



Relevanța conținutului web și a comportamentului utilizatorilor în analiza traficului

Raport de cercetare

Student doctorand:

Diana - Florina Haliță (căș. Șotropa)

Coordonator științific:

Prof. Dr. Florian Mircea Boian

9 Noiembrie 2016

Universitatea Babeș-Bolyai, Cluj-Napoca, România

1. Cercetări anterioare
2. Rezultate și contribuții proprii
3. Analiză Formală Conceptuală n-adică
4. Concluzii și direcții viitoare

Cercetări anterioare

- determinarea tiparelor comportamentale cantitative: *relaxat*, *normal* și *intens*;
- identificarea tiparelor comportamentale calitative: *atractori*.

Rezultate și contribuții proprii

- utilizarea FCA 4-adic pentru detectarea tiparelor comportamentale repetitive [ICCS 2016, SOFTCOMM 2016];
- utilizarea FCA 5-adic pentru detectarea trendsetter-ilor și a follower-ilor [AI4KM 2016];

Analiză Formală Conceptuală n- adică

- detectarea tiparelor comportamentale repetitive folosind FCA;
- detectarea comportamentelor online similare și gruparea utilizatorilor în *bundle-uri*;
- identificarea sub-claselor de comportamente complexe;
- observarea și înțelegerea comportamentelor utilizatorilor și a interacțiunii lor în folosirea unei platforme de e-learning;
- determinarea utilizatorilor care inițiază anumite tipuri de comportamente, i.e. *trendsetter*;
- determinarea utilizatorilor care preiau comportamentul unui trendsetter, i.e. *follower*.

lanț

succesiune de pagini vizitate de către un utilizator într-o sesiune;

- compararea lanțurilor de pagini;

⇒ *obiecte* = lanțuri de pagini cu similaritate ≥ 80

$([L - LPs, L - LPs - WE], [w1 = 15, w1 = 16], [w2 = 16])$

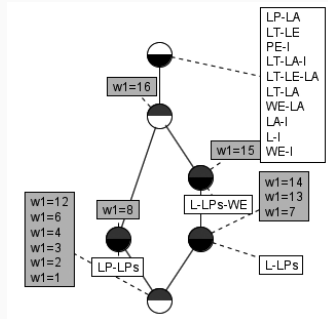


Figure 1: Proiecție pe condiția $w2=16$

$([L - LPs, L - LPs - WE], [w1 = 15, w1 = 16], [w2 = 16])$

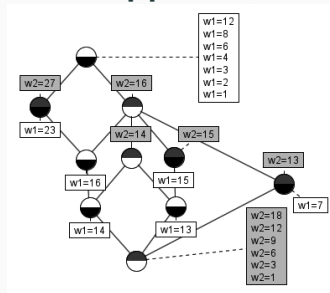


Figure 2: Proiecție pe atributul $[L-LPs]$

Definiție

Un n -context este un $(n + 1)$ -uplu $\mathbb{K} := (K_1, \dots, K_n, Y)$ cu K_1, \dots, K_n mulțimi și $Y \subseteq K_1 \times \dots \times K_n$ relația de incidență n -ară. Un n -concept al unui n -context \mathbb{K} este un n -uplu (A_1, \dots, A_n) cu proprietatea că $A_1 \times \dots \times A_n \subseteq Y$ și pentru orice n -uplu (C_1, \dots, C_n) cu $C_i \subseteq A_i, \forall i \in \{1, \dots, n\}$ relația de incidență: $C_1 \times \dots \times C_n \subseteq Y \Rightarrow C_i = A_i, \forall i \in \{1, \dots, n\}$ este satisfăcută.

Setul de date

Obiecte, Atribute, Condiții, Stări

Generarea 4-concepte: Generalizarea algoritmului Trias

Pentru fiecare stare considerăm tricontextul corespunzător. Cuboidele 3D se completează cu a 4-a dimensiune ținând cont de condiția de maximalitate.

Generarea 4-concepte: Algoritm

Let O, A, C be three sets, representing objects, attributes and conditions which are common for all the tricontexts.

Let $\mathbb{K}_1 := (K_{11}, K_{12}, K_{13}, Y_1), \dots, \mathbb{K}_n := (K_{n1}, K_{n2}, K_{n3}, Y_n)$ be n tricontexts with $K_{i1} \subseteq O, K_{i2} \subseteq A, K_{i3} \subseteq C$ and $Y_i \subseteq K_{i1} \times K_{i2} \times K_{i3}$ the ternary relation between them, $\forall i \in \{1, \dots, n\}$.

$\Theta = \emptyset$;

$\forall \mathbb{K}_i, i \in \{1, \dots, n\}$

$A_4 = \text{state}(\mathbb{K}_i)$;

for all $(A_1, A_2, A_3) \subseteq Y_i$ do $\text{add}(\Theta, (A_1, A_2, A_3, A_4))$;

$\forall (\theta_i, \theta_j) \subseteq \Theta \times \Theta$

if $(\theta_i(A_4) \subseteq \theta_j(A_4) \ \& \ \exists l, m \in \{1, 2, 3\}, l \neq m$ so that

$\theta_i(A_l) \cap \theta_j(A_l) \neq \emptyset \ \& \ \theta_i(A_m) \cap \theta_j(A_m) \neq \emptyset)$ then $\theta_i = \theta_i \cup \theta_j$;

Generarea 4-concepte: Algoritm

$\forall (\theta_i, \theta_j) \subseteq \Theta \times \Theta$

if $(\theta_i(A_4) == \theta_j(A_4) \ \& \ \exists l, m \in \{1, 2, 3\}, l \neq m$ so that
 $\theta_i(A_l) == \theta_j(A_l) \ \& \ \theta_i(A_m) == \theta_j(A_m))$ then $\theta_i = \theta_i \cup \theta_j$; &
delete(θ_j);

$\forall (\theta_i, \theta_j) \subseteq \Theta \times \Theta$

if $(\theta_i == \theta_j)$ then delete(θ_i);

$\forall (\theta_i, \theta_j) \subseteq \Theta \times \Theta$

if $(\theta_i(A_4) == \theta_j(A_4) \ \& \ \exists l, m, n \in \{1, 2, 3\}, l \neq m \neq n \neq l$ so that
 $\theta_i(A_l) == \theta_j(A_l) \ \& \ \theta_i(A_m) == \theta_j(A_m) \ \& \ \theta_i(A_n) \subseteq \theta_j(A_n))$ then
delete(θ_i);

Generarea 4-concepte: Algoritm

$\forall \theta_i \subseteq \Theta \ \& \ \forall \mathbb{K}_j, j \in \{1, \dots, n\} \ \& \ \forall (A_1, A_2, A_3) \subseteq Y_j$
if($A_1 == \theta(A_1) \ \& \ A_2 == \theta(A_2) \ \& \ A_3 == \theta(A_3) \ \&$
 $state(\mathbb{K}_j) == \theta(A_4)$) then save(i)=true;

$\forall \theta_i \subseteq \Theta$
if(!save(i)) then delete(θ_i);

Θ is the list of all tetraconcepts derived from all the triconcepts.

Exemplu: 4-context generat pentru utilizatorul X

- *obiecte*: mulțimea utilizatorilor care au un comportament similar cu cel al utilizatorului X , inclusiv X ;
- *attribute*: mulțimea săptămânilor în care comportamentul a apărut la utilizatorul X ;
- *conditii*: mulțimea săptămânilor în care un comportament similar cu al utilizatorului X a fost detectat la un utilizator din mulțimea de obiecte;
- *stări*: lista comportamentelor similare detectate.

Exemplu: Analiză tetrică a utilizatorului K pentru lanțul "LA-I"

- ([u2=H,K,O,T], [w1=2], [w2=2], [LA-I])
- ([u2=E], [w1=1,w1=2,w1=3], [w2=6], [LA-I])
- ([u2=K], [w1=2], [w2=2], [LA-I,WE-LA-I])
- ([u2=B,H,T], [w1=3], [w2=2], [LT-LA-I])
- ([u2=H], [w1=3], [w2=2], [LT-LA-I,LT-LA])
- ([u2=K], [w1=3], [w2=3], [LT-LA-I,L-LA-I,LT-LA,L-LA,L-WE-LA])
- ([u2=O], [w1=3], [w2=2], [L-LA-I])
- ([u2=B], [w1=9], [w2=7], [L-PE-LA-I])
- ([u2=K], [w1=16], [w2=16],
[LP-LP_s-PE-LA-I,LP-LP_s,L-LP_s,L-WE,L_s-L-LP])

Bundle-ul utilizatorilor care au comportamente similare cu utilizatorii K si H pentru lanțul "L-LP"

Student K

- utilizatori cu comportamente similare = G, H, K
- w_1 utilizator K = 7, 13
- w_2 pentru utilizatorii care au comportament similar cu cel al utilizatorului K = 5, 6, 7, 11, 14

Student H

- utilizatori cu comportamente similare = B, D, F, G, H, I, W, K, L, M, O, Q, T, X
- w_1 utilizator H = 4, 5, 6, 7, 9
- w_2 pentru utilizatorii care au comportament similar cu cel al utilizatorului H = 3, 4, 5, 6, 7, 8, 9, 11, 12, 14, 15, 17, 20

Setul de date

Obiecte, Attribute, Condiții, Stări1, Stări2

- *obiecte*: mulțimea de *followeri*;
- *attribute*: lista comportamentelor similare;
- *conditii*: mulțimea săptămânilor în care comportamentul a apărut la utilizatorul X ;
- *stări1*: mulțimea săptămânilor în care un comportament similar cu al utilizatorului X a fost detectat la un utilizator din mulțimea de obiecte; detectate.
- *stări2*: mulțimea de *trendsetteri*.

Utilizatorul F - trendsetter pentru "Ls-LP-LT-LA"

Obiecte	Atribute	Conditii (w1)	Stări 1 (w2)	Stări 2
F, B, D, H, Q	Ls-LP-LT-LA	4	4	F
D	Ls-LP-LT-LA	4	4, 6, 7	F
...	Ls-LP-LT-LA	4	...	F
X	Ls-LP-LT-LA	4	8	F
S	Ls-LP-LT-LA	4	13	F
...	Ls-LP-LT-LA	4	...	F
G	Ls-LP-LT-LA	8	15	X
S	Ls-LP-LT-LA	13	13	S
L,W	Ls-LP-LT-LA	13	14	S
G	Ls-LP-LT-LA	13	15	S

Concluzii și direcții viitoare

- utilizarea FCA 4-adic pentru detectarea tiparelor comportamentale repetitive;
- utilizarea FCA 5-adic pentru detectarea trendsetter-ilor și a follower-ilor;
- formalizarea noțiunilor definite;
- determinarea schimbărilor tiparelor comportamentale în timp folosind TCA - studiul traiectoriilor;

Schimbarea tiparelor comportamentale în timp

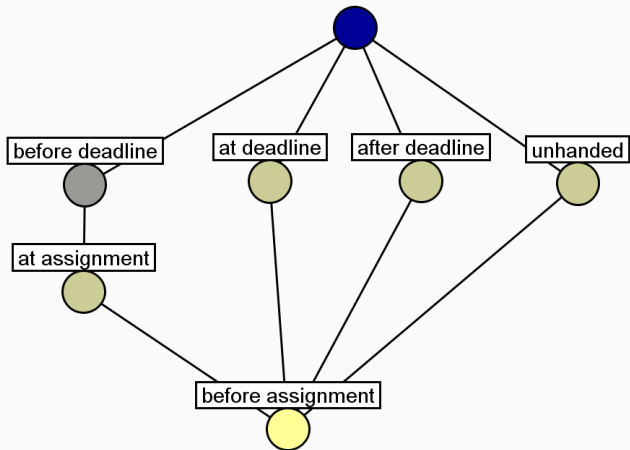


Figure 3: Early Student

Schimbarea tiparelor comportamentale în timp

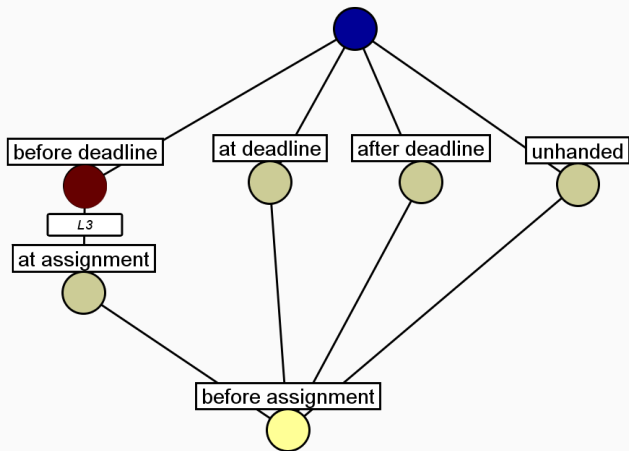


Figure 3: Early Student

Schimbarea tiparelor comportamentale în timp

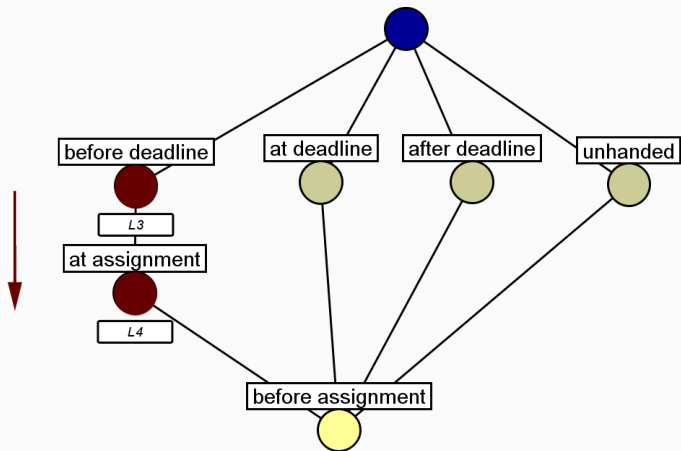


Figure 3: Early Student

Schimbarea tiparelor comportamentale în timp

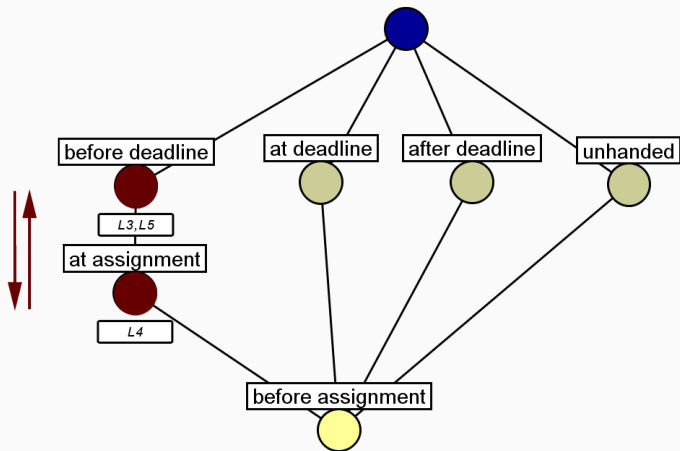


Figure 3: Early Student

Schimbarea tiparelor comportamentale în timp

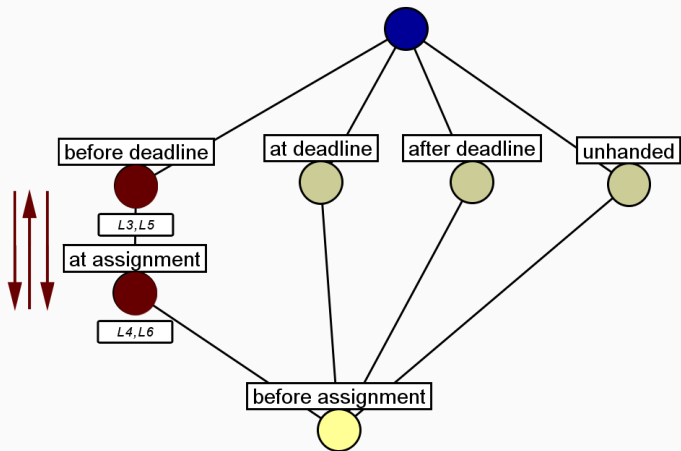


Figure 3: Early Student

Schimbarea tiparelor comportamentale în timp

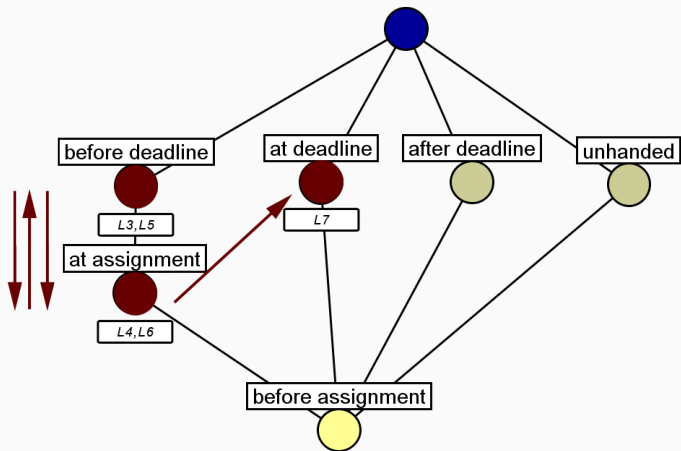


Figure 3: Early Student

Schimbarea tiparelor comportamentale în timp

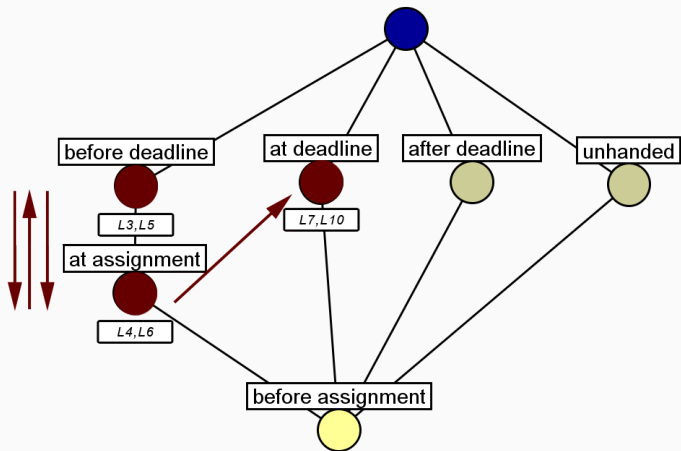


Figure 3: Early Student

Schimbarea tiparelor comportamentale în timp

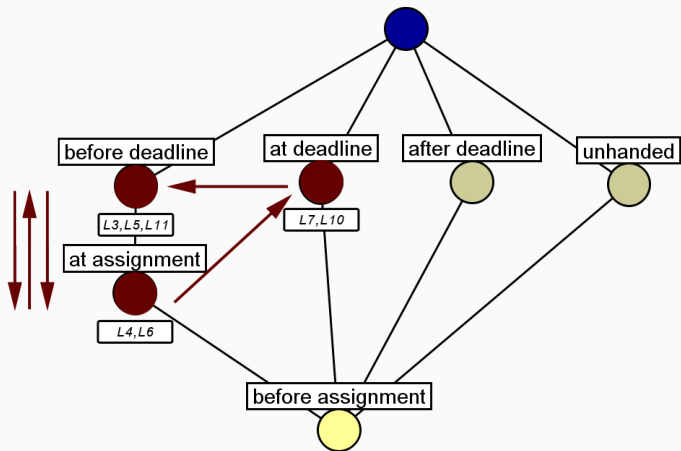


Figure 3: Early Student

Future Work - discover user trails

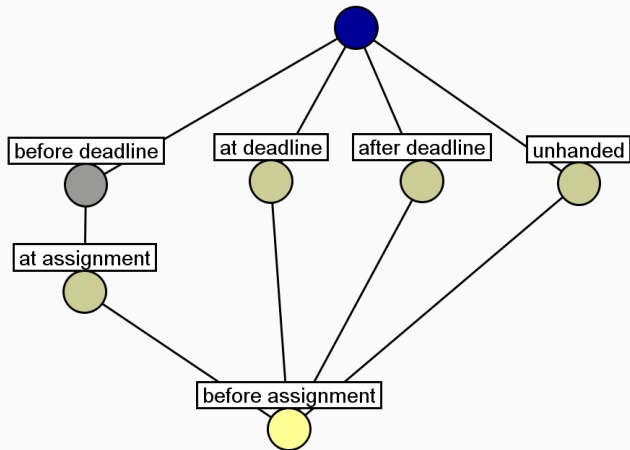


Figure 4: Common Student

Future Work - discover user trails

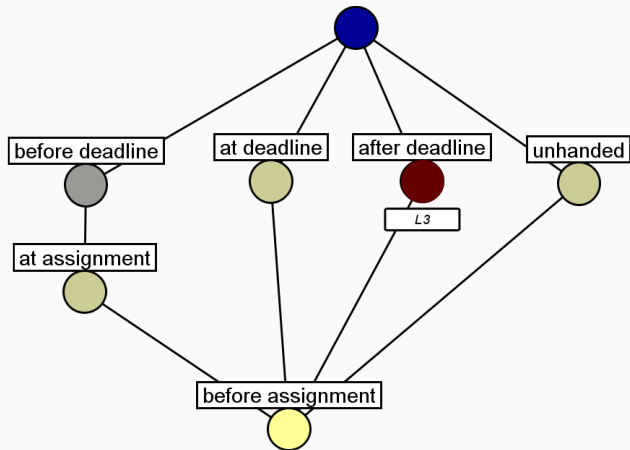


Figure 4: Common Student

Future Work - discover user trails

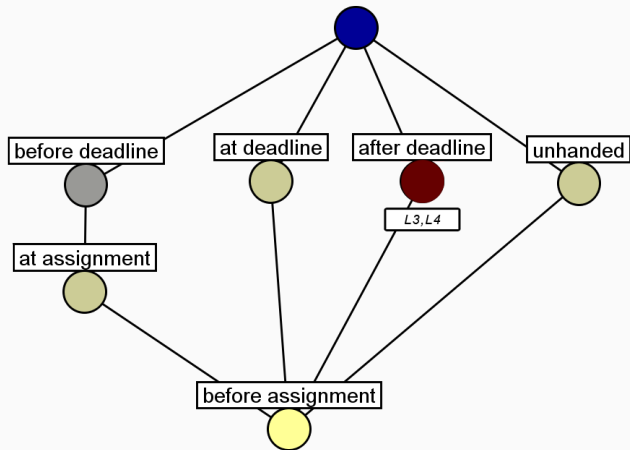


Figure 4: Common Student

Future Work - discover user trails

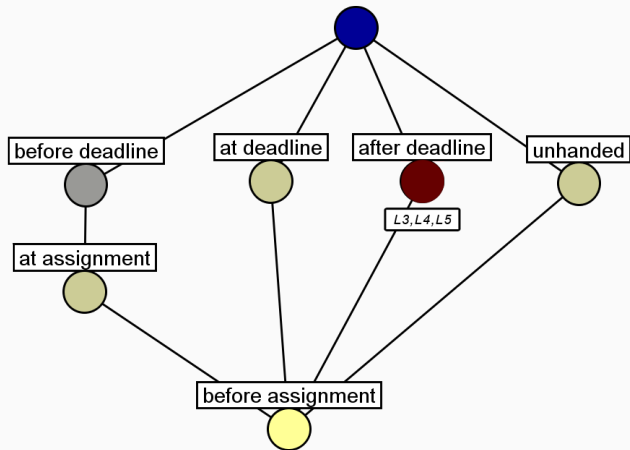


Figure 4: Common Student

Future Work - discover user trails

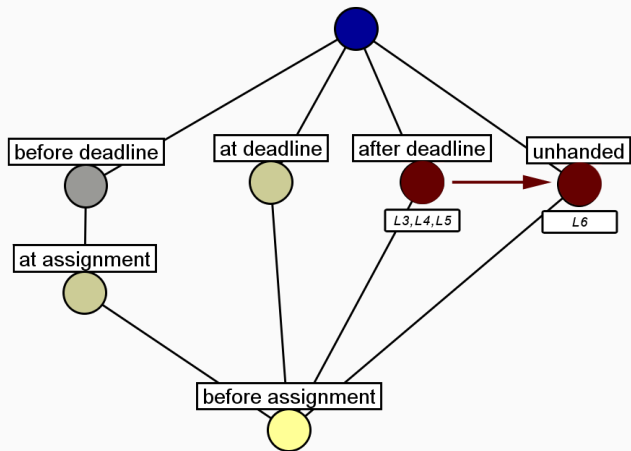


Figure 4: Common Student

Future Work - discover user trails

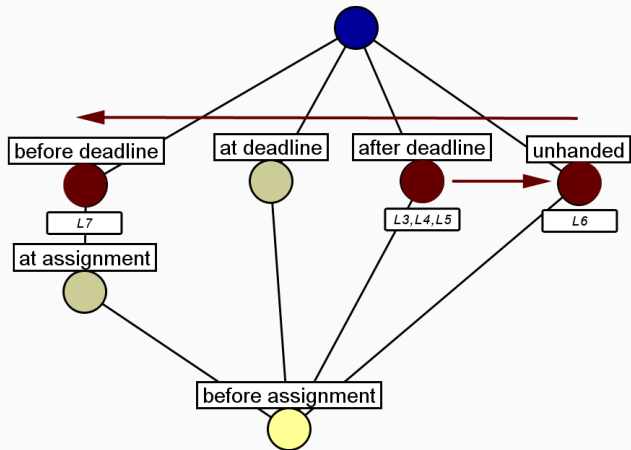


Figure 4: Common Student

Future Work - discover user trails

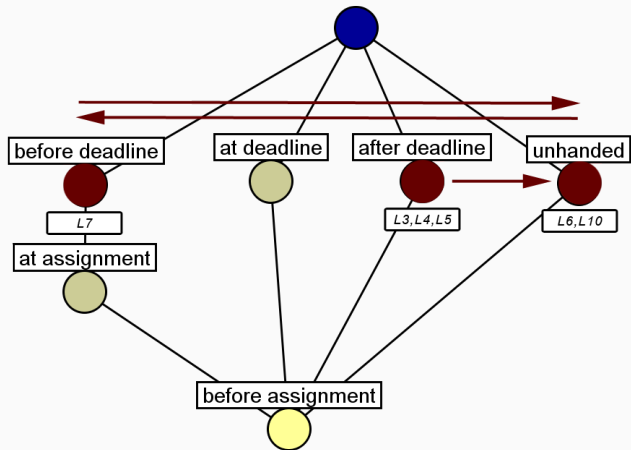


Figure 4: Common Student

Future Work - discover user trails

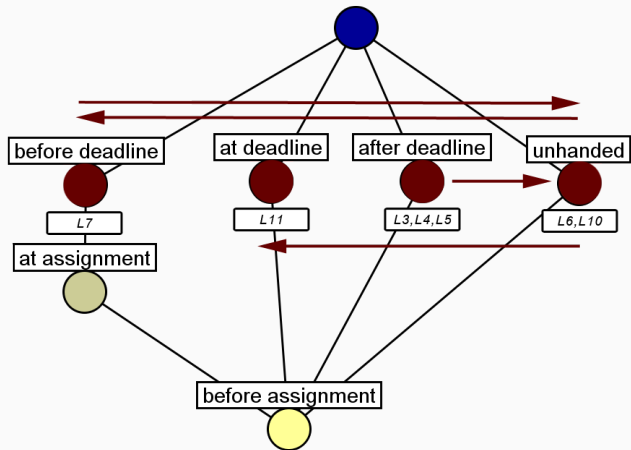


Figure 4: Common Student

Future Work - discover user trails

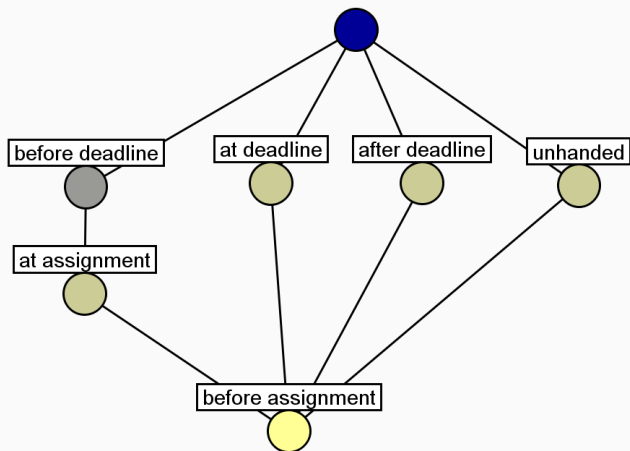


Figure 5: Late Rise Student

Future Work - discover user trails

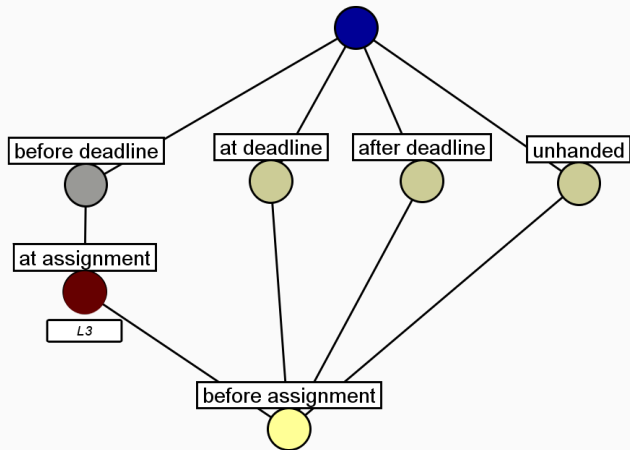


Figure 5: Late Rise Student

Future Work - discover user trails

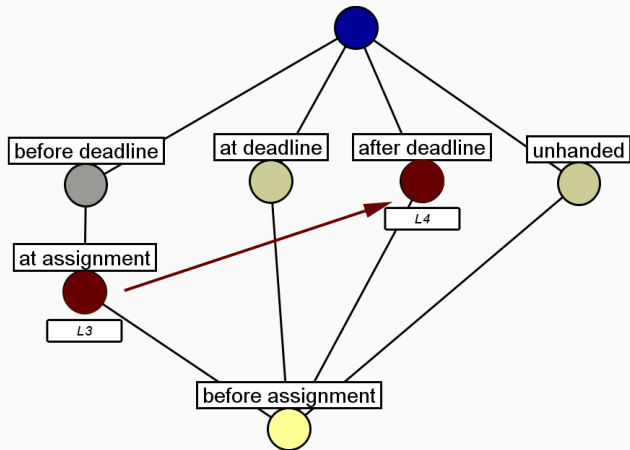


Figure 5: Late Rise Student

Future Work - discover user trails

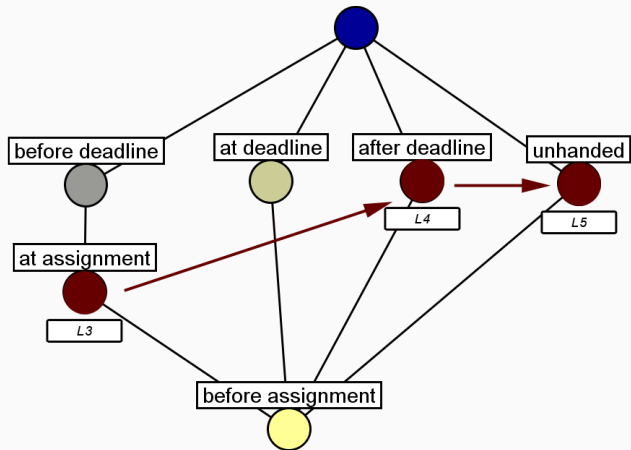


Figure 5: Late Rise Student

Future Work - discover user trails

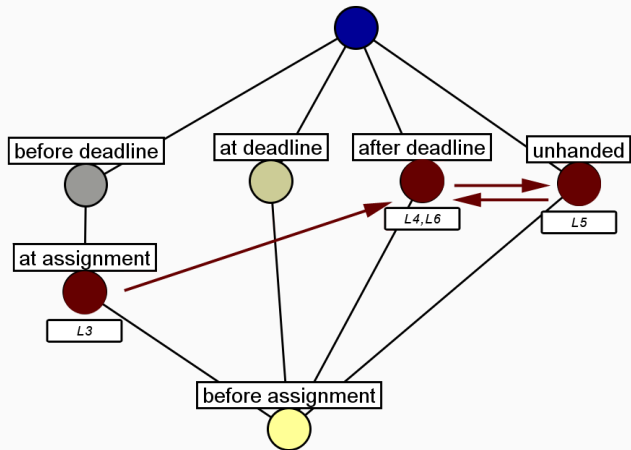


Figure 5: Late Rise Student

Future Work - discover user trails

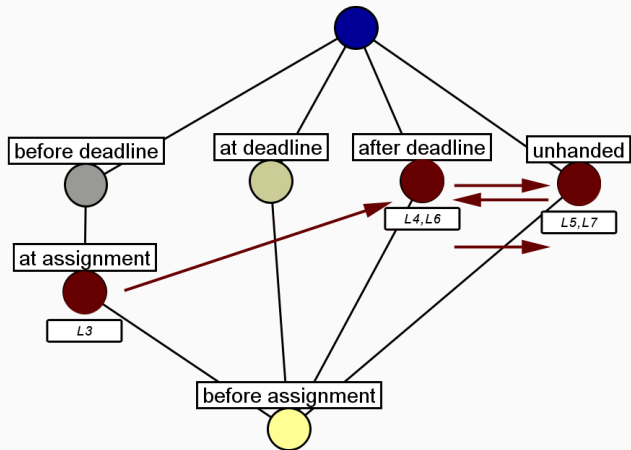


Figure 5: Late Rise Student

Future Work - discover user trails

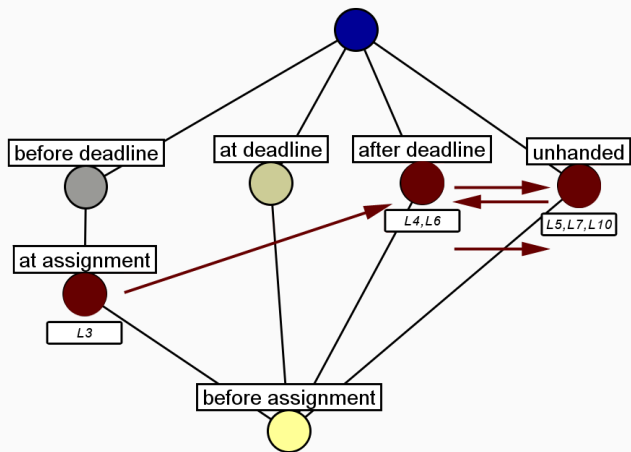


Figure 5: Late Rise Student

Future Work - discover user trails

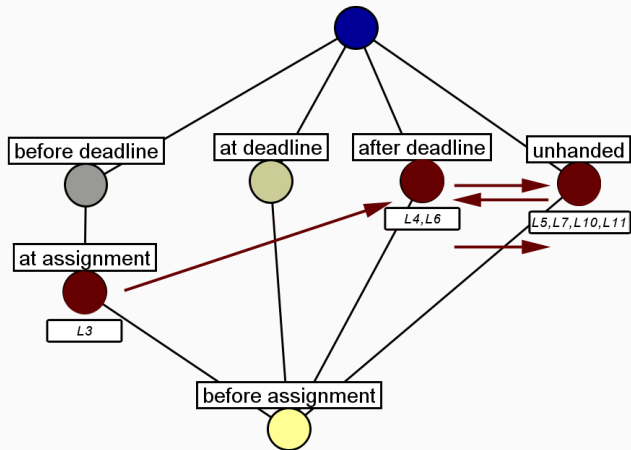


Figure 5: Late Rise Student

Muğumesc