Przetwarzanie wiedzy niepewnej

Halina Kwaśnicka
halina.kwasnicka@pwr.edu.pl

Wnioskowanie z niepewnością

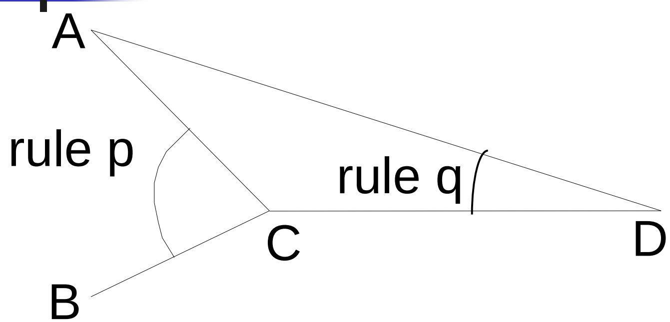
Typy niepewności:

1. Wejściowe fakty są niepewne lub mają przypisane prawdopodobieństwa
2. Reguły, nawet kiedy fakty są absolutnie pewne, generują nowe nowe fakty z pewnym stopniem ufności; np.,

IF organism is a gram positive coccus growing in chains (100% certainty)
THAN organism is a streptococcus (with 70% certainty)

3. Kombinacje przypadków 1 i 2

Podejście probabilistyczne



Logiczna reprezentacja:

$A \cap B \rightarrow C$; $A \cap C \rightarrow D$;

Probabilistyczna reprezentacja:

Dane P(A) i P(B), P(D)=?

$P(A \wedge B)=P(C)$; $P(A \wedge C)=P(D)$;

- Propagacja prawdopodobieństw
- założenie:
  - A, B: niezależne,
  - $P(A)=P(B)=0{,}5$.
- $P(C)=P(A)\cdot P(B)=0{,}5\cdot 0{,}5=0{,}25$,
- $P(D)=P(A)\cdot P(C)=0{,}5\cdot 0{,}25=0{,}125$
- Wynik jest zbyt rygorystyczny!
- Jeśli A i C są zależne:
  - $P(D)=P(A \wedge C)=P(A)\cdot P(C/A)$

- $P(C/A)=P(C \wedge A)/P(A)=P(A \wedge B \wedge A)/P(A)=P(A \wedge B)/P(A)=$
  $=(P(A)\cdot P(B))/\ P(A)=P(B)$
- Otrzymujemy: $P(D) = P(A) \cdot P(C/A) = P(A) \cdot P(B)$

Prawdopodobieństwa – zalety i wady

- Rachunek prawdopodobieństwa jest formalnym, poprawnym mechanizmem wnioskowania
- Lecz:
  - Stosowanie rachunku prawdopodobieństwa wymaga od użytkownika dostarczenia pewnej liczby prawdopodobieństw warunkowych
  - Nawet niezależne fakty początkowe nie propagują niezależności w procesie wnioskowania (patrz przykład)
  - Wnioskowanie z założeniem niezależności produkuje rygorystyczne (złe) wyjścia
  - ...

Certainty Factor (CF) – Czynnik Pewności

- Załóżmy:
- R1: IF INTERETS RATES =FALL (CF=0,6)
  AND TAXES=REDUCED (CF=0,8)
  THEN STOCK MARCET=RISE (Reguła CF=0,9)
- CF(h)=MB(h)–MD(h)
  - (MB – *Measure of Belief*),
  - (MD – *Measure of Disbelief*)
  - $0 \leq MB \leq 1$; $0 \leq MD \leq 1$; $\mathbf{-1 \leq CF \leq 1}$

CF – propagacja niepewności

- **$CF(h{:}reguła_1, reguła_2) = CF(h{:}reguła_1) + CF(h{:}reguła_2) * [1 – CF(h{:}reguła_1)]$**
- Reguła$_1$: IF A (MB(A)=0.7) THEN B.
  - **MB(B: Reguła$_1$)=MB(A)**
- Reguła$_2$: IF A (MB(A)=0.7) THEN B (MB(Reguła$_2$)= 0.6).
  - **MB(B: Reguła$_2$)=MB(A)$\cdot$MB(Reguła$_2$)**
- Reguła$_3$: IF A1(MB(A1)= 0.7) AND A2 (MB(A2)=0.5) THEN B (MB(Rule$_3$) = 0.8)
  - **MB(B: Reguła$_3$)=min{MB(A1), MB(A2)}$\cdot$MB(Reguła$_3$)**

- MB(B:Reguła)=**0,7**
- MB(B: Reguła$_1$ Reguła$_2$)=0,7+(0,7$\cdot$0,6)$\cdot$[1-0,7]=0,7+0,42$\cdot$0,3=**0,826**
- MB(B:(Reguła$_{1,}$ Reguła$_2$), Reguła$_3$)= = 0,826+min{0,7, 0,5}$\cdot$0,8$\cdot$[1-0,826]=**0,8516**

Czynnik pewności gdy jego wartości generowane z różnych źródeł mają różne znaki

$$CF(h,r_1,r_2) = \begin{cases} CF(h,r_1) + CF(h,r_2)[1 - CF(h,r_1)] & \text{dla } CF(h,r_1), CF(h,r_2) > 0 \\ CF(h,r_1) + CF(h,r_2)[1 + CF(h,r_1)] & \text{dla } CF(h,r_1), CF(h,r_2) < 0 \\ \dfrac{CF(h,r_1) + CF(h,r_2)}{1 - \min\left(|CF(h,r_1)|, |CF(h,r_2)|\right)} & \text{dla } CF(h,r_1) * CF(h,r_2) < 0 \end{cases}$$

CF – zalety i wady

- Ważne cechy CF:
  - Przesuwa zaufanie produkowanych hipotez asymptotycznie do pewności kumulując miary otrzymywane z kolejnych reguł produkujących rozważane hipotezy.
  - CF jest symetryczną miarą, tzn., jest niezależny od uporządkowania zapalanych reguł
- Lecz:
  - CF$\approx$0 – niemożliwe jest rozróżnienie czy MB i MD są prawie takie same (konflikt) czy obie miary bliskie zeru (hipoteza nie może być potwierdzona lub zanegowana)

Świat rozmyty?

- zimno – ciepło; wysoki-niski
- początek wieku, Bertrand Russel
- grecki paradoks jest kluczem do zrozumienia teorii zbiorów i logiki matematycznej
- stara łamigłówka: **pewien Kreteńczyk stwierdza, że** *„wszyscy Kreteńczycy kłamią”*
- pytanie, czy wobec powyższego on też kłamie?
- odpowiedź „tak”
- odpowiedź „nie”
- zdanie jest jednocześnie prawdziwe i fałszywe

W starej łamigłówce pewien **Kreteńczyk stwierdza, że** „wszyscy Kreteńczycy kłamią”. Nasuwa się pytanie, **czy wobec powyższego on też kłamie**?

Jeśli odpowiedź będzie „tak”, to znaczy, że mówi prawdę, a zatem kłamie.

Jeśli ten Kreteńczyk mówi prawdę (odpowiedź „nie”), to znaczy, że jest kłamcą mówiąc, że wszyscy kłamią.

Oba przypadki prowadzą do sprzeczności wg logiki klasycznej, ponieważ zdanie jest jednocześnie prawdziwe i fałszywe.

Logika w świecie rozmytym

- logika rozmyta gałąź informatyki
- cel - umożliwienie komputerom tworzenia zdroworozsądkowego, zróżnicowanego obrazu niepewnej rzeczywistości
- podstawę stworzyli matematycy zajmujący się logiką matematyczną:
  - w latach dwudziestych sformułowali zasadę, że wszystko można stopniować
  - Jan Łukasiewicz (1878-1956) sformułował zasady logiki wielowartościowej - wartość zdań określa się wartościami pośrednimi między zerem a jedynką

Logika wielowartościowa

$$V(p)=\left\{0,\frac{1}{p-1},\frac{2}{p-1},\ldots,\frac{p-2}{p-1},1\right\}$$

- *V(p)* - zbiór możliwych wartości logicznych
- *p* – liczba wartości logicznych w p-wartościowej logice

$$T(\neg A)=1-T(A) \qquad T(A)\wedge 1=T(A)$$

$$T(A)\wedge 0=0 \qquad T(A)\vee 1=1$$

$$T(A)\vee 0=A$$

- Jak zrobić tabelę wartości np. dla *p*=3?
- W oparciu o własności logiki klasycznej – nie da się

Jak działają operatory logiczne?

	$\neg$
0	1
0,5	0,5
1	0

$\vee$	0	0,5	1
0	0	0,5	1
0,5	0,5	?	1
1	1	1	1

$\wedge$	0	0,5	1
0	0	0	0
0,5	0	?	0,5
1	0	0,5	1

- Przykład

  B = A to: $A \wedge B = A \wedge A = A$ (1)

  B = $\neg$ A to: $A \wedge B = A \wedge \neg A = 0$ (2)
- (1) i (2) prawdziwe w logice klasycznej
- Niech A=0,5 ($\neg$ A =0,5)

  z (1): $A \wedge B = 0{,}5 \wedge 0{,}5 = 0{,}5$ (3)

  z (2): $A \wedge B = 0{,}5 \wedge 0{,}5 = 0$ (4)

Wybrane operatory w logice rozmytej

Wybrane operatory w logice rozmytej:

silna koniunkcja: $T(A \wedge B) = \min\{T(A), T(B)\}$

słaba koniunkcja: $T(A \otimes B) = \max\{0, A+B -1\}$

silna alternatywa: $T(A \vee B) = \max\{T(A), T(B)\}$

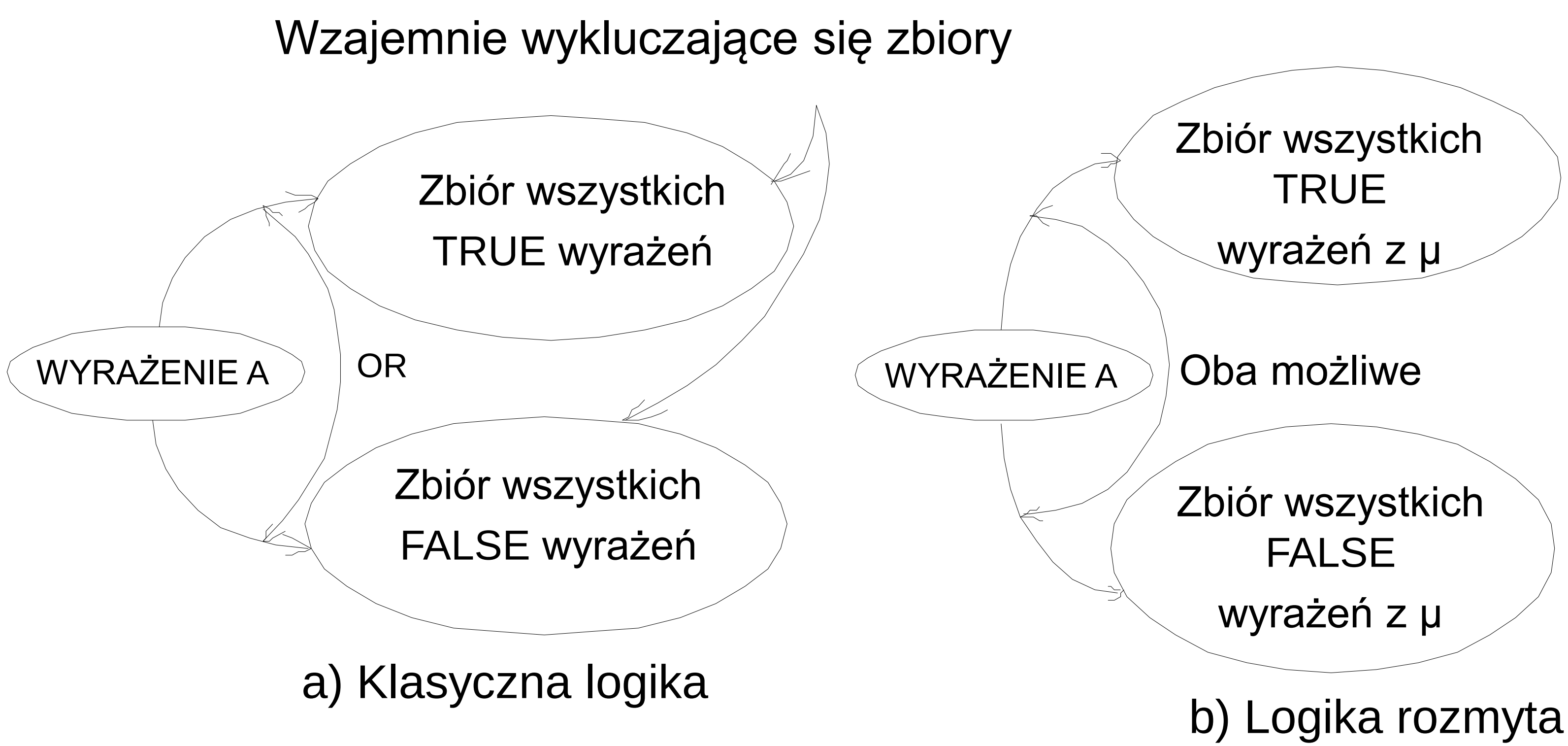
słaba alternatywa: $T(A \oplus B) = \min\{1, T(A)+T(B)\}$

Negacja i implikacja są jednakowo definiowane:

negacja: $T(A) = 1 – T(A)$,

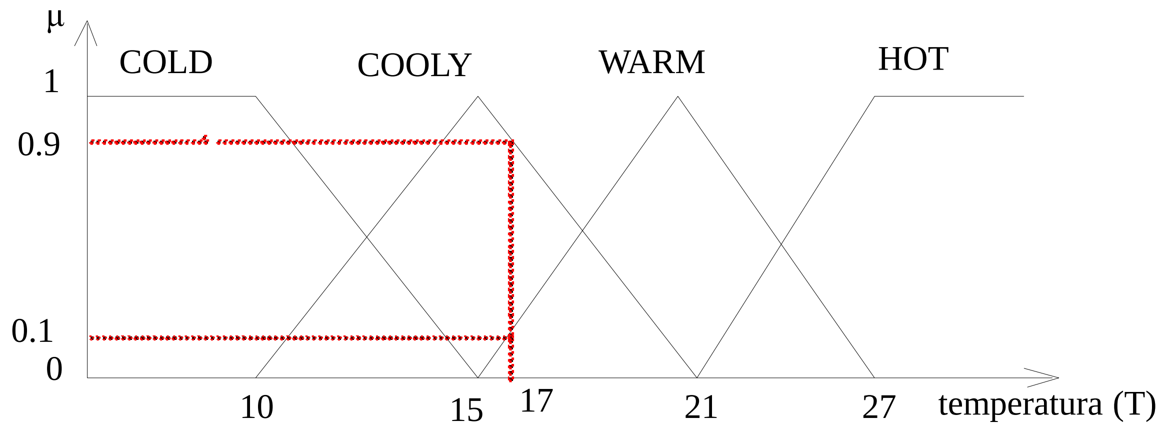
implikacja: $T(AB) = T(AB)$

Fuzzy versus klasyczna logika



Rozmywanie i wnioskowanie rozmyte

- Funkcja przynależności μ dla zmiennej *temperatura* (T)

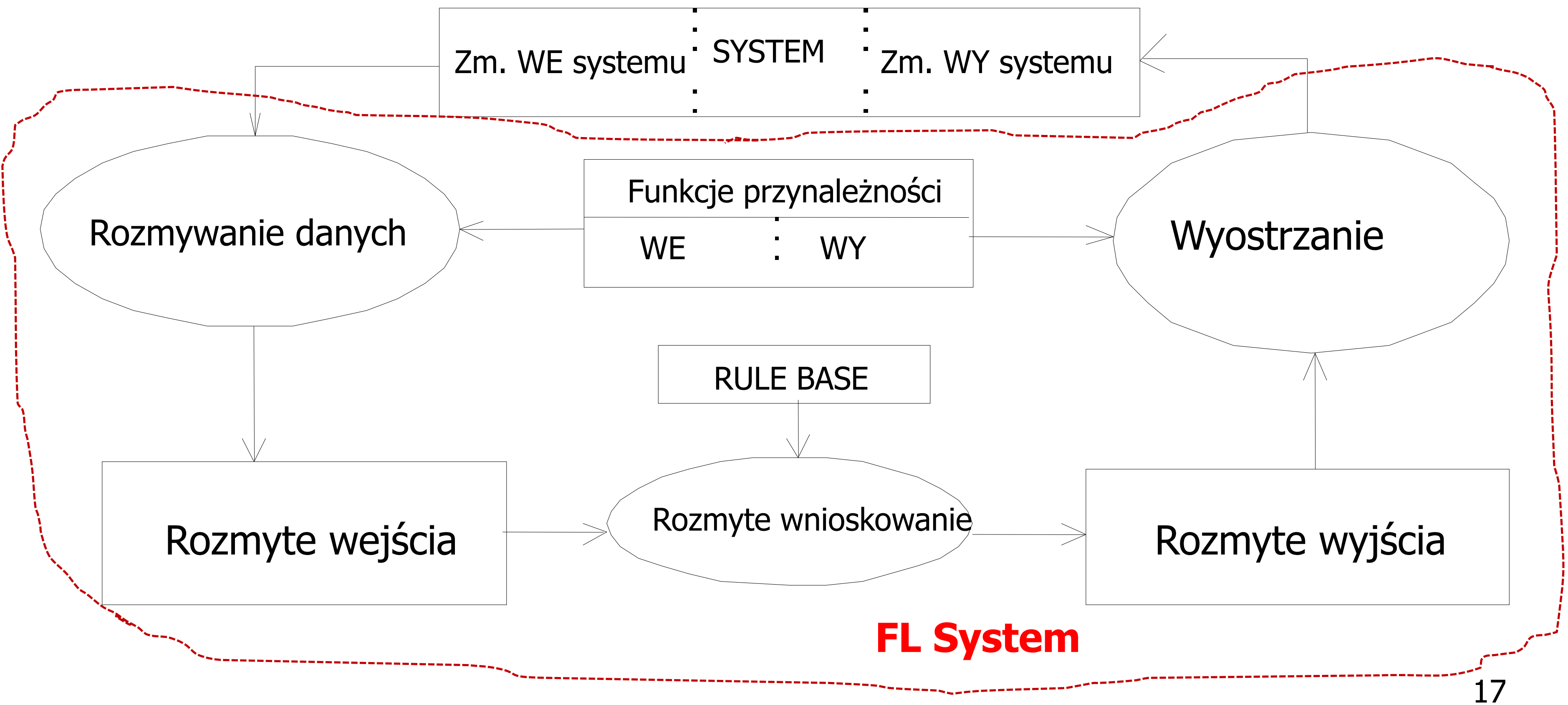


- Reguła 1: if **A** and **B** then **Z** and **X**; Reguła 2: if **C** and **D** then **Z** and **Y**
- Siła(Reguła 1)=min{μ(A), μ(B)},
- Siła(Reguła 2)= min{μ(C), μ(D)},
- Siła akcji X = Strength(Reguła 1),
- Siła akcji Y = Strength(Reguła 2),
- Siła akcji Z = max**{**Strength(Reguła 1), Strength(Reguła 2)**}**

  = max**{**min**(**μ(A), μ(B)**)**, min**(**μ(C), μ (D)**)}**.

Logika rozmyta

- Rozmyty system ekspertowy



Rozmyte reguły – przykłady

- **If** *cena _akcji_spada* **and** *obroty_są_duże* **then** *sprzedaj_akcje*.
- **If** jest *ciepło* **then** obroty silnika ustaw *małe*.
- **If** jest *chłodno* **then** obroty silnika ustaw *średnie*.
- Podczas wnioskowania sprawdza się, czy dana reguła może być zastosowana, tzn. czy spełnione są jej przesłanki.
- Jeśli tak, to w oparciu o przypisane przesłankom wartości przynależności, wyliczana jest moc każdej akcji produkowanej przez rozpatrywana regułę.

Rozmyte reguły – przykłady

- Kilka reguł może produkować tę samą akcję, każda z różną siłą – jaka jest ostateczna siła produkowanej akcji.
- Korzysta się z interpretacji ***koniunkcji*** i ***alternatywy*** dla logiki wielowartościowej.
- Przykład:

  R1: jeśli A i B *to* Z i Y i *R2: jeśli* C i D *to* Z i Y

- Siła *R1* = min(wartość przynależności A, wartość przynależności B),
- Siła *R2* = min(wartość przynależności C, wartość przynależności D),
- Siła akcji X = Siła *R1*,
- Siła akcji Y = Siła *R2*,
- Siła akcji Z = max(Siła *R1*, Siła *R2*) =
- = max{min(wartość przynależności A, wartość przynależności B),
- min(wartość przynależności C, wartość przynależności D)}.

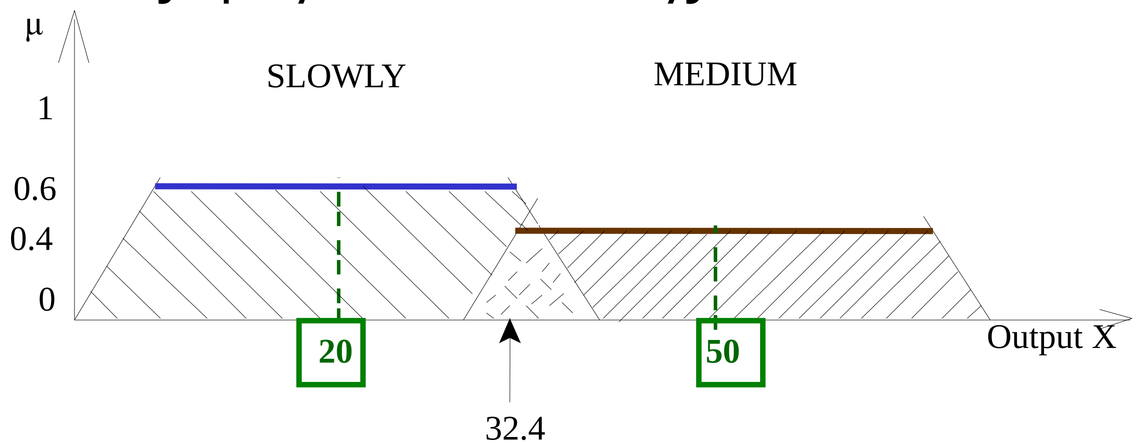
Przykład rozmytego wnioskowania

Załóżmy, że:

- Rozmyty zbiór SLOWLY zdefiniowany jest dla $X \in [0, 40]$,
- Rozmyty zbiór MEDIUM zdefiniowany jest dla $X \in [30, 70]$,
- Punkty obcięcia dla SLOWLY są: $X1=6$, $X2=34$, $\mu=0.6$,
- Punkty obcięcia dla MEDIUM są: $X3= 34$, $X4= 66$, $\mu=0.4$,
- Inne rozmyte wartości nie zostały wywnioskowane (ich siła jest równa zero),
- Centroid dla SLOWLY = 20,
- Centroid dla MEDIUM=50.

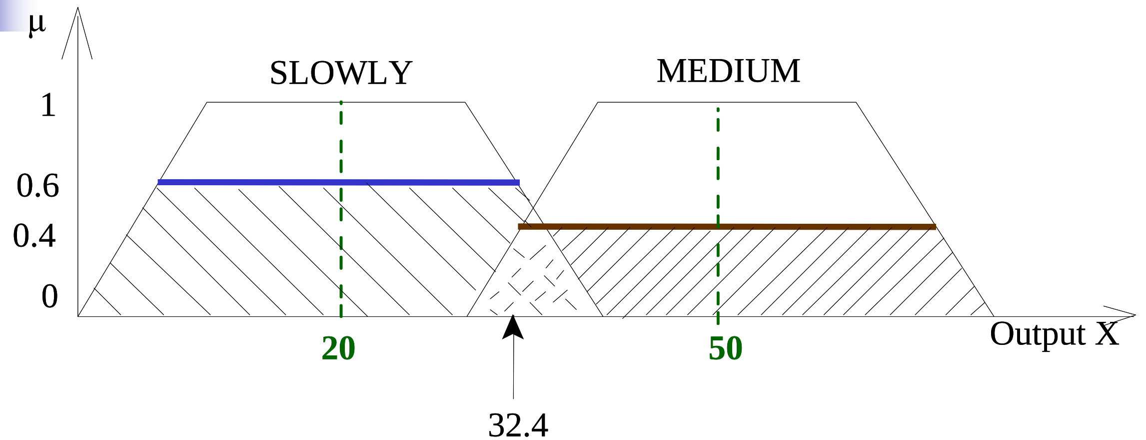
Proces wyostrzania – przykład

- Funkcje przynależności – wyjście X



- Obliczenie środka ciężkości (centroidu) każdej funkcji przynależności (rozmyte wartości wyjść) – wykonane na etapie przygotowywania systemu!
- Obcięcie funkcji przynależności do poziomu siły, z jaka dane wyjście jest produkowane (niebieska i brązowa linia na rysunku).
- Obliczenie pola powierzchni obciętych funkcji przynależności
- Ostra wartość wyjścia jest obliczana jako średnia ważona współrzędnych centroidów x i obliczonych pól, z polami jako wagami

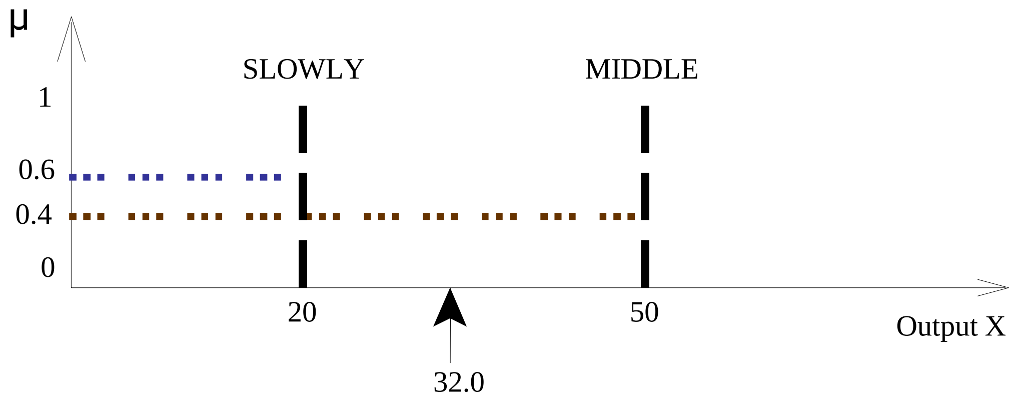
Przykład wyostrzania



- Centroid (T=SLOWLY)=20,
- Centroid (X=MEDIUM)=50,
- Pole obciętej figury SLOWLY=0,6·(40+28)/2 = 20,4,
- Pole obciętej figury MEDIUM=0,4·(40+32)/2 = 14,4.
- Ostre wyjście $= \dfrac{20{,}4 \cdot 20 + 14{,}4 \cdot 50}{20{,}4 + 14{,}4} = 32{,}4$

Uproszczony proces wyostrzania

- Nie musimy obliczać pola obciętych figur
- Wuykorzystujemy singletony (zobacz rysunek niżej)



- Ostra wartośc wyjścia jest obliczana jako średnia wazona centroidów (singletonów) z siłą jako wagi.
- Ostre Wyjście = $\dfrac{0{,}6 \cdot 20 + 0{,}4 \cdot 50}{0{,}6 + 0{,}4} = 32{,}0$

Sieci Bayesowskie (BN) - definicja

- Inne nazwy:
  - belif network, Bayesan network, probabilistic network, causal network, knowledge map
- Sieć Bayesowska to acykliczny graf (DAG, ***D**irected **A**cyclic **G**raph*), składający się z:
  - zbioru wierzchołków odpowiadających zbiorowi zmiennych
  - zbioru skierowanych krawędzi łączących pary węzłów – intuicyjne znaczenie połączenia od węzła *A* do *B* oznacza, że *A bezpośrednio wpływa na B*
  - graf nie zawiera cykli
- Węzły, od których dochodzą krawędzie do danego węzła to węzły rodzicielskie
- Każdy węzeł zawiera tabelę prawdopodobieństw warunkowych, określających wpływ węzłów 'rodzicielskich' na dany węzeł

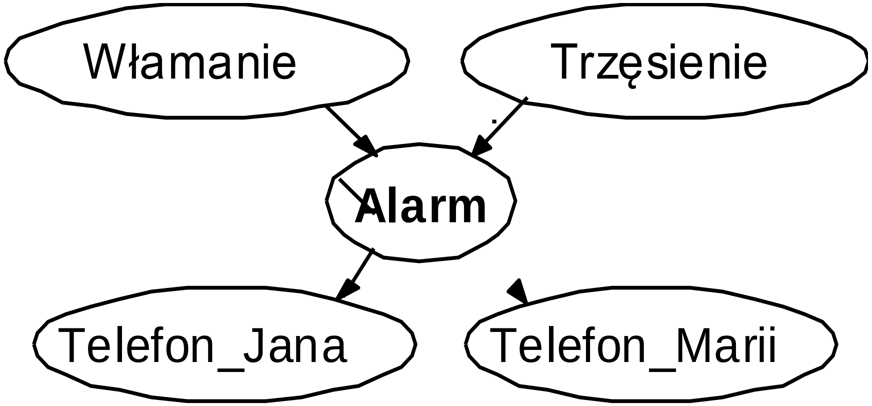
Cechy BN

- łatwiej ekspertom określić bezpośrednie zależności warunkowe w dziedzinie, niż podawać aktualne prawdopodobieństwa
- po zbudowaniu topologii BN, należy określić prawdopodobieństwa warunkowe dla węzłów, które są ze sobą bezpośrednio połączone
- te prawdopodobieństwa są wykorzystywane do obliczania każdego innego prawdopodobieństwa
- topologia sieci bayesowskiej - abstrakcyjna baza wiedzy: obejmuje ogólną strukturę przyczynowych procesów w rozpatrywanym wycinku dziedziny

Przykład CPT

Prawdopodobieństwa dla węzła Alarm (CPT, *Conditional Probability Table*):

Włamanie	Trzęsienie	P(Alarm/Włamanie,Trzęsienie)	
		True	False
True	True	0,950	0,050
Truse	False	0,950	0,050
False	True	0,290	0,710
False	False	0,001	0,999



Suma prawdopodobieństw w wierszach musi być równa **jeden**

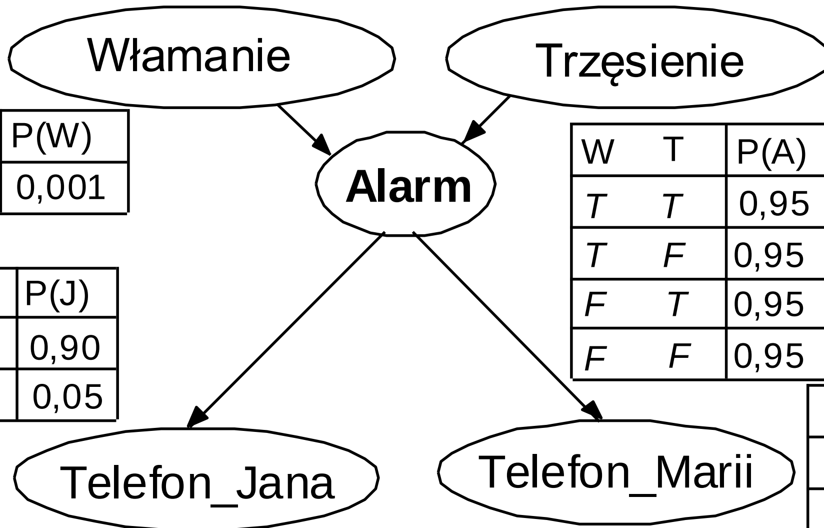
- w tej sytuacji określenie jednej wartości w wierszu wystarczy (druga - dopełnienie do jedynki)
- węzeł o *n* rodzicach - 2^n niezależnie określanych prawdopodobieństw
- węzeł bez rodzicielskich ma tabelę jednowierszową - wartości prawdopodobieństw **a priori** dla każdej możliwej wartości zmiennej

BN - przetwarzanie

- sieć bayesowska dostarcza pełnego opisu wycinka dziedziny
- łączne prawdopodobieństwo jest koniunkcją poszczególnych przypisań każdej zmiennej, czyli:
- P(X_1=x_1 ∧ X_2=x_2 ∧ ... ∧ X_n=x_n), zapis w krótszej formie: P(x_1, ... , x_n)
- wartość takiego prawdopodobieństwa można wyliczyć ze wzoru:

$$P(x_1, ..., x_n) = \prod_{i=1}^{n} P(x_i \mid Parents(X_i))$$

Przykład sieci bayesowskiej



P(W)
0,001

P(T)
0,002

W	T	P(A)
T	*T*	0,95
T	*F*	0,95
F	*T*	0,95
F	*F*	0,95

A	P(J)
T	0,90
F	0,05

A	P(M)
T	0,70
F	0,01

- CPT - zdekomponowana reprezentacja łącznych prawdopodobieństw
- można wyliczyć np., prawdopodobieństwo zdarzenia: alarm zadziałał, nie było włamania, ani trzęsienia, zadzwonił Jan i Maria
- $P(J \wedge M \wedge A \wedge \neg W \wedge \neg T) = P(J|A)P(M|A)P(A|\neg W \wedge \neg T)P(\neg W)P(\neg T) =$
  $= 0{,}90 \cdot 0{,}70 \cdot 0{,}001 \cdot 0{,}999 \cdot 0{,}998 = 0{,}00062$

Przyrostowa konstrukcja BN

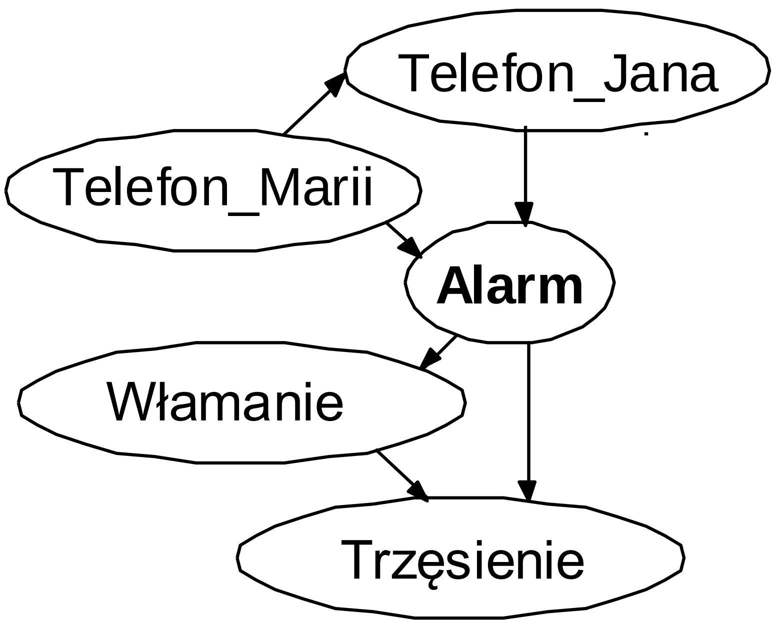
- Wybierz zbiór odpowiednich zmiennych X_i opisujących daną dziedzinę
- Wybierz **uporządkowanie** zmiennych (wykorzystaj tu wiedzę o dziedzinie)
- Dopóki zbiór zmiennych nie jest pusty, dla i=1 do n:
  - a)weź kolejną zmienną X_i i dodaj dla niej węzeł w sieci
  - b)ustal zbiór (*Parents*(X_i) jako pewien minimalny zbiór węzłów **istniejących już w sieci** w ten sposób, aby spełniony był warunek dany równaniem $P(X_i | X_{i-1}, ..., x_1) = P(X_i | Parents(X_i))$
  - c)zdefiniuj tablicę prawdopodobieństw warunkowych dla X_i
- Mamy gwarancję, że:
  - sieć jest acykliczna
  - nie zawiera nadmiarowych wartości prawdopodobieństw i nie jest sprzeczna z aksjomatami prawdopodobieństwa

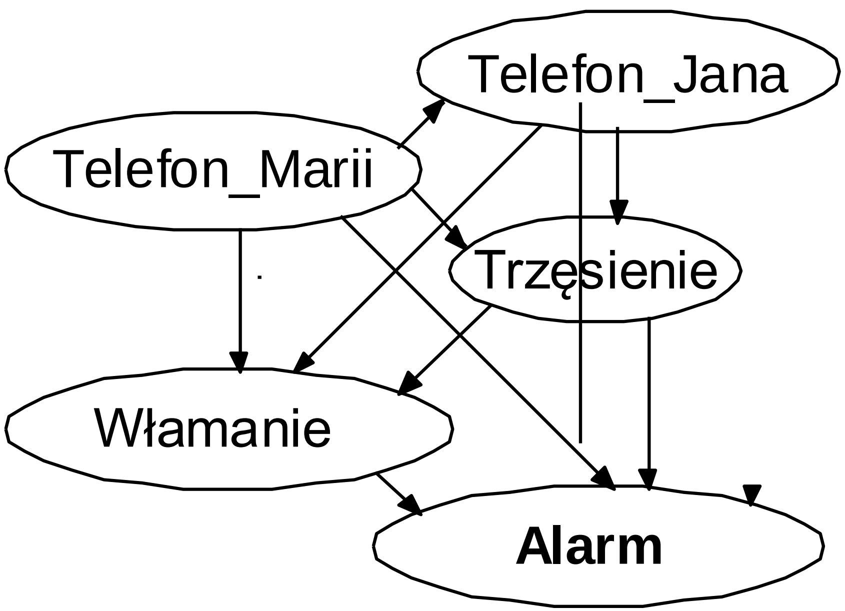
Zwartość a uporządkowanie węzłów

- BN może być bardziej zwarta, niż sieć w pełni połączona (można uwzględnić dużą liczbę zmiennych bez eksponencjalnego wzrostu liczby prawdopodobieństw warunkowych)
- BN to przykład ogólnej własności
  - *systemów lokalnie strukturyzowanych* (*locally structured systems*), inaczej: *systemów rzadkich* (*sparce systems*)
  - w systemach rzadkich, każdy składnik jest połączony tylko z ograniczoną liczbą innych składników
  - zwykle w systemach rzadkich złożoność obliczeniowa rośnie raczej liniowo a nie eksponencjalnie
- jeśli każda zmienna bezpośrednio wpływa na wszystkie inne, to sieć jest w pełni połączona - wymaga tylu samo wartości, co podejście z prawdopodobieństwem łącznym

Wpływ uporządkowania - przykład

Budowa rzadkiej BN nie jest zadaniem trywialnym – ważne uporządkowanie zmiennych, wpływa na stopień komplikacji BN



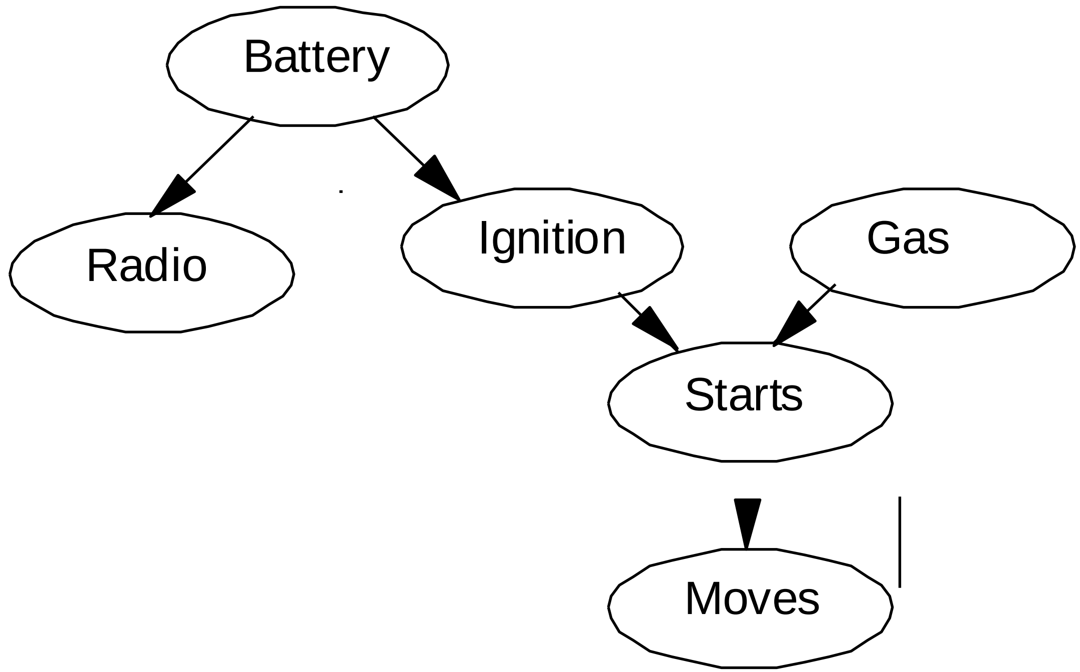


Zależność zmiennych

To, czy jest *paliwo* w samochodzie oraz to, czy gra w nim *radio*, są niezależne przy zdarzeniu *działa zapłon*

- *Paliwo* i *radio* są niezależne jeśli wiadomo, że *akumulator* działa
- *Paliwo* i *radio* są niezależne bez dowodów (zbioru zdarzeń *E*), ale są zależne przy wiedzy, czy silnik *zastartował*. Np. jeśli silnik nie *zastartował*, to grające *radio* zwiększa wiarygodność tego, że nie ma paliwa. *Paliwo* i *radio* są również zależne przy znanym węźle *Moves* (*samochód ruszył*), ponieważ ta wartość informuje, iż silnik *zastartował*

Przykład sieci bayesowskiej



Rysunek 1. Przykład sieci bayesowskiej: system elektryczny silnika samochodowego

Węzły deterministyczne

- *deterministyczny węzeł* – jest najprostszy, ma wartość określoną dokładnie przez wartość rodziców, bez niepewności, powiązania mogą być:
  - logiczne – np. węzły-rodzice: *Canadian*, *US*, *Mexican*; węzeł-dziecko: *NorthAmerican*; powiązanie: wartość potomka jest dysjunkcją rodziców
  - numeryczne: np., węzły-rodzice: ceny poszczególnych modeli samochodów u kilku dilerów; węzeł-dziecko: cena transakcji; powiązanie: wartość potomka jest minimum węzłów rodzicielskich; węzły-rodzice: miary dopływu i wypływu z jeziora; węzeł-dziecko: poziom wody w jeziorze; powiązanie: różnica między wpływem a wypływem
- *Niepewne relacje* mogą być charakteryzowane przez tzw. ‘zaszumione’ relacje logiczne: *noisy-OR* - uogólnienie logicznej *OR*

noisy-OR - uogólnienie logicznej *OR*

- W logice: *przeziębienie=true* lub *grypa=true* lub *malaria=true* to *gorączka=true*
- Model *noisy-OR* dodaje pewną niepewność, czyni trzy założenia
  - każda przyczyna ma niezależną szansę powodować efekt
  - wszystkie możliwe przyczyny są wymienione (założenie to może być osłabione przez dodanie tzw. węzła nieszczelności (*leak node*), który pokrywa ‘opuszczone przyczyny’
  - cokolwiek wstrzymującego, powiedzmy, *przeziębienie* powodującego *gorączkę* jest niezależne od czegokolwiek wstrzymującego *grypę* powodującą gorączkę
  - te ‘inhibitory’ nie są reprezentowane jako węzły, są raczej uwzględniane jako „parametry szumu”

Parametry szumu (*noise parameters*)

- Jeśli
  - P(*gorączka*|*przeziębienie*)=0,4
  - P(*gorączka*|*grypa*)=0,8
  - P(*gorączka*|*malaria*)=0,9
- to parametry szumu wynoszą odpowiednio 0,6, 0,2 i 0,1
- Jeśli żaden rodzic nie jest *true* to potomek=*false* w 100%
- Jeśli dokładnie jeden rodzic ma wartość *true*, to potomek ma *false* (prawdopod.= parametrowi szumu tego węzła)
- W ogólności

  Prawdopodobieństwo, że potomek ma wartość *FALSE* jest iloczynem parametrów szumu wszystkich węzłów rodzicielskich, które są *TRUE*
- Dla omawianego tu przykładu gorączki, mamy

Przykład

Cold	**Flu**	**Malaria**	**P(Fever)**	**P(¬*Fever*)**
F	F	F	0,0	1,0
F	F	T	0,9	0,1
F	T	F	0,8	0,2
F	T	T	0,98	0,02 = 0,2 · 0,1
T	F	F	0,4	0,6
T	F	T	0,94	0,06 = 0,6 · 0,1
T	T	F	0,88	0,12 = 0,6 · 0,2
T	T	T	0,988	0,012 = 0,6 · 0,2 · 0,1

P(*Fever*|*Cold*)=0,4 — Parametr szumu=0,6

P(*Fever*|*Flu*)=0,8 — Parametr szumu=0,2

P(*Fever*|*Malaria*)=0,9 — Parametr szumu=0,1

Wnioskowanie w sieciach bayesowskich

Podstawowym zadaniem każdego systemu wnioskowania probabilistycznego jest wyliczenie rozkładu prawdopodobieństwa a posteriori dla zbioru zmiennych zapytań (*query variables*), przy zadanych dokładnie wartościach pewnych zmiennych dowodzących (*evidence variables*).

Oznacza to, iż system oblicza P(*Query*|*Evidence*)

Za **samodzielne** przygotowanie szczegółowej wiedzy nt. wnioskowania w BN, z przykładami, duży BONUS (wykracza poza zakres obowiązkowego materiału)