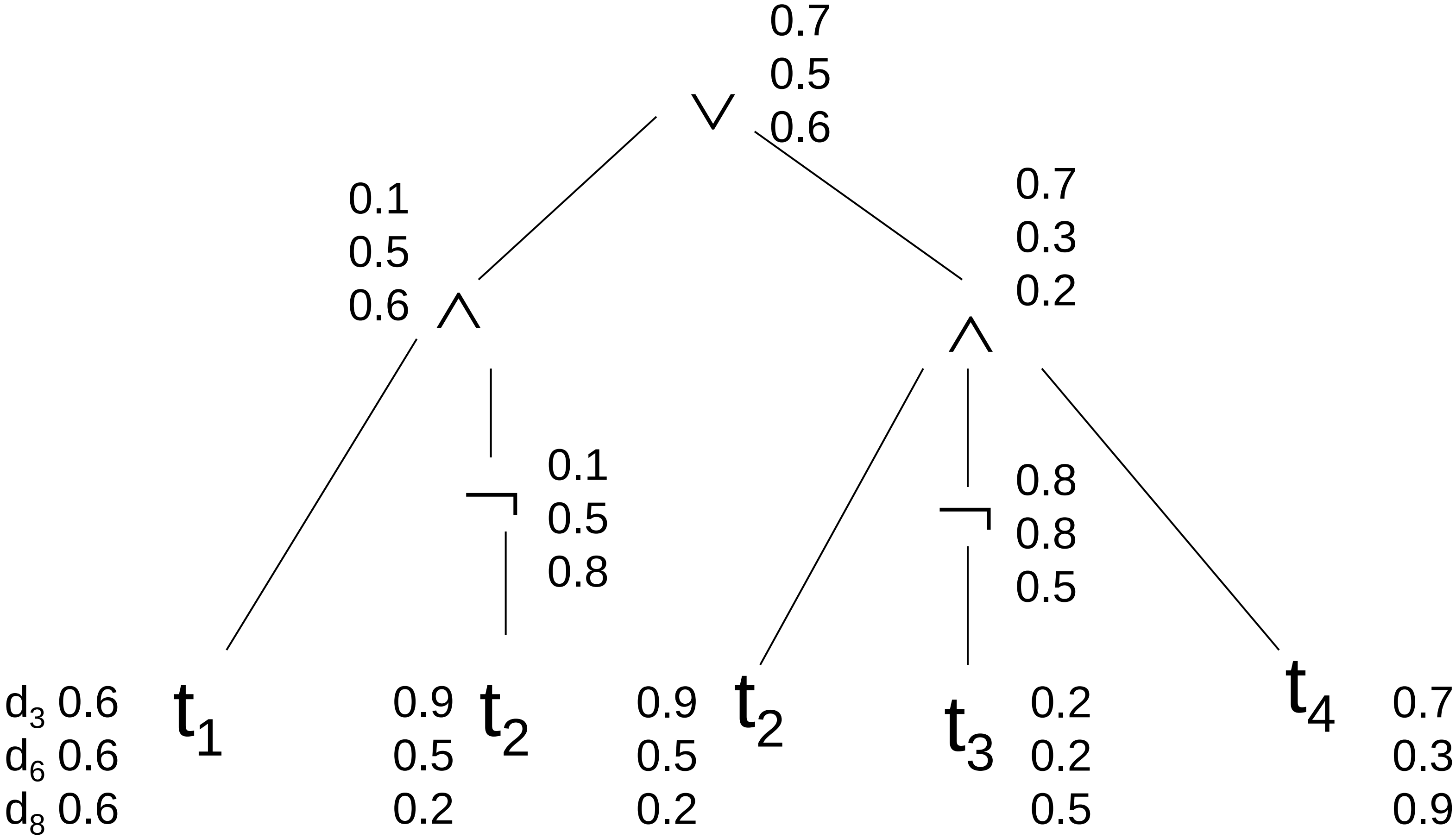
Fuzzy example

$$q = (t_1 \wedge \neg t_2) \vee (t_2 \wedge \neg t_3 \wedge t_4))$$



Fuzzy example
Max, Min

$q = (t_1 \wedge \neg t_2) \vee (t_2 \wedge \neg t_3 \wedge t_4))$



Inverted File Processing (Method 1)

- $D_t = \{<d_\alpha, \mu_{\mathcal{D}t}(d_\alpha)> \mid \mu_{\mathcal{D}t}(d_\alpha) > 0\}$

- $D_{\neg e} = \{<d_\alpha, 1\text{-}RSV_e(d_\alpha)> \mid 1\text{-}RSV_e(d_\alpha) > 0\}$

- $D_{e1 \wedge e2} = \{<d_\alpha, RSV_{e1 \wedge e2}(d_\alpha)> \mid RSV_{e1 \wedge e2}(d_\alpha) > 0\}$

- $D_{e1 \vee e2} = \{<d_\alpha, RSV_{e1 \vee e2}(d_\alpha)> \mid RSV_{e1 \vee e2}(d_\alpha) > 0\}$

Processing Fuzzy Query (Method 1)

Output	Query
$\mathcal{D}_{t}$	t
$\mathcal{D}_{e1}$	e1
$\mathcal{D}_{e2}$	e2
$\mathcal{D}_{e1} \cap \mathcal{D}_{e2}$	$e1 \wedge e2$
$\mathcal{D}_{e1} \cup \mathcal{D}_{e2}$	$e1 \vee e2$
$\mathcal{D} \setminus \mathcal{D}_{e1}$	$\neg e1$

Set Issues

Set Identities

- $A \cup B = B \cup A$
- $(A \cup B) \cup C = A \cup (B \cup C)$
- $A \cup (B \cap C) =$
  $(A \cup B) \cap (A \cup C)$
- $A \cup \varnothing = A$
- $A \cup A' = S$
- $A \cap B = B \cap A$
- $(A \cap B) \cap C = A \cap (B \cap C)$
- $A \cap (B \cup C) =$
  $(A \cap B) \cup (A \cap C)$
- $A \cap S = A$
- $A \cap A' = \varnothing$

Definitions of Union and Intersection

Name	Union	Intersection
max, min	$\cup$	$\cap$
max, product	$\cup$	$\cdot$
probabilistic sum, product	$\hat{+}$	$\cdot$
Einstein	$\overset{+}{\varepsilon}$	$\overset{\bullet}{\varepsilon}$
bold	$\overset{\bullet}{\cup}$	$\overset{\bullet}{\cap}$
Hamacher	$\overset{+}{\gamma}$	$\overset{\cdot}{\gamma}$
Yager	$\overset{\lambda}{v}$	$\overset{Y}{v}$
Schweizer-Sklar	$\perp_p$	$\top_p$

Operations –
Union & Intersection

- $a \cdot b = ab$
- $a \mathbin{\overset{\uparrow}{+}} b = a + b - ab$
- $a \overset{+}{\varepsilon} b = (a + b) / (1 + ab)$
- $a \overset{\bullet}{\varepsilon} b = (ab) / (1 + (1-a)(1-b)$
- $a \mathbin{\dot{\cup}} b = \min(a + b, 1)$
- $a \mathbin{\dot{\cap}} b = \max(0, a + b - 1)$

Operations – Union & Intersection

- $a \overset{+}{\gamma} b = (a + b - (1 - \gamma)ab) / (\gamma + (1-\gamma)(1- ab))$
  - $\gamma \in [0, +\infty)$
- $a \overset{\cdot}{\gamma} b = (ab) / (\gamma + (1-\gamma)(a+b))$
  - $\gamma \in [0, +\infty)$
- $a \underset{v}{\curlywedge} b = \min(1, (a^v + b^v)^{(1/v)})$
  - $v= [1, +\infty)$
- $a \underset{v}{\curlyvee} b = 1 - \min(1, [(1-a)^v + (1-b)^v]^{(1/v)})$
  - $v= [1, +\infty)$

Operations – Union & Intersection

$$a \perp_p b = \begin{cases} 1 - ((1-a)^{-p} + (1-b)^{-p} - 1)^{(-1/p)} & p > 0 \\ a \uparrow b & p = 0 \\ 1 - ((1-a)^{-p} + (1-b)^{-p} - 1)^{(-1/p)} & p < 0,\ (1-a)^{-p} + (1-b)^{-p} > 1 \\ 1 & p < 0,\ (1-a)^{-p} + (1-b)^{-p} \leq 1 \end{cases}$$

Operations – Union & Intersection

$$
a\ T_p\ b = \begin{cases} (a^{-p} + b^{-p} - 1)^{(-1/p)} & p > 0 \\ ab & p = 0 \\ (a^{-p} + b^{-p} - 1)^{(-1/p)} & p < 0,\ a^{-p} + b^{-p} > 1 \\ 0 & p < 0,\ a^{-p} + b^{-p} \leq 1 \end{cases}
$$

Example of Set Operations

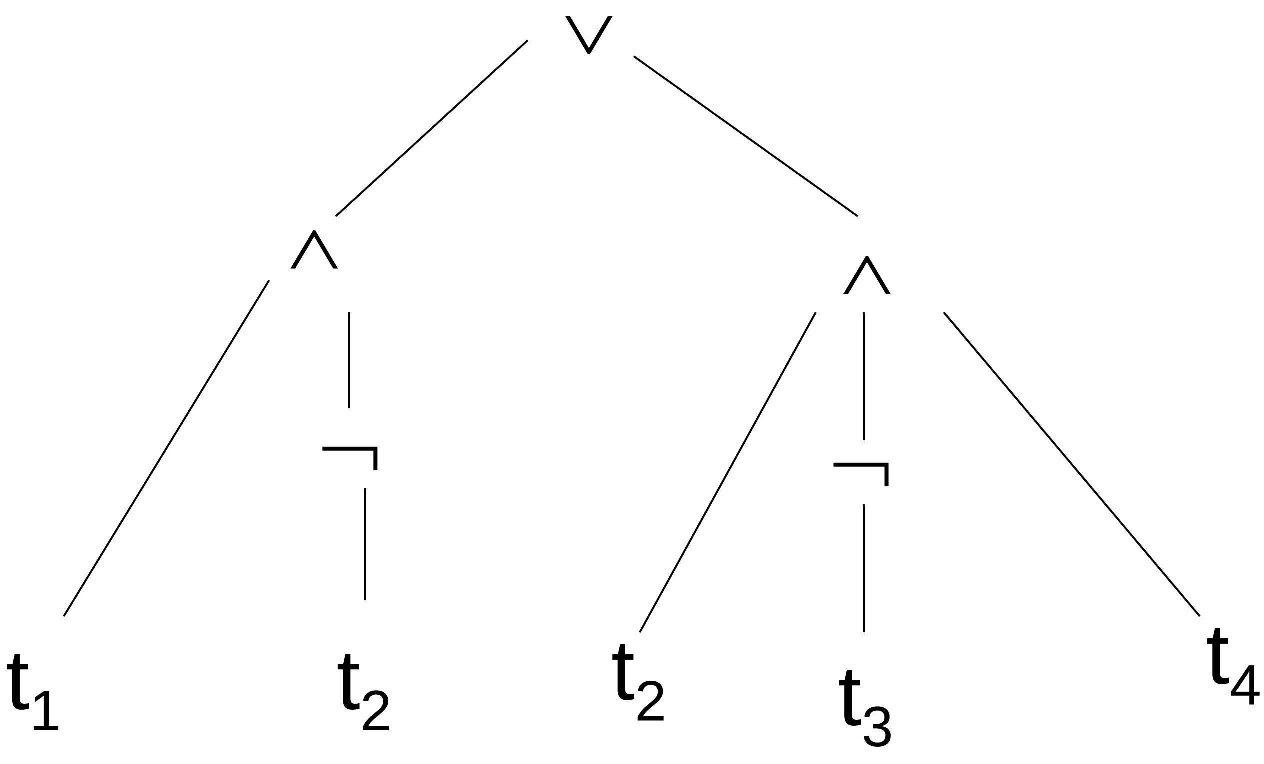
- Assume
  - Union = Probabilistic Sum
    - $\vee = a \hat{+} b = a + b - ab$
  - Intersection = Product
    - $\wedge = a \cdot b = ab$

Term-Document Incidence Matrix

	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8
t_1	0.1	0.7	0.6	0.4	0.8	0.6	0.3	0.6
t_2	0.3	0.8	0.9	0.8	0.6	0.5	0.7	0.2
t_3	0.8	1	0.2	0.6	0.8	0.2	0.8	0.5
t_4	0.4	0.6	0.7	0.5	0.7	0.3	0.1	0.9
t_5	0.7	0.2	0.8	0.4	0.6	0.2	0.8	0.4

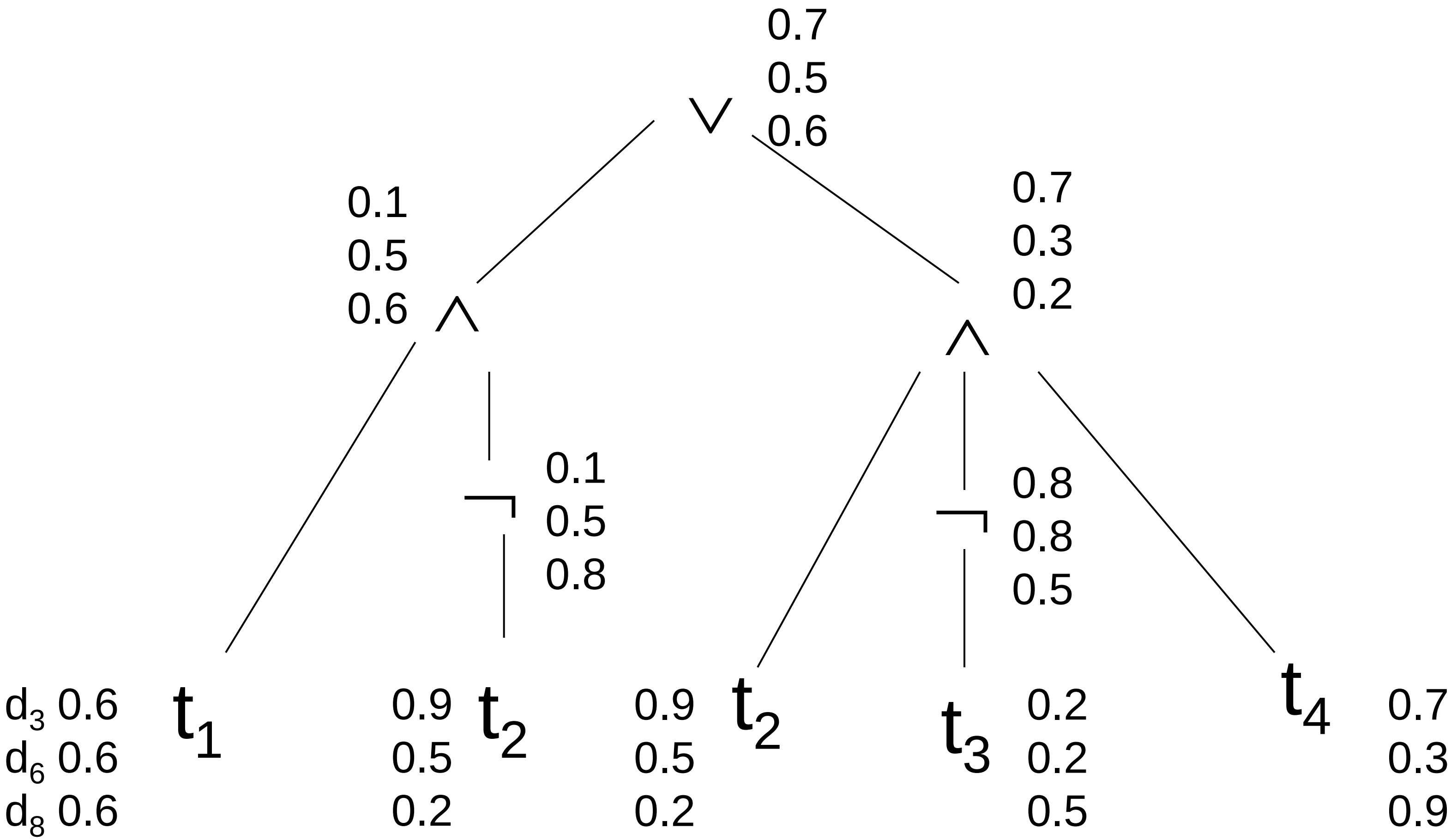
Fuzzy example

$$q = (t_1 \wedge \neg t_2) \vee (t_2 \wedge \neg t_3 \wedge t_4))$$



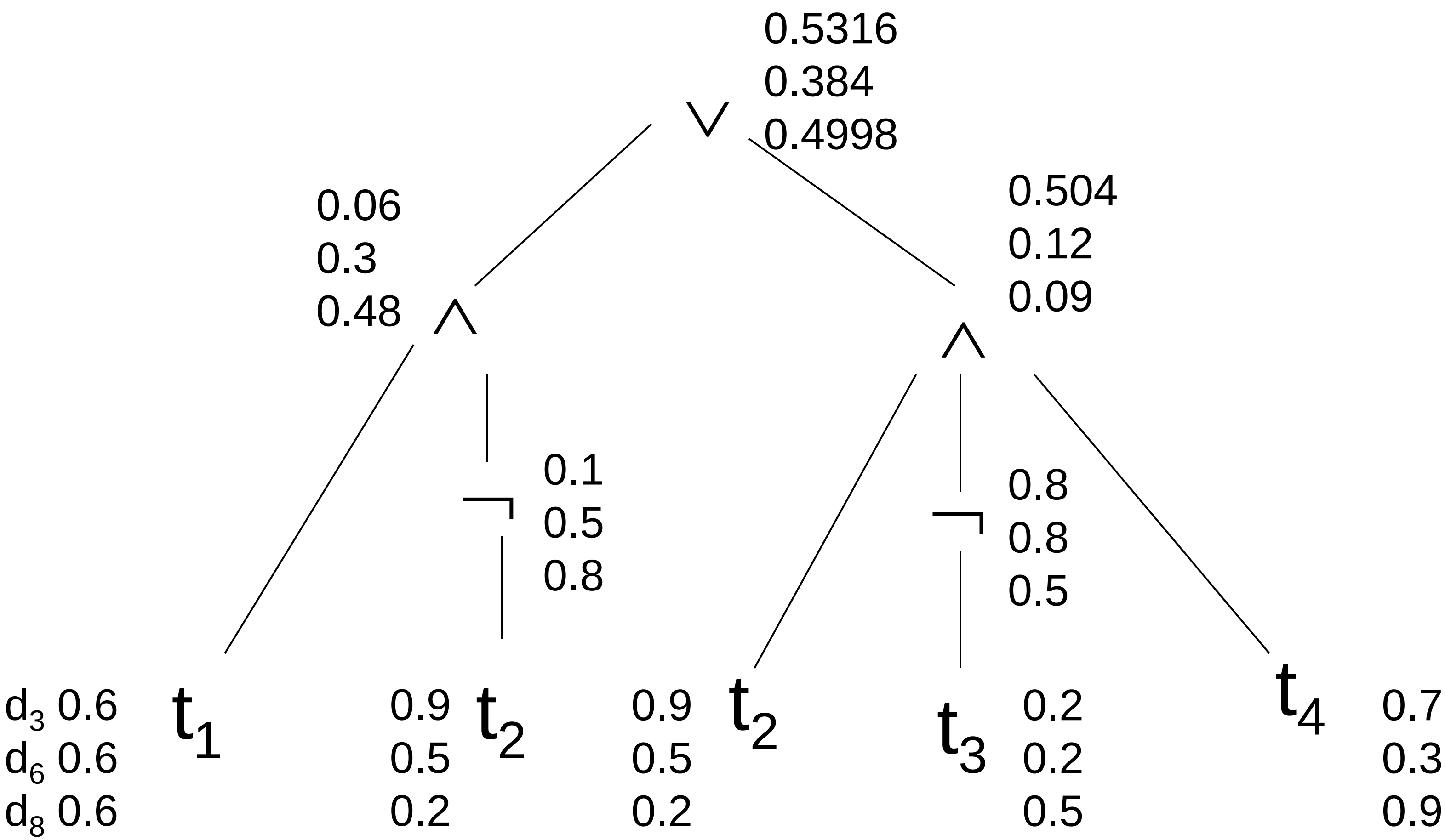
Fuzzy example
Max, Min

$q = (t_1 \wedge \neg t_2) \vee (t_2 \wedge \neg t_3 \wedge t_4))$



Fuzzy example
probabilistic sum and product

$q = (t_1 \wedge \neg t_2) \vee (t_2 \wedge \neg t_3 \wedge t_4))$



Differences

Document	Max, Min	Prob. Sum, Product
d_3	0.7	0.5316
d_6	0.5	0.384
d_8	0.6	0.4998

λ*-level Sets

λ^*-level and λ^*-level fuzzy sets

- λ^*-level Set
  - $D(\lambda^*) = \{<d,t> \mid \mu_{\mathcal{D}} (d_\alpha, t) \geq \lambda^*\}$
  - $D_t(\lambda^*) = \{ d \mid <d, t> \in D (\lambda^*)\}$

- Then meaning of $\mathcal{D}_{t(\lambda*)}$, λ^*-level fuzzy Set, of descriptor $t \in T \subset T^*$
  - $\mathcal{D}_{t(\lambda*)} = \{<d, \mu_{\mathcal{D}_{t(\lambda*)}}(d) = \mu_{\mathcal{D}} (d_\alpha, t) > \mid d \in D_t(\lambda^*)\}$

  Note: T* has same meaning as E used in Boolean IR
- If $t' \in T^*$, then meaning of $D_{t(\lambda*)}$ of descriptor $t = \neg t'$
  - $\mathcal{D}_{t(\lambda*)} = \{<d, \mu_{\mathcal{D}_{t(\lambda*)}}(d)=1-\mu_{\mathcal{D}_{t'(\lambda*)}}(d)> \mid$ $1-\mu_{\mathcal{D}_{t'(\lambda*)}}(d) > \lambda^*, d \in D_{t'}(\lambda^*) \}$

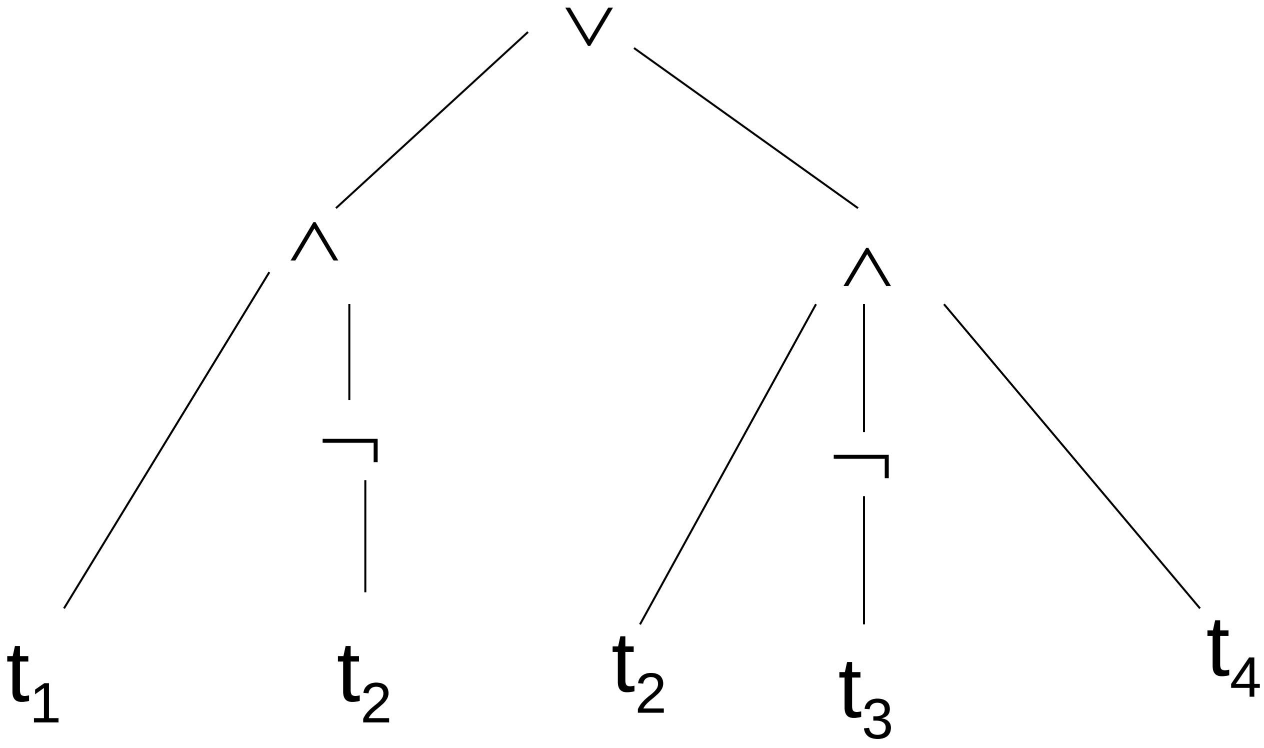
λ*-level fuzzy sets

- If t', t'' $\in$ T*, then meaning of $D_{t(\lambda*)}$ of descriptor t = t' $\wedge$ t''
  - $\mathcal{D}_{t(\lambda*)} = \{<d, \mu_{\mathcal{D}_{t(\lambda*)}}(d)=\min(\mu_{\mathcal{D}_{t'(\lambda*)}}(d), \mu_{\mathcal{D}_{t''(\lambda*)}}(d))> | d \in D_{t'}(\lambda^*) \cap D_{t''}(\lambda^*) \}$

- If t', t'' $\in$ T*, then meaning of $D_{t(\lambda*)}$ of descriptor t = t' $\vee$ t''
  - $\mathcal{D}_{t(\lambda*)} = \{<d, \mu_{\mathcal{D}_{t(\lambda*)}}(d)=\max(\mu_{\mathcal{D}_{t'(\lambda*)}}(d), \mu_{\mathcal{D}_{t''(\lambda*)}}(d))> | d \in D_{t'}(\lambda^*) \cup D_{t''}(\lambda^*) \}$

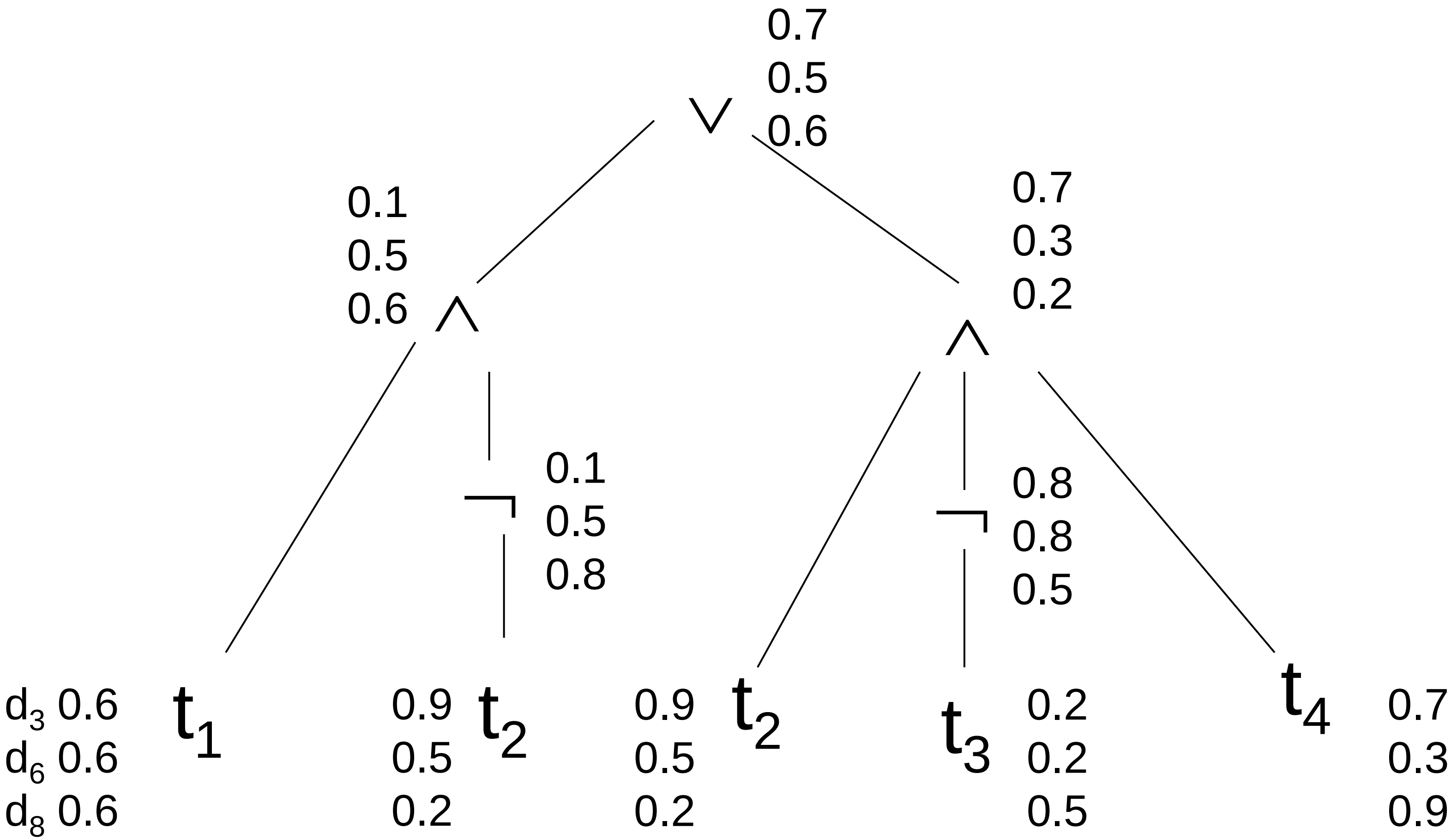
Fuzzy example

$$q = (t_1 \wedge \neg t_2) \vee (t_2 \wedge \neg t_3 \wedge t_4))$$



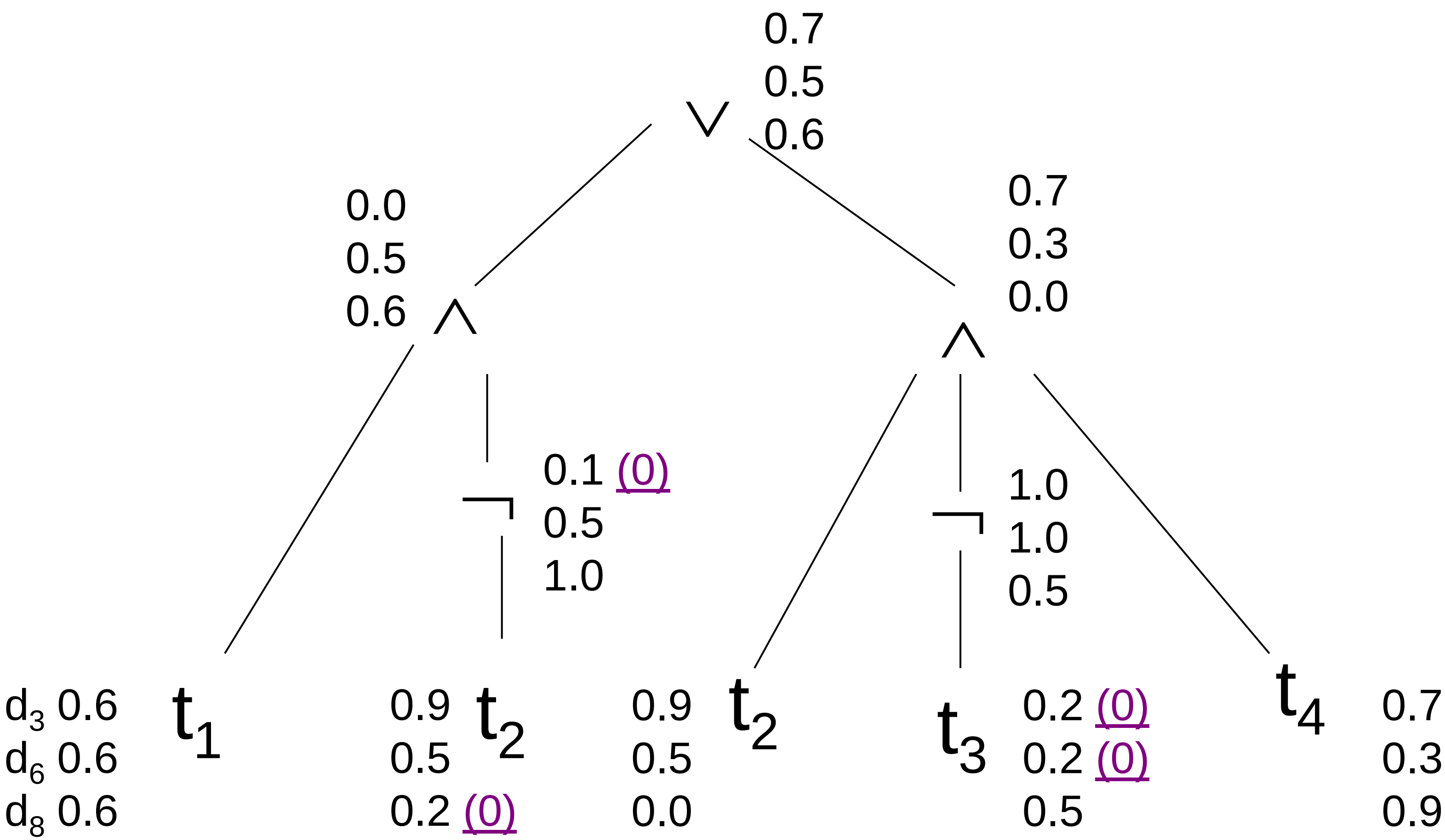
Fuzzy example
Max, Min

$q = (t_1 \wedge \neg t_2) \vee (t_2 \wedge \neg t_3 \wedge t_4))$



Fuzzy example
Max, Min, λ*=0.3

$q = (t_1 \wedge \neg t_2) \vee (t_2 \wedge \neg t_3 \wedge t_4))$



Questions?