Rozdział 6

R_2U – regułowy system produkcji z niepewnością drugiego rzędu

6.1 Wstęp

Z rozważań zamieszczonych w podrozdziale 5.1 oraz z definicji 5.1 metryki jakości systemu regułowego z niepewnością wynika, że jedną z głównych cech predysponujących ten system do osiągnięcia wysokiej jakości jest zastosowanie do reprezentacji wiedzy metody z dwupunktowym estymatorem stopnia pewności. Spośród metod omawianych w podrozdziale 3.7, żądane kryteria spełnia nie-aksjomatyczny system wnioskowania NARS według Wanga, z regułami w formacie $P \rightarrow C\ \langle p_1;\ p_2 \rangle$, zwanymi przekonaniami. W regule Wanga dwupunktowy estymator stopnia pewności jest zaimplementowany za pomocą parametrów: częstotliwości (p_1), określającego prawdopodobieństwo zajścia konkluzji C przy pewnym zajściu przesłanki P, oraz zaufania (p_2), określającego stopień wiarygodności parametru p_1. Oba parametry mają interpretację częstotliwościową i przyjmują wartości z przedziałów, odpowiednio, $\langle 0; 1 \rangle$ i $(0; 1)$. Sposób implementacji funkcji propagacji niepewności f_{2i} oraz f_{3i} w systemie NARS gwarantuje zróżnicowanie wpływu poszczególnych przekonań na wynik końcowy wnioskowania. Jeśli w procesie wnioskowania przekonanie $P_i \rightarrow C_i\ \langle p_{i1};\ p_{i2} \rangle$ będzie mieć wpływ na postać przekonania końcowego $P_k \rightarrow C_k\ \langle p_{k1};\ p_{k2} \rangle$, to wielkość tego wpływu na częstotliwość p_{k1}będzie tym większa, im większe będzie zaufanie p_{i2}.

System NARS posiada jednak i poważne ograniczenia, które zmniejszają jego praktyczną użyteczność (patrz: podrozdz. 4.3.3). Najważniejsze z nich dotyczą budowy przekonania oraz specyfiki działania silnika wnioskującego.

W regule Wanga zarówno przesłanka, jak i konkluzja mają postać formuł atomowych prostych (niezanegowanych). Ta reguła może wyrażać tylko relację dwuargumentową. Tymczasem większość relacji pomiędzy obiektami, stanami i zjawiskami obserwowalnymi w świecie rzeczywistym ma charakter wieloargumentowy. Do ich wyrażania potrzebne są reguły z przesłankami złożonymi, sformułowanymi w postaci koniunkcji formuł atomowych. Takie reguły odkrywa się w procesie uczenia maszynowego, z danych zgromadzonych w repozytoriach o schemacie atrybutowym płaskim (relacyjne bazy danych) lub strukturalnym (obiektowe bazy danych). Użycie w nich parametru zaufania ($0 < p_2 < 1$) jest w pełni uzasadnione.

Relacje dwuargumentowe (i reguły z przesłanką atomową) dobrze nadają się do modelowania zależności semantycznych, na czele z klasycznymi relacjami hiponimii, hiperonimii i meronimii. Tych relacji nie odkrywa się jednak z danych, lecz zadaje aksjomatycznie, z maksymalną możliwą pewnością, np.

$$\text{Rodzaj\_studiow} = \text{uniw\_II\_st} \rightarrow \text{Tytuł\_zawodowy} = \text{mgr}\ \langle 0.98; 1 \rangle.$$

W takiej sytuacji parametr zaufania ($p_2 = 1$) staje się nadmiarowy.

Podobnie jak w innych systemach regułowych, także i w systemie NARS wszystkie reguły w bazie wiedzy systemu muszą wyrażać zależności dodatnie monotoniczne. Spełnienie tego warunku jest niezbędne dla prawidłowego przebiegu wnioskowania i prawidłowej propagacji niepewności w systemie. Przy tym wymogu, ograniczenie w regułach Wanga postaci formuł atomowych do prostych (niezanegowanych) jeszcze bardziej zawęża zbiór zależności, które da się wyrazić przy ich użyciu. Przykładowo, następująca reguła:

$$\text{Typ\_uczelni} = \text{uniwersytet} \rightarrow \text{Liczba\_kierunkow} = (\leq 40)\ \langle 0.35; 1 \rangle,$$

gdzie symboliczny zapis (≤ 40) oznacza liczbę całkowitą nie większą niż 40, byłaby uznana za niepoprawną. W tej regule zależność pomiędzy przesłanką a konkluzją jest bowiem zależnością ujemną monotoniczną. W wypadku dopuszczenia operatora negacji, można by natomiast zaproponować poprawną regułę:

$$\text{Typ\_uczelni} = \text{uniwersytet} \rightarrow \text{Liczba\_kierunkow} = \neg(\leq 40)\ \langle 0.65; 1 \rangle.$$

Silnik wnioskujący w systemie NARS działa przy założeniu zmiennej zawartości bazy wiedzy. Przyjmuje się, że w trakcie prowadzenia wnioskowania baza ta może być uzupełniana o dowolne reguły Wanga, których treść wynika z nowo udostępnionej ewidencji. W szczególności, możliwa jest sytuacja, w której do bazy wiedzy zawierającej pewną regułę $P \rightarrow C\ \langle p_{11};\ p_{12} \rangle$ dodaje się nową regułę $P \rightarrow C\ \langle p_{21};\ p_{22} \rangle$, z tymi samymi przesłanką i konkluzją. Takie działanie jest uzasadnione – między innymi – w wypadku, gdy system działa w warunkach dynamicznego otoczenia i w trybie rzeczywistym wyciąga wnioski na temat jego zmieniających się parametrów. W opisanej sytuacji należałoby jednak usunąć z bazy wiedzy regułę pierwotną $P \rightarrow C\ \langle p_{11};\ p_{12} \rangle$. Następnie, w trosce o zachowanie wiarygodności systemu, w ślad za regułą $P \rightarrow C\ \langle p_{11};\ p_{12} \rangle$ należałoby usunąć/zmodyfikować w bazie danych systemu wszystkie te fakty, na których postać wpłynęła ta reguła, następnie fakty zależne od tych poprzednich itd., aż do momentu osiągnięcia stanu stabilnego sytemu. Zważywszy nieustaloną (w systemie regułowym) kolejność wykonywania reguł, realizacja tego procesu musiałaby się opierać na algorytmie heurystycznym, a jego złożoność miałaby charakter wykładniczy. Oczywistym jest fakt, że proces ten byłby niemożliwy do przeprowadzenia w reżimie czasu rzeczywistego.

Aby rozwiązać powyższy problem, Wang wzbogacił regułę i fakt o dwa dodatkowe parametry, zwane pilnością i trwałością, które determinują początkowy priorytet reguły/faktu i jej/jego trwałość. Oba parametry mają charakter względny i przyjmują wartości z przedziału $\langle 0; 1\rangle$. Przy tym, o ile częstotliwość i zaufanie stanowią uniwersalne wskaźniki jakości pozyskanej wiedzy, o tyle pilność i trwałość określa się na potrzeby konkretnego wnioskowania, bez związku z tą jakością. Przy uwzględnieniu pilności i trwałości, nie ma potrzeby wyposażania systemu NARS w podsystem utrzymywania wiarygodności. Jeśli jednak system nie posiada właściwości lokalnej zgodności, to te dodatkowe parametry będą wpływać nie tylko na kolejność wykonywania reguł, lecz także na wynik końcowy wnioskowania. Z tego powodu, przedstawione rozwiązanie – choć bardzo wydajne – wydaje się ryzykowne: efektywność działania takiego systemu zależy w znacznym stopniu od samego użytkownika.

Z powyższych rozważań wynika, że system NARS, podobnie jak większość systemów regułowych z niepewnością, jest raczej predestynowany do działania w warunkach statycznych. Do oceny statycznych właściwości obiektu/stanu/sytuacji wystarcza jednak baza wiedzy o stałej, niezmieniającej się w czasie zawartości. Zależnie od potrzeb, można ją aktualizować synchronicznie lub asynchronicznie w okresach bezczynności systemu. Przy takim użyciu systemu subiektywne parametry pilności i trwałości reguł stałyby się zbędne, a jakość odpowiedzi systemu zależałaby wyłącznie od jakości jego bazy wiedzy i działania silnika wnioskującego.

Zgodnie ze wzorem 4.11, jakość bazy wiedzy systemu regułowego z niepewności zależy od liczby reguł w bazie wiedzy i od średniego poziomu wiarygodności tych reguł. Choć w obu wypadkach ta zależność ma charakter dodatni monotoniczny, odpowiednio logarytmiczny i liniowy, to jednak powiększanie bazy wiedzy nie może się odbywać w sposób bezwzględny. W szczególności, ze względu na wymóg niesprzeczności, baza wiedzy nie powinna zawierać dwóch różnych reguł z tą samą przesłanką i tą samą konkluzją, np. $P \to C\ \langle p_{11};\ p_{12}\rangle$ i $P \to C\ \langle p_{21};\ p_{22}\rangle$. Takie reguły należy scalić w jedną na etapie projektowania bazy. Jeśli połączenie zostanie wykonane przy użyciu reguły rewizji (3.9), to parametr zaufania w wypadkowej regule $P \to C\ \langle p_1;\ p_2\rangle$ będzie miał wartość większą od wartości parametrów zaufania w obu regułach cząstkowych $(p_2 > p_{12}, p_2 > p_{22})$.

Zapewnienie niesprzeczności bazy wiedzy systemu nie jest równoznaczne z nadaniem systemowi właściwości lokalnej zgodności. Przy rezygnacji z posługiwania się subiektywnymi parametrami reguł – pilnością i trwałością, na jakość odpowiedzi systemu wpływałby nowy algorytm rozstrzygania konfliktów w agendzie. W celu zwiększenia stopnia poprawności odpowiedzi, jego działanie należałoby oprzeć na regułach cechujących się najwyższą jakością, wyrażaną w tym wypadku za pomocą parametru zaufania.

6.2 Reguła 2U - składnia i semantyka

W pracy [79] zaproponowano model systemu regułowego z niepewnością oparty na formacie reguły z niepewnością drugiego rzędu, zwanej w skrócie regułą 2U (ang. *rule with 2nd order uncertainty*). Stanowi on wynik poszukiwań takiej metody reprezentacji wiedzy niepewnej i takiego algorytmu wnioskowania przybliżonego, które w połączeniu gwarantują:

- konstrukcję systemu cechującego się wysoką jakością,
- możliwość pozyskiwania wiedzy do bazy wiedzy systemu w sposób półautomatyczny, z danych zapisanych w popularnym formacie atrybutowym.

W powyższym modelu zaadaptowano niektóre z rozwiązań zastosowanych przez Wanga w systemie NARS. Wprowadzone modyfikacje miały na celu zniesienie pewnych ograniczeń tego systemu, zarówno tych dotyczących efektywności i wydajności działania systemu, jak i tych, które zawężają obszar jego zastosowań.

Niech $\mathcal{D}_u$ oznacza dziedzinę problemową, dla której należy zaprojektować system regułowy z niepewnością, $\mathcal{T}_{\mathcal{D}u}$ – wybraną konceptualizację tej dziedziny, a $C_{\mathcal{TD}u} = \{\mathrm{C}_{\mathrm{TDu}1}, \mathrm{C}_{\mathrm{TDu}2}, \dots, \mathrm{C}_{\mathrm{TDu}n}\}$ – skończony zbiór kategorii użytych w konceptualizacji $\mathcal{T}_{\mathcal{D}u}$. Powyższym kategoriom nadano unikalne nazwy, zwane atrybutami zbiorowymi, pochodzące ze zbioru $\mathbf{A_S} = \{\mathrm{A_{s1}}, \mathrm{A_{s2}}, \dots, \mathrm{A}_{sn}\}$.

Definicja 6.1. Schematem danych w dziedzinie $\mathcal{D}_u$, zgodnym z konceptualizacją $\mathcal{T}_{\mathcal{D}u}$, nazywa się dowolny niepusty podzbiór $\mathbf{S_S} = \{\mathrm{A}_{si1}, \mathrm{A}_{si2}, \dots, \mathrm{A}_{sim}\}$ zbioru atrybutów zbiorowych $\mathbf{A_S}$.

Definicja 6.2. Klasyczną formułą atomową z atrybutem zbiorowym A_{si} nazywa się dowolne wyrażenie w postaci:

$$\mathrm{A}_{si} \in \{v_{i1}, v_{i2}, \dots, v_{ik}\}, \tag{6.1}$$

$$\neg(\mathrm{A}_{si} \in \{v_{i1}, v_{i2}, \dots, v_{ik}\}), \tag{6.2}$$

$$\{v_{i1}, v_{i2}, \dots, v_{ik}\} \subseteq \mathrm{A}_{si}, \tag{6.3}$$

$$\text{lub } \neg(\{v_{i1}, v_{i2}, \dots, v_{ik}\} \subseteq \mathrm{A}_{si}), \tag{6.4}$$

gdzie $0 \leq k \leq |\mathrm{C}_{\mathrm{TDu}i}|$, a v_{i1}, v_{i2}, …, v_{ik} oznaczają parami różne wartości – przykłady kategorii $\mathrm{C}_{\mathrm{TDu}i}$. Dla ujednolicenia zapisu wyrażeń 6.1 – 6.4, przyjęto dla nich wspólną notację:

$$\mathrm{A}_{si} \blacksquare \{v_{i1}, vi_2, \dots, v_{ik}\} qlr, \tag{6.5}$$

w której ■ oznacza symbol relacyjny = lub $\neq$, a qlr – kwalifikator $\odot$ lub $\oplus$. Powyższy kwalifikator nadaje zbiorowi $\{v_{i1}, v_{i2}, \dots, v_{ik}\}$ interpretację, odpowiednio koniunkcyjną lub dysjunkcyjną. Znaczenie formuły (6.5) dla poszczególnych kombinacji wartości ■ i qlr jest przy tym następujące: $\oplus$ i =

– wyrażenie 6.1, $\oplus$ i $\neq$ – wyrażenie 6.2, $\odot$ i $=$ – wyrażenie 6.3, $\odot$ i $\neq$ – wyrażenie 6.4. Zbiór wartości z kwalifikatorem $\{v_{i1}, v_{i2}, \ldots, v_{ik}\}qlr$ nazywa się w dalszym ciągu zbiorem kwalifikowanym i stosuje dla niego oznaczenia vq, vq_1, vq_2, …, vq_i, itd.

Definicja 6.3. Regułą 2U opartą na schemacie danych $\mathbf{S_S} = \{A_{si1}, A_{si2}, \ldots, A_{sim}\}$ nazywa się dowolne wyrażenie w postaci:

$$\text{it is declared with } \mathrm{grf}(p_r) : P_1, P_2, \ldots, P_k \rightarrow C \text{ with } \mathrm{irf}(p_c)\,, \tag{6.6}$$

gdzie $0 \leq k < m$, P_1, P_2, …, P_k, C oznaczają klasyczne formuły atomowe z parami różnymi atrybutami ze zbioru $\{A_{si1}, A_{si2}, \ldots, A_{sim}\}$, irf oraz grf – nazwy współczynników wiarygodności, odpowiednio wewnętrznego i zewnętrznego, a p_c i p_r – wartości współczynników, odpowiednio irf i grf, pochodzące z przedziału $(0;\ 1\rangle$. W dalszym tekście regułę 2U zapisuje się w skróconej postaci:

$$\mathrm{grf}(p_r) : P_1, P_2, \ldots, P_k \rightarrow C : \mathrm{irf}(p_c)\,. \tag{6.7}$$

Reguła wyraża dodatnią monotoniczną zależność między przesłanką $P_1 \wedge P_2 \wedge \ldots \wedge P_k$ a konkluzją C. Użyte w regule współczynniki wiarygodności irf i grf oznaczają, odpowiednio: stopień pewności konkluzji C przy założeniu 100% pewności przesłanki $P_1 \wedge P_2 \wedge \ldots \wedge P_k$, oraz stopień pewności (wiarygodności) reguły jako całości w kontekście dostępnej ewidencji, czyli zbioru dostępnych danych z dziedziny $\mathcal{D}_u$, zbudowanych według schematu $\mathbf{S_S} = \{A_{si1}, A_{si2}, \ldots, A_{sim}\}$. Współczynnik irf służy w regule 2U do oznaczenia poziomu niepewności pierwszego rzędu, a współczynnik grf – poziomu niepewności drugiego rzędu. Współczynniki irf i grf stanowią odpowiedniki częstotliwości i zaufania w regule Wanga.

Definicja 6.4. Faktem 2U opartym na schemacie danych $\mathbf{S_S} = \{A_{si1}, A_{si2}, \ldots, A_{sim}\}$ nazywa się dowolną klasyczną formułę atomową z atrybutem zbiorowym ze zbioru $\{A_{si1}, A_{si2}, \ldots, A_{sim}\}$. Dla ujednolicenia rozważań fakt ten można wyrazić jako regułę 2U w postaci:

$$\mathrm{grf}(p_r) : \text{true} \rightarrow A_{sij} \blacksquare \{v_{ij1}, v_{ij2}, \ldots, v_{ijk}\}qlr : \mathrm{irf}(p_c) \tag{6.8}$$

ze stałą logiczną true w roli przesłanki i rozważaną formułą w roli konkluzji, gdzie $1 \leq j \leq m$, a pozostałe symbole zachowują znaczenia z definicji 6.2 i 6.3.

Zakłada się, że reguły 2U będą pozyskiwane z danych zgromadzonych na drodze doświadczalnej i zapisanych w formacie atrybutowym (patrz podrozdz. 7.2.1).

Reguła 2U, choć posługuje się do oznaczenia stopnia pewności podobnymi parametrami jak reguła Wanga, ma większą od niej siłę wyrazu. Zapewnia ją:

– używanie zbiorów wartości w miejsce wartości pojedynczych,
– używanie w formułach atomowych dwóch operatorów relacyjnych $\in$ i $\subseteq$ w miejsce jednego równościowego $=$,
– dopuszczenie w formułach atomowych operatora negacji $\neg$,

– użycie w przesłance reguły koniunkcji formuł atomowych w miejsce pojedynczej formuły atomowej.

Posługiwanie się w klasycznych formułach atomowych zbiorami kwalifikowanymi powoduje, że te formuły stają się złożone. Kwalifikator zbioru $\odot$ oznacza, że rozważany atrybut przyjmuje równolegle wszystkie wartości z kwalifikowanego zbioru (stwierdzenie równoważne koniunkcji formuł elementarnych), a kwalifikator zbioru $\oplus$ – że przyjmuje on jedną z wartości wymienionych w kwalifikowanym zbiorze (stwierdzenie równoważne dysjunkcji formuł elementarnych). Podczas gdy formuła zapisana z użyciem kwalifikatora $\oplus$ przypomina hipotezę z teorii Dempstera-Shafera, to formuła zapisana z użyciem kwalifikatora $\odot$ ma charakter nowatorski. Szersze uzasadnienie potrzeby posługiwania się oboma kwalifikatorami można znaleźć w rozdziale 9, prezentującym postulaty dotyczące budowy medycznych systemów eksperckich.

Zastosowanie zbiorów kwalifikowanych jest równoznaczne z przyjęciem za zbiór operatorów relacyjnych – zbioru dwuelementowego $\{\in, \supseteq\}$. Zbiór ten jest podzbiorem właściwym zbioru operatorów relacyjnych z logiki ALSV(FD) (ang. *Attributive Logic with Set Values*) [105]. Większość operatorów nieuwzględnionych w formacie reguły 2U jest trudno implementowalna (np. operator $\sim$) lub słabo aplikowalna (np. operator $=$).

O operatorze negacji i jego wpływie na siłę wyrazu reguły była mowa w podrozdziale 6.1. Z drugiej strony, szczególną rolę tego operatora należy rozpatrywać w kontekście założeń przyjmowanych odnośnie do zasad wnioskowania w systemie typu $\boldsymbol{R\_2U}$ – regułowym systemie produkcji z regułami 2U (patrz podrozdz. 6.4).

W zamyśle, także parametr wiarygodności grf ma głębsze znaczenie od parametru zaufania w regule Wanga. Przyjmuje się, że grf może uwzględniać różne aspekty ewidencji, z której pozyskano regułę, oprócz wielkości ewidencji – także stopień jej homogeniczności, czy źródło pochodzenia. Zgodnie z planem, parametr ten ma odgrywać ważną rolę podczas wnioskowania w systemach z regułami 2U. W szczególności, w procesach wnioskowania w przód może służyć do rozstrzygania konfliktów w agendzie.

6.3 Baza wiedzy i baza danych systemu R_2U

6.3.1. Projektowanie bazy wiedzy

Zbiór reguł 2U pozyskanych ze zbioru danych atrybutowych o schemacie $\mathbf{S_S}$ (def. 6.1) stanowi punkt wyjścia do konstrukcji bazy wiedzy systemu $\boldsymbol{R\_2U}$ dedykowanego dziedzinie $\mathcal{D}_u$.

Definicja 6.5. Bazą wiedzy regułowego systemu produkcji $\boldsymbol{R\_2U}$ dla dziedziny $\mathcal{D}_u$ charakteryzowanej zbiorem atrybutów $\mathbf{A_S} = \{\mathrm{A}_{s1}, \mathrm{A}_{s2}, \dots, \mathrm{A}_{sn}\}$ nazywamy dowolny niepusty zbiór reguł 2U opartych na pewnym schemacie $\mathbf{S_{Si}} = \{A_{si1}, A_{si2}, \dots, A_{sim}\} \subseteq \mathbf{A_S}$ i spełniających obowiązkowe wymagania jakościowe (patrz str. 80). W celu zagwarantowania tych wymagań należy:

- zweryfikować dla każdej reguły ze zbioru charakter zależności łączącej jej przesłankę i konkluzję: jeśli zależność nie ma charakteru monotonicznego, to taką regułę należy usunąć ze zbioru; jeśli ma charakter monotoniczny ujemny, to regułę należy „odwrócić” poprzez zanegowanie konkluzji i odpowiednią zmianę parametrów liczbowych; jeśli ma charakter monotoniczny dodatni, to regułę należy pozostawić w zbiorze w niezmienionej postaci;
- usunąć ze zbioru pary reguł sprzecznych; ewentualnie, po konsultacji z ekspertem – reguły uznane za poprawne przywrócić do zbioru;
- usunąć ze zbioru reguły, dla których można wskazać w zbiorze reguły subsumujące.

Weryfikacja charakteru zależności pomiędzy przesłanką a konkluzją reguły jest problemem ogólnym, dotyczącym każdego regułowego systemu z niepewnością. Podstawowa trudność weryfikacji polega na tym, że należy ją prowadzić globalnie, na podstawie równoczesnej obserwacji całej bazy wiedzy. Szczegółowy przebieg weryfikacji zależy od formatu reguły z niepewnością, czyli od liczby i znaczenia parametrów liczbowych definiujących stopień pewności.

Definicja 6.6. Niech $\leq_{AS}$ oznacza relację zdefiniowaną na zbiorze klasycznych formuł atomowych z atrybutami zbiorowymi ze zbioru $\mathbf{A_S}$. Formuła P_i pozostaje w relacji $\leq_{AS}$ z formułą P_j wtedy i tylko wtedy, gdy:

- $P_i \stackrel{\text{def}}{=} (A_{si} \in \{v_{i1}, v_{i2}, \dots, v_{iki}\})$ i $P_j \stackrel{\text{def}}{=} \left(A_{sj} \in \left\{v_{j1}, v_{j2}, \dots, v_{jkj}\right\}\right)$ i $A_{si} = A_{sj}$ i $\left\{v_{j1}, v_{j2}, \dots, v_{jkj}\right\} \subseteq \{v_{i1}, v_{i2}, \dots, v_{iki}\}$, lub
- $P_i \stackrel{\text{def}}{=} \neg(A_{si} \in \{v_{i1}, v_{i2}, \dots, v_{iki}\})$ i $P_j \stackrel{\text{def}}{=} \neg\left(A_{sj} \in \left\{v_{j1}, v_{j2}, \dots, v_{jkj}\right\}\right)$ i $A_{si} = A_{sj}$ i $\{v_{i1}, v_{i2}, \dots, v_{iki}\} \subseteq \left\{v_{j1}, v_{j2}, \dots, v_{jkj}\right\}$, lub
- $P_i \stackrel{\text{def}}{=} (\{v_{i1}, v_{i2}, \dots, v_{iki}\} \subseteq A_{si})$ i $P_j \stackrel{\text{def}}{=} \left(\left\{v_{j1}, v_{j2}, \dots, v_{jkj}\right\} \subseteq A_{sj}\right)$ i $A_{si} = A_{sj}$ i $\{v_{i1}, v_{i2}, \dots, v_{iki}\} \subseteq \left\{v_{j1}, v_{j2}, \dots, v_{jkj}\right\}$, lub
- $P_i \stackrel{\text{def}}{=} \neg(\{v_{i1}, v_{i2}, \dots, v_{iki}\} \subseteq A_{si})$ i $P_j \stackrel{\text{def}}{=} \neg\left(\left\{v_{j1}, v_{j2}, \dots, v_{jkj}\right\} \subseteq A_{sj}\right)$ i $A_{si} = A_{sj}$ i $\left\{v_{j1}, v_{j2}, \dots, v_{jkj}\right\} \subseteq \{v_{i1}, v_{i2}, \dots, v_{iki}\}$, lub
- $P_i \stackrel{\text{def}}{=} (A_{si} \in \{v_{i1}\})$ i $P_j \stackrel{\text{def}}{=} \left(\left\{v_{j1}\right\} \subseteq A_{sj}\right)$ i $A_{si} = A_{sj}$ i $v_{i1} = v_{j1}$, lub
- $P_i \stackrel{\text{def}}{=} \neg(\{v_{i1}\} \subseteq A_{si})$ i $P_j \stackrel{\text{def}}{=} \neg\left(A_{sj} \in \left\{v_{j1}\right\}\right)$ i $A_{si} = A_{sj}$ i $v_{i1} = v_{j1}$ lub
- $\exists_{P_k} (P_i \leq_{AS} P_k) \land \left(P_k \leq_{AS} P_j\right)$.

Twierdzenie 6.1. Relacja $\leq_{AS}$ jest relacją porządku częściowego.

Dowód. Z definicji 6.6 jasno wynika, że relacja $\leq_{\mathrm{AS}}$ jest zwrotna, antysymetryczna i przechodnia. Tym samym, ma wszystkie własności relacji porządku częściowego.

Twierdzenie 6.2. Niech R_i i R_j oznaczają dwie reguły 2U oparte na schemacie danych $\mathbf{S_S} = \{\mathrm{A}_{si1}, \mathrm{A}_{si2}, \dots, \mathrm{A}_{sim}\}$, zredagowane w postaci:

$$R_i \stackrel{\text{def}}{=} \mathrm{grf}(p_{ri}) : P_{i1}, P_{i2}, \dots, P_{ihi} \rightarrow C_i : \mathrm{irf}(p_{ci}) \tag{6.9}$$

$$\text{i } R_j \stackrel{\text{def}}{=} \mathrm{grf}(p_{rj}) : P_{j1}, P_{j2}, \dots, P_{jhj} \rightarrow C_j : \mathrm{irf}(p_{cj}) \tag{6.10}$$

Jeśli zachodzi którakolwiek z poniższych zależności:

– $\exists_{1 \leq k \leq hi, 1 \leq l \leq hj}(P_{ik} = P_{jl}) \wedge (C_j = \neg C_i)$, (6.11)

– $\exists_{1 \leq k \leq hi, 1 \leq l \leq hj}(P_{jl} = \neg P_{ik}) \wedge (C_i = C_j)$, (6.12)

– $\forall_{1 \leq k \leq hi} \exists_{1 \leq l \leq hj} \ (P_{ik} \leq_{\mathrm{AS}} P_{jl}) \wedge (C_j \leq_{\mathrm{AS}} C_i) \wedge (p_{ci} > p_{cj})$, (6.13)

– $\forall_{1 \leq k \leq hi} \exists_{1 \leq l \leq hj} \left((P_{jl} \leq_{\mathrm{AS}} P_{ik}) \wedge (C_i \leq_{\mathrm{AS}} C_j) \wedge (p_{cj} > p_{ci})\right)$, (6.14)

to przynajmniej jedna z reguł R_i i R_j jest zbudowana nieprawidłowo – nie wyraża zależności dodatniej monotonicznej.

W zapisach reguł, P_{ik} $(1 \leq k \leq h_i)$, C_i, P_{jl} $(1 \leq l \leq h_j)$ i C_j oznaczają klasyczne formuły atomowe z atrybutami zbiorowymi ze zbioru $\mathbf{S_S}$, p_{ci} i p_{cj} – wartości współczynników irf, a p_{ri} i p_{rj} – wartości współczynników grf.

Dowód. Jeśli spełnienie formuły klasycznej zawartej w przesłance reguły wpływa dodatnio na spełnienie konkluzji w tej regule, to znaczy, że zależność między tą formułą a konkluzją reguły jest dodatnia monotoniczna, a zależność między tą formułą a negacją konkluzji jest ujemna monotoniczna (6.11). W opisanej sytuacji, ujemny monotoniczny charakter ma także zależność między negacją tej formuły a konkluzją (6.12).

Ponadto, jeśli wzmocnienie przesłanki reguły przez wzmocnienie dowolnej z jej formuł klasycznych lub przez dodanie nowej formuły klasycznej wpływa ujemnie na spełnienie konkluzji tej reguły, to znaczy, że:

– zależność pomiędzy przesłanką reguły a jej konkluzją nie była dodatnia monotoniczna, lub
– warunek wzmacniający przesłankę niszczy istniejącą zależność dodatnią monotoniczną (6.13 i 6.14).

W celu zweryfikowania poprawności budowy wskazanej reguły R_i należy wziąć pod uwagę wszystkie te reguły R_j z wyjściowego zbioru reguł, które pozostają z nią w nieprawidłowej zależności 6.11– 6.14. Ostateczną decyzję można podjąć na podstawie procentowego udziału takich reguł w ogólnej liczbie wszystkich reguł ze zbioru.

Definicja 6.7. Niech R_i i R_j oznaczają dwie reguły 2U oparte na schemacie danych $\mathbf{S_S} = \{\mathrm{A}_{si1}, \mathrm{A}_{si2}, \dots, \mathrm{A}_{sim}\}$ i zredagowane w postaci, odpowiednio 6.9 i 6.10. Reguły R_i i R_j pozostają w sprzeczności wtedy i tylko wtedy, gdy:

– zbiory formuł atomowych $\{P_{i1}, P_{i2}, \dots, P_{ihi}\}$ i $\{P_{j1}, P_{j2}, \dots, P_{jhj}\}$ są identyczne oraz

– $(C_i \leq_{\mathrm{AS}} C_j$ i $p_{ci} < p_{cj})$ lub $(C_j \leq_{\mathrm{AS}} C_i$ i $p_{cj} < p_{ci})$.

Sformułowany w definicji 6.7 warunek sprzeczności reguł 2U jest przypadkiem szczególnym zależności 6.13 lub 6.14. Prawidłowe przeprowadzenie procesu eliminacji reguł niewyrażających zależności dodatniej monotonicznej daje więc gwarancję niewystępowania w bazie wiedzy jakiejkolwiek pary reguły sprzecznych.

Ostatnią ważną czynnością projektową jest oczyszczanie bazy wiedzy z reguł nadmiarowych. Choć nie jest to czynność obligatoryjna, to jednak warto ją przeprowadzić z uwagi na większą wydajność późniejszych procesów wnioskowania, a także – z uwagi na możliwość szybkiej oceny jakości bazy wiedzy systemu.

Definicja 6.8. Niech R_i i R_j oznaczają dwie reguły 2U oparte na schemacie danych $\mathbf{S_S} = \{\mathrm{A}_{si1}, \mathrm{A}_{si2}, \dots, \mathrm{A}_{sim}\}$ i zredagowane w postaci, odpowiednio 6.9 i 6.10. Reguła R_i subsumuje regułę R_j wtedy i tylko wtedy, gdy:

– zbiory formuł atomowych $\{P_{i1}, P_{i2}, \dots, P_{ihi}\}$ i $\{P_{j1}, P_{j2}, \dots, P_{jhj}\}$ nie są identyczne,

– $\forall_{1 \leq k \leq hi} \exists_{1 \leq l \leq hj} (P_{ik} \leq_{\mathrm{AS}} P_{jl})$,

– $C_i = C_j$ i $p_{ci} = p_{cj}$
oraz

– $p_{ri} \geq p_{rj}$.

Po zweryfikowaniu zależności między przesłankami i konkluzjami reguł oraz usunięciu reguł niewyrażających zależności dodatniej monotonicznej i reguł nadmiarowych, bazę wiedzy systemu $\boldsymbol{R\_2U}$ można poddać ocenie za pomocą metryki jakości 4.11. Funkcja avg przyjmuje w tym wypadku postać średniej arytmetycznej z wartości współczynników grf wszystkich reguł.

Przykładowy algorytm konstrukcji bazy wiedzy systemu $\boldsymbol{R\_2U}$ ze zbiorczych danych atrybutowych zamieszczono w podrozdziale 7.3.

6.3.2. Obsługa bazy danych

Baza danych systemu $\boldsymbol{R\_2U}$ jest zbiorem faktów w postaci 6.8. Nowy stan bazy danych jest ustalany na początku każdego procesu wnioskowania, tak progresywnego, jak i regresywnego. W wypadku wnioskowania progresywnego, stan ten będzie podlegał obowiązkowym aktualizacjom w trakcie prowadzenia wnioskowania i opcjonalnej normalizacji po jego zakończeniu.

Zgodnie z zaleceniem sformułowanym w podrozdziale 4.4, w bazie danych systemu $\boldsymbol{R\_2U}$ będą przechowywane tylko fakty z konkluzjami w formie prostej, czyli niezanegowanej. Za spełnienie tego wymogu odpowiadają algorytmy wnioskowania progresywnego i regresywnego, w szczególności – funkcje propagacji niepewności (patrz: podrozdział 6.4). Pomimo takiej gwarancji, stan bazy danych w procesie wnioskowania progresywnego powinien być przez cały czas monitorowany przez podsystem utrzymywania wiarygodności. Ta potrzeba wynika ze specyfiki formuł atomowych w faktach/regułach 2U, używających wartości zbiorowych w miejsce tradycyjnych wartości pojedynczych. Podobnie jak nie należy dopuszczać do występowania w bazie wiedzy par reguł sprzecznych, tak też nie należy dopuszczać do równoczesnego występowania w bazie danych par faktów sprzecznych.

Definicja 6.9. Niech F_i i F_j oznaczają dwa fakty 2U oparte na schemacie danych $\mathbf{S_S} = \{\mathrm{A}_{si1}, \mathrm{A}_{si2}, \dots, \mathrm{A}_{sim}\}$ i zredagowane w postaci:

$$F_i \stackrel{\text{def}}{=} \mathrm{grf}(p_{ri}) : \mathrm{true} \rightarrow C_i : \mathrm{irf}(p_{ci}) \quad (6.15)$$

$$\text{i } F_j \stackrel{\text{def}}{=} \mathrm{grf}(p_{rj}) : \mathrm{true} \rightarrow C_j : \mathrm{irf}(p_{cj}) \,. \quad (6.16)$$

gdzie C_i i C_j oznaczają klasyczne formuły atomowe z atrybutami zbiorowymi zapisane w formie prostej (niezanegowanej).

Fakty F_i i F_j pozostają w sprzeczności wtedy i tylko wtedy, gdy zachodzą zależności:

- $C_i \leq_{\mathrm{AS}} C_j$ i $p_{ci} < p_{cj}$, lub
- $C_j \leq_{\mathrm{AS}} C_i$ i $p_{ci} > p_{cj}$.

Wykrycie takiej sprzeczności powinno uruchamiać natychmiastowy proces jej eliminacji. Może on być wykonywany automatycznie lub pół-automatycznie, przy współudziale użytkownika. Podstawą rozstrzygnięć w procesie automatycznym może być wartość współczynnika grf: z dwóch znajdujących się w bazie danych i pozostających w sprzeczności faktów 2U należy raczej usunąć ten, który ma mniejszą wartość tego współczynnika.

6.4 Wnioskowanie w systemie R_2U

6.4.1 Założenia

Wnioskowanie w systemie $\boldsymbol{R\_2U}$ opiera się na kilku istotnych założeniach. Przede wszystkim, przyjmuje się, że będzie ono prowadzone przy założeniu otwartego świata OWA.

Kolejne założenia są związane z interpretacją parametrów liczbowych używanych w regule/fakcie 2U: współczynnika wiarygodności irf i współczynnika wiarygodności grf. W istocie parametry te odnoszą się do różnych właściwości ewidencji, z której pozyskano regułę. Dlatego nic nie stoi na przeszkodzie, by silnik wnioskujący systemu używał ich niezależnie od siebie, gdy zachodzi potrzeba. Taką odrębną, ważną rolę przewidziano w systemie dla współczynnika grf. W procesie wnioskowania w przód będzie on sterował procesem konstrukcji agendy, lecz także – brał aktywny udział przy rozstrzyganiu konfliktów w agendzie.

Ze względu na planowaną pół-automatyczną akwizycję wiedzy ze zbioru dostępnych danych należy zakładać, że p_c i p_r będą wyrażać oszacowanie, a nie dokładne wartości współczynników irf i grf reguły w postaci 6.7. Z drugiej strony, te wartości muszą pochodzić z pewnych ustalonych zakresów $\langle Bel_c;\ Pls_c \rangle$ i $\langle Bel_r\ ;\ Pls_r \rangle$ (patrz podrozdz. 3.5.2). Oczywiście, takim samym ograniczeniom podlegają wartości współczynników we wszystkich faktach 2U, które będą generowane lub weryfikowane w trakcie prowadzonego wnioskowania. To zaś oznacza konieczność takiego zdefiniowania funkcji propagacji niepewności, które gwarantuje spełnienie tych ograniczeń.

Aby zrealizować podane założenia przyjęto, że w procesie wnioskowania progresywnego:

- do agendy będzie kwalifikowana każda i tylko taka reguła, które spełnia następujące trzy warunki:
  - reguła nie była wcześniej wykonywana w bieżącym procesie wnioskowania lub
    po ostatnim jej wykonaniu wystąpiła przejściowo sytuacja niespełnienia jednego z dwóch kolejnych warunków kwalifikacji do agendy,
  - dla każdej zawartej w przesłance reguły klasycznej formuły atomowej P_i $(1 \leq i \leq n)$ w formie prostej (niezanegowanej): $\mathrm{grf}(p_{ri}) : P_1, \dots, P_i, \dots, P_n \rightarrow C : \mathrm{irf}(p_{ci})$, baza danych musi zawierać fakt: $\mathrm{grf}(p_{rj}) : \mathrm{true} \rightarrow P_j : \mathrm{irf}(p_{cj})$ taki, że:
    - wartości p_{ri} i p_{rj} są nie niższe od wymaganej wartości progowej τ_1,
    - spełniona jest relacja: $P_i \leq_{\mathrm{AS}} P_j$,
    - wartość p_{cj} jest nie niższa od wymaganej wartości progowej τ_2;
  - dla każdej zawartej w przesłance reguły klasycznej formuły atomowej P_i $(1 \leq i \leq n)$ w formie zanegowanej: $\mathrm{grf}(p_{ri}) : P_1, \dots, \neg P_i, \dots, P_n \rightarrow C : \mathrm{irf}(p_{ci})$, baza danych musi zawierać fakt: $\mathrm{grf}(p_{rj}) : \mathrm{true} \rightarrow P_j : \mathrm{irf}(p_{cj})$ taki, że:
    - wartości p_{ri} i p_{rj} są nie niższe od wymaganej wartości progowej τ_1,
    - spełniona jest relacja: $P_j \leq_{\mathrm{AS}} P_i$,
    - wartość p_{cj} jest nie wyższa od wartości progowej $(1 - \tau_2)$;

- konflikt w agendzie będzie rozstrzygany na rzecz reguły o najwyższej wartości współczynnika grf; jeśli zastosowanie kryterium grf nie będzie wystarczające do ustalenia kolejności reguł w agendzie, to zostanie wzięta pod uwagę dodatkowa cecha reguły, np. jej pozycja porządkowa w bazie wiedzy;
- przy wykonywaniu reguły $\mathrm{grf}(p_{ri}) : P_1, \dots, P_i, \dots, P_n \rightarrow C : \mathrm{irf}(p_{ci})$, spośród wszystkich znajdujących się w bazie danych faktów z konkluzją P_j spełniających wymagania klasycznej formuły atomowej P_i $(1 \leq i \leq n)$ w formie prostej (niezanegowanej), w procesie propagacji niepewności zostanie uwzględniony fakt z najwyższą wartością p_{cj} współczynnika irf;
- przy wykonywaniu reguły $\mathrm{grf}(p_{ri}) : P_1, \dots, \neg P_i, \dots, P_n \rightarrow C : \mathrm{irf}(p_{ci})$, spośród wszystkich znajdujących się w bazie danych faktów z konkluzją P_j spełniających wymagania klasycznej formuły atomowej P_i $(1 \leq i \leq n)$ w formie zanegowanej, w procesie propagacji niepewności zostanie uwzględniony fakt z najniższą wartością p_{cj} współczynnika irf.

Reasumując, na danym etapie wnioskowania progresywnego będzie wykonywana najbardziej wiarygodna spośród tych reguł, które spełniają minimalne żądania dotyczące istnienia i szczegółowej postaci faktów odpowiadających formułom atomowym z przesłanki reguły. Dla każdej takiej formuły atomowej, spośród wszystkich odpowiadających jej faktów w procesie propagacji niepewności będzie uwzględniany ten, który jest najbardziej „korzystny".

Skutek wykonania reguły będzie zależał od stanu bazy danych na danym etapie procesu wnioskowania. Przyjąwszy, że reguła jest w postaci 6.7, należy rozpatrzyć następujące trzy przypadki:

- fakt z konkluzją C (negacją konkluzji C) znajduje się w bazie danych i jest oznaczony jako aksjomat,
- nie ma w bazie danych faktu z konkluzją C (negacją konkluzji C),
- fakt z konkluzją C (negacją konkluzji C) znajduje się w bazie danych i nie ma żadnego oznaczenia (nie jest aksjomatem, lecz został wygenerowany w poprzednich krokach procesu wnioskowania).

W wymienionych sytuacjach silnik wnioskujący zadziała jak następuje:

- nie podejmie żadnego działania – fakty aksjomatyczne to dane wejściowe dla procesu wnioskowania; w związku z tym nie powinny podlegać żadnym modyfikacjom;
- wstawi odpowiedni, nie-aksjomatyczny fakt z konkluzją C (negacją konkluzji C) do bazy danych,
- zastąpi w bazie danych nie-aksjomatyczny fakt z konkluzją C (negacją konkluzji C) nowym faktem z tą samą konkluzją.

W drugim i trzecim wypadku, konstrukcja faktu z konkluzją C (negacją konkluzji C) będzie się odbywać przy wykorzystaniu odpowiednio zaimplementowanych funkcji propagacji niepewności f_{1i}, f_{2i}, f_{3i}, f_{4i} $(1 \leq i \leq 2)$ oraz pewnej ustalonej wartości stałej g (patrz podrozdz. 6.4.2). Wszystkie planowane zmiany w bazie danych powinny być poddane ocenie systemu utrzymywania wiarygodności.

Wnioskowanie regresywne będzie przebiegać zgodnie z uniwersalnym schematem przyjętym dla systemów regułowych z niepewnością. Szczegółowe założenia tyczące tego procesu będą zrealizowane poprzez odpowiednią implementację funkcji propagacji niepewności f_{5i}, f_{60j}, f_{6jk}, f_{7i} ($0 \leq i \leq 2$, $1 \leq j \leq n$, $1 \leq k \leq 2$; patrz podrozdz. 6.4.3). Spośród możliwych operatorów relacyjnych, w pytaniach będą używane jedynie $\leq$ i $\geq$.

Pesymistyczną złożoność czasową pojedynczego kroku wnioskowania progresywnego (aktualizacja zawartości agendy, wybór reguły i jej wykonanie) w systemie $\boldsymbol{R\_2U}$ można wyrazić wzorem $\mathrm{O}(z \cdot c_1 \cdot c_2 \cdot n^2)$, gdzie n oznacza wielkość bazy wiedzy systemu, c_1 – maksymalną liczbę formuł atomowych w przesłance reguły, c_2 – maksymalną liczbę faktów w bazie danych, a z – współczynnik narzutu. Liczbowy współczynnik $z > 1$ reprezentuje narzut czasowy związany z testowaniem ograniczeń progowych oraz wykonywaniem obliczeń funkcji propagacji niepewności. Jego wartość może być znacząca w stosunku do iloczynu $c_1 \cdot c_2$. Złożoność całego procesu wnioskowania regresywnego jest natomiast wykładnicza względem wielkości bazy wiedzy systemu.

6.4.2. Propagacja niepewności w procesie wnioskowania progresywnego

Zamieszczona w poprzednim podrozdziale charakterystyka silnika wnioskującego systemu $\boldsymbol{R\_2U}$ wymaga uzupełnienia o definicje funkcji propagacji niepewności. Oto definicje funkcji używanych w procesie wnioskowania progresywnego.

Propagacja niepewności przesłanek składowych do przesłanki złożonej. W systemie $\boldsymbol{R\_2U}$ dopuszczono przesłanki złożone, sformułowane w postaci koniunkcyjnej. Aby określić stopień pewności całej przesłanki F_v: $\mathrm{grf}(p_{rv}) : \mathrm{true} \rightarrow P_1 \wedge P_2 \wedge \ldots \wedge P_n : \mathrm{irf}(p_{cv})$ na podstawie znanych stopni pewności n faktów składowych F_i: $\mathrm{grf}(p_{ri}) : \mathrm{true} \rightarrow P_i : \mathrm{irf}(p_{ci})$ $(1 \leq i \leq n)$, należy skorzystać z poniższych funkcji f_{11} i f_{12}:

$$\mathrm{f}_{11}, \mathrm{f}_{12} \colon ([0;1] \times [0;1])^n \rightarrow [0;1] ;$$

$$\mathrm{f}_{11}(p_{c1}, p_{r1}, p_{c2}, p_{r2}, \ldots, p_{cn}, p_{rn}) = \min\{\mathrm{p}_{c1}, \mathrm{p}_{c2}, \ldots, \mathrm{p}_{cn}\} ;$$

$$\mathrm{f}_{12}(p_{c1}, p_{r1}, p_{c2}, p_{r2}, \ldots, p_{cn}, p_{rn}) = \min\{\mathrm{p}_{r1}, \mathrm{p}_{r2}, \ldots, \mathrm{p}_{rn}\}.$$

Z powyższych wzorów wynika, że współczynniki irf i grf przesłanki złożonej reprezentującej wirtualny fakt F_v oblicza się jako minimum z wartości współczynników składowych, odpowiednio irf i grf. Zatem, każda z funkcji f_{11} i f_{12} ma w istocie n jednorodnych parametrów. Obliczeniom tych funkcji nie towarzyszy ani tworzenie, ani modyfikacja rzeczywistego faktu 2U – analizowana przesłanka pełni rolę faktu wirtualnego.

Zgodnie z założeniem (patrz podrozdz. 4.3.2), wartość p_{cv} współczynnika irf faktu wirtualnego F_v powinna pochodzić z zakresu $\langle Bel_{cv}; Pls_{cv}\rangle$ takiego, że: $Bel_{cv} = \max\{Bel_{c1} + \ldots + Bel_{cn} - (n-1), 0\}$ i $Pls_{cv} = \min\{Pls_{c1}, \ldots, Pls_{cn}\}$, gdzie Bel_{ci}, Pls_{ci} $(1 \le i \le n)$ oznaczają, odpowiednio globalny stopień przekonania i stopień możliwości współczynnika irf i-tego faktu składowego F_i. Podobnie, wartość p_{rv} współczynnika grf faktu wirtualnego F_v powinna pochodzić z zakresu $\langle Bel_{rv}; Pls_{rv}\rangle$ takiego, że: $Bel_{rv} = \min\{Bel_{r1}, \ldots, Bel_{rn}\}$ i $Pls_{rv} = \min\{Pls_{r1}, \ldots, Pls_{rn}\}$, gdzie Bel_{ri}, Pls_{ri} $(1 \le i \le n)$ oznaczają, odpowiednio globalny stopień przekonania i stopień możliwości współczynnika grf i-tego faktu składowego F_v.

Przyjąwszy, że prawdziwe są nierówności: $Bel_{ci} \le p_{ci} \le Pls_{ci}$ oraz $Bel_{ri} \le p_{ri} \le Pls_{ri}$, można łatwo wykazać, że: $p_{cv} \ge Bel_{cv}$. Skoro $p_{cv} = \min\{\mathrm{p}_{c1}, \mathrm{p}_{c2}, \ldots, \mathrm{p}_{cn}\}$, to $p_{cv} = p_{ci}$ dla pewnego i z przedziału $\langle 1; n\rangle$. Z tego wynika, że: $p_{cv} = p_{ci} \ge Bel_{ci} \ge Bel_{ci} + b$, gdzie $b \le 0$. Kontynuując, $p_{cv} \ge Bel_{ci} + \Big((Bel_{c1} - 1) + \cdots + (Bel_{ci-1} - 1) + (Bel_{ci+1} - 1) + \cdots + (Bel_{cn} - 1)\Big) = Bel_{c1} + \cdots + Bel_{cn} - (n-1)$. Z drugiej strony, skoro $p_{cv} = \min\{\mathrm{p}_{c1}, \mathrm{p}_{c2}, \ldots, \mathrm{p}_{cn}\}$, to zachodzi: $p_{cv} \ge 0$. Podsumowując, $p_{cv} \ge \max\{Bel_{c1} + \cdots + Bel_{cn} - (n-1), 0\} = Bel_{cv}$. Prawdziwość nierówności $p_{cv} \le Pls_{cv}$, a także nierówności $Bel_{rv} \le p_{rv} \le Pls_{rv}$ można łatwo wykazać z definicji funkcji f_{11} i f_{12}.

Propagacja niepewności hipotezy – przez regułę – do nowej hipotezy. Podstawowym działaniem wykonywanym w procesie wnioskowania w przód jest wykonywanie reguł. Jeśli wykonywana reguła R ma postać $\mathrm{grf}(p_{rr}) : P \to C : \mathrm{irf}(p_{cr})$, a oba współczynniki wiarygodności faktu F_p: $\mathrm{grf}(p_{rp}) : \mathrm{true} \to P : \mathrm{irf}(p_{cp})$ przyjmują wartość 1, to w wyniku uzyskuje się fakt cząstkowy F_c: $\mathrm{grf}(p_{rc}) : \mathrm{true} \to C : \mathrm{irf}(p_{cc})$. W większości wypadków wartości współczynników p_{cp} i p_{rp} są jednak różne od 1. Dlatego, przy wykonywaniu reguły zachodzi potrzeba skorzystania z poniższych funkcji f_{21} i f_{22}:

$$\mathrm{f}_{21}, \mathrm{f}_{22}: ([0;1] \times [0;1])^2 \to [0;1] ;$$

$$\mathrm{f}_{21}(p_{cp}, p_{rp}, p_{cr}, p_{rr}) = p_{cp} \cdot p_{cr} ;$$

$$\mathrm{f}_{22}(p_{cp}, p_{rp}, p_{cr}, p_{rr}) = \min\{\mathrm{p}_{rp}, \mathrm{p}_{rr}\} .$$

W istocie, każda z funkcji f_{21} i f_{22} zależy nie od czterech, ale od dwóch jednorodnych parametrów. Zależnie od stanu bazy danych, utworzony fakt cząstkowy F_c ignoruje się, wstawia w uzyskanej/zanegowanej postaci do bazy danych, lub dołącza do występującego tam faktu z konkluzją C/negacją konkluzji C. Łatwo zauważyć, że wraz ze wzrostem długości łańcucha wnioskowania, maleją współczynniki irf i grf kolejnych generowanych faktów cząstkowych.

Przyjmuje się (patrz podrozdz. 4.3.2), że wartość p_{cc} współczynnika irf faktu F_c powinna pochodzić z zakresu $\langle Bel_{cc}; Pls_{cc} \rangle$ takiego, że: $Bel_{cc} = Bel_{cp} \cdot Bel_{cr}$ i $Pls_{cc} = Pls_{cp} \cdot Pls_{cr}$, gdzie Bel_{cp}, Bel_{cr} oznaczają globalny stopień przekonania współczynnika irf, odpowiednio faktu F_p i reguły R, a Pls_{cp}, Pls_{cr} – stopień możliwości współczynnika irf, odpowiednio faktu F_p i reguły R. Przy obowiązujących założeniach $Bel_{cp} \leq p_{cp} \leq Pls_{cp}$ oraz $Bel_{cr} \leq p_{cr} \leq Pls_{cr}$, można łatwo dowieść spełnienia żądanego wymogu: $Bel_{cc} \leq p_{cc} \leq Pls_{cc}$. Podobne rozumowanie da się przeprowadzić dla współczynnika grf. W tym wypadku zakłada się, że: $Bel_{rc} = \min\{Bel_{rp}, Bel_{rr}\}$ oraz $Pls_{rc} = \min\{Pls_{rp}, Pls_{rr}\}$. Jeśli przy tym, zgodnie z założeniem, prawdziwe są nierówności: $p_{rp} \geq Bel_{rp}$ oraz $p_{rr} \geq Bel_{rr}$, to: $p_{rc} = \min\{p_{rp}, p_{rr}\} \geq \min\{Bel_{rp}, Bel_{rr}\} = Bel_{rc}$. W podobny sposób można dowieść nierówności $p_{rc} \leq Pls_{rc}$.

Zaproponowane definicje funkcji f_{21} i f_{22} pozostają w zgodności z kluczowym założeniem na temat dodatniego monotonicznego charakteru zależności między przesłanką reguły a jej konkluzją.

Propagacja niepewności hipotez wielokrotnych do hipotezy zbiorczej. Jeśli na pewnym etapie procesu wnioskowania w przód został wygenerowany fakt cząstkowy F_c: $\mathrm{grf}(p_r) : \mathrm{true} \rightarrow C : \mathrm{irf}(p_c)$ z konkluzją C taką, że w bazie danych znajduje się nie-aksjomatyczny fakt $F_{c(n-1)}$: $\mathrm{grf}(p_{r(n-1)}) : \mathrm{true} \rightarrow C : \mathrm{irf}(p_{c(n-1)})$ z tą samą konkluzją, to silnik wnioskujący zastąpi w bazie danych fakt $F_{c(n-1)}$ – nowym faktem F_{cn}: $\mathrm{grf}(p_{rn}) : \mathrm{true} \rightarrow C : \mathrm{irf}(p_{cn})$ z konkluzją C (zgodnie z przyjętym założeniem, w opisanej sytuacji konkluzja C musi być zredagowana w formie prostej, czyli niezanegowanej). Powyższą akcję silnik przeprowadzi przy wykorzystaniu następujących funkcji f_{31} i f_{32}:

$$\mathrm{f}_{31}, \mathrm{f}_{32}: ([0; 1] \times [0; 1])^2 \rightarrow [0; 1] ;$$

$$\mathrm{f}_{31}(p_{c(n-1)}, p_{r(n-1)}, p_c, p_r) = (1 - w_n) \cdot p_{c(n-1)} + w_n \cdot p_c ;$$

$$f_{32}(p_{c(n-1)}, p_{r(n-1)}, p_c, p_r) = \max\{p_{r(n-1)}, p_r\} ,$$

gdzie n ($n \geq 2$) oznacza numer porządkowy faktu cząstkowego z konkluzją C wygenerowanego w bieżącym kroku procesu wnioskowania (fakt F_c), a w_n – wagę tego faktu, spełniającą – wraz z wagami w_1, w_2, …, w_{n-1} przypisanymi wcześniej wygenerowanym faktom z konkluzją C – następujące zależności:

- $w_1 \geq w_2 \geq \cdots \geq w_n$ oraz
- $w_1 + w_2 + \cdots + w_n = 1$.

Przykładową postać wzoru na obliczanie wagi faktu cząstkowego zaproponowała Szymkowiak w pracy [77]:

$$w'_k = \begin{cases} 1 & dla\ k = 1, \\ \frac{w'_{k-1}}{g + w'_{k-1}}, & dla\ 2 \le k \le n, \end{cases} \tag{6.17}$$

$$\begin{cases} w_n = w'_n, \\ w_k = w'_k \cdot (1 - w'_{k+1}) \cdot \ldots \cdot (1 - w'_n) \quad dla\ 1 \le k \le n-1, \end{cases} \tag{6.18}$$

gdzie w'_k $(1 \le k \le n)$ oznacza wagę pierwotną, w_k $(1 \le k \le n)$ – wagę znormalizowaną, gwarantującą spełnienie zależności: $w_1 + \cdots + w_n = 1$, a g – stałą proporcji, reprezentującą żądany stosunek $\frac{w_{k-1}}{w_k}$ $(2 \le k \le n)$ znormalizowanych wag dwóch kolejnych faktów cząstkowych z tą samą konkluzją.

Przykładowo, dla $g = 1.5$ wartości kolejnych pomocniczych wag wynoszą: $w'_1 = 1$, $w'_2 = \frac{1}{2.5} = 0.4$, $w'_3 = \frac{0.4}{1.9} \cong 0.21$, $w'_4 \cong \frac{0.21}{1.71} \cong 0.12$, itd. Po wygenerowaniu ostatniego, czwartego faktu cząstkowego następuje obliczenie wartości wag właściwych: $w_4 \cong 0.12$, $w_3 \cong 0.21 \cdot (1 - 0.12) \cong 0.18$, $w_2 \cong \cong 0.4 \cdot (1 - 0.21) \cdot (1 - 0.12) \cong 0.28$, $w_1 \cong 1 \cdot (1 - 0.4) \cdot (1 - 0.21) \cdot \cdot (1 - 0.12) \cong 0.42$, przy zachowanym stosunku $g \cong \frac{w_1}{w_2} \cong \frac{w_2}{w_3} \cong \frac{w_3}{w_4} \cong 1.5$.

Wagi faktów cząstkowych stają się wagami ich współczynników irf, a przez to wpływają na wartość współczynnika irf faktu zbiorczego. Wagi te nie mają natomiast wpływu na wartość współczynnika grf faktu zbiorczego.

Zgodnie z założeniem (patrz podrozdz. 4.3.2), wartość p_{cn} współczynnika irf faktu F_{cn} powinna pochodzić z zakresu $\langle Bel_{cn}; Pls_{cn} \rangle$ takiego, że: $Bel_{cn} = = \min\{Bel_{c(n-1)}, Bel_c\}$ i $Pls_{cn} = \max\{Pls_{c(n-1)}, Pls_c\}$, gdzie $Bel_{c(n-1)}$ i $Pls_{c(n-1)}$ oznaczają, odpowiednio globalny stopień przekonania i stopień możliwości współczynnika irf faktu $F_{c(n-1)}$, a Bel_c i Pls_c – globalny stopień przekonania i stopień możliwości współczynnika irf faktu F_c. Z definicji funkcji f_{31} wynika, że $p_{cn} = (1 - w_n) \cdot p_{c(n-1)} + w_n \cdot p_c \ge \min\{p_{c(n-1)}, p_c\} \ge \ge \min\{Bel_{c(n-1)}, Bel_c\} = Bel_{cn}$ oraz $p_{cn} = (1 - w_n) \cdot p_{c(n-1)} + w_n \cdot p_c \le \le \max\{p_{c(n-1)}, p_c\} \le \max\{Pls_{c(n-1)}, Pls_c\} = Pls_{cn}$. W takim razie, prawdziwa jest podwójna nierówność: $Bel_{cn} \le p_{cn} \le Pls_{cn}$.

Z kolei, wartość p_{rn} współczynnika grf faktu F_{cn} musi pochodzić z zakresu $\langle Bel_{rn}; Pls_{rn} \rangle$ takiego, że: $Bel_{rn} = \max\{Bel_{r(n-1)}, Bel_r\}$ i $Pls_{rn} = = \max\{Pls_{r(n-1)}, Pls_r\}$, gdzie $Bel_{r(n-1)}$ i $Pls_{r(n-1)}$ oznaczają, odpowiednio globalny stopień przekonania i stopień możliwości współczynnika grf faktu $F_{c(n-1)}$, a Bel_r, Pls_r – globalny stopień przekonania i stopień możliwości współczynnika grf faktu F_c. Przy obowiązujących założeniach $Bel_{r(n-1)} \le \le p_{r(n-1)} \le Pls_{r(n-1)}$ oraz $Bel_r \le p_r \le Pls_r$, można dowieść nierówności

$Bel_{rn} \leq p_{rn} \leq Pls_{rn}$ w podobny sposób, jak to przeprowadzono dla funkcji f_{22} propagacji niepewności hipotezy – przez regułę – do nowej hipotezy.

Zastosowana w systemie $\boldsymbol{R\_2U}$ metoda obliczania współczynnika wiarygodności irf hipotezy zbiorczej różni się zasadniczo od metody obliczania analogicznego współczynnika (CF) w systemie MYCIN. W systemie $\boldsymbol{R\_2U}$ waga poszczególnych hipotez wielokrotnych jest zróżnicowana i zależy od porządku generacji tych hipotez. Zakłada się przy tym, że hipoteza ma tym większą wartość, im wcześniej ją wygenerowano ($g > 1$). Z kolei, w systemie MYCIN wszystkie hipotezy wielokrotne mają równorzędne znaczenie ($g = 1$) i jednakowy wpływ na wartość końcową współczynnika pewności. Założenie poczynione dla systemu $\boldsymbol{R\_2U}$ bazuje na spostrzeżeniu, że silnik wnioskujący w pierwszej kolejności wykonuje te reguły, które w największym stopniu zależą od faktów aksjomatycznych. Dalsze reguły mają charakter wtórny: zależą od wniosków pośrednich, wygenerowanych na wcześniejszych etapach procesu wnioskowania.

Metoda obliczania współczynnika wiarygodności irf w systemie $\boldsymbol{R\_2U}$ wykazuje natomiast duże podobieństwo do metody obliczania współczynnika częstotliwości w nie-aksjomatycznym systemie wnioskowania NARS. Wartość współczynnika częstotliwości hipotezy zbiorczej zależy od współczynników częstotliwości poszczególnych hipotez wielokrotnych, lecz także od współczynników zaufania tych hipotez: im większa jest wartość zaufania hipotezy wielokrotnej, tym większy jest wpływ częstotliwości tej hipotezy na częstotliwość hipotezy zbiorczej. O ile jednak w systemie NARS uwzględnia się w tych obliczeniach bezwzględne wartości współczynników zaufania poszczególnych hipotez wielokrotnych, o tyle w systemie $\boldsymbol{R\_2U}$ bierze się pod uwagę tylko uporządkowanie współczynników grf tych hipotez przez relację $\geq$. Powyższa różnica wynika z odmiennych interpretacji znaczenia współczynnika zaufania w regule Wanga i zewnętrznego współczynnika wiarygodności grf w regule 2U. Pierwszy współczynnik ma interpretację częstotliwościową: jego wielkość można wyliczyć na podstawie samej wielkości ewidencji. Drugi współczynnik ma interpretację złożoną: przyjmuje się, że wiarygodność reguły zależy od wielkości ewidencji, z której tę regułę pozyskano, lecz także od stopnia homogeniczności tej ewidencji oraz od źródła jej pochodzenia. Wpływ wymienionych czynników jest trudny do zmierzenia, dlatego wartości grf nie da się obliczyć – można ją co najwyżej oszacować. Zakłada się przy tym, że błąd oszacowania nie ma wpływu na porządek $\geq$, w jakim pozostają wiarygodności poszczególnych reguł.

Z tego samego założenia wynika przyjęta w systemie $\boldsymbol{R\_2U}$ postać funkcji f_{32}. Ze wszystkich współczynników grf hipotez wielokrotnych, przy szacowaniu wielkości współczynnika grf hipotezy zbiorczej bierze się pod uwagę tylko jeden – najwyższy. W ten sposób hipoteza zbiorcza przechowuje pamięć o „najmocniejszym” wywodzie konkluzji. Pozostałe dopuszczone wywody, spełniające wymogi progów τ_1 i τ_2, wpłyną na wartość współczynnika irf tego faktu.

Propagacja niepewności hipotezy do hipotezy przeciwnej. Z definicji 6.3 wynika, że w regule 2U dopuszcza się użycie formuły atomowej z negacją (6.2 lub 6.4), zarówno po stronie przesłanki, jak i konkluzji. Z drugiej strony, konkluzje wszystkich faktów przechowywanych w bazie systemu $\boldsymbol{R\_2U}$ są zredagowane w formie prostej, czyli niezanegowanej. Z tego powodu, każde wykonanie reguły 2U zawierającej w przesłance formułę zanegowaną musi być poprzedzone „odwróceniem" tej formuły. Proces taki odbywa się w locie i nie powoduje żadnej zmiany w stanie bazy danych (nowo uzyskany fakt ma charakter wirtualny). Podobny proces zachodzi po wykonaniu reguły 2U z konkluzją w postaci formuły zanegowanej. Zanim zostanie podjęta właściwa obsługa faktu cząstkowego z tą konkluzją, wcześniej nastąpi jej odwrócenie. Aby określić stopień pewności hipotezy F_c': $\mathrm{grf}(p_r')$: true → $\neg C$: $\mathrm{irf}(p_c')$ uzyskanej w procesie zmiany hipotezy F_c: $\mathrm{grf}(p_r)$: true → C : $\mathrm{irf}(p_c)$ na przeciwną, należy skorzystać z funkcji propagacji niepewności f_{41} i f_{42}, zdefiniowanych jak następuje:

$$\mathrm{f}_{41}, \mathrm{f}_{42}: [0;1] \times [0;1] \rightarrow [0;1] ;$$

$$\mathrm{f}_{41}(p_c, p_r) = 1 - p_c ;$$

$$\mathrm{f}_{42}(p_c, p_r) = p_r .$$

Przy założeniu $Bel_c \leq p_c \leq Pls_c$, ze względu na postać funkcji f_{41}, prawdziwa jest podwójna nierówność: $(1 - Pls_c) \leq p_c' \leq (1 - Bel_c)$. W konsekwencji, ze względu na zachodzenie oczywistych równości: $Bel_c' = (1 - Pls_c)$ i $Pls_c' = (1 - Bel_c)$, prawdziwa jest podwójna nierówność: $Bel_c' \leq p_c' \leq Pls_c'$. Z drugiej strony, ze względu na zachodzenie równości: $Bel_r' = Bel_r$ i $Pls_r' = Pls_r$, prawdziwe są nierówności: $Bel_r' \leq p_r' = = p_r \leq Pls_c'$.

Należy w tym miejscu podkreślić, że funkcja propagacji niepewności f_{41} może być stosowana tylko w odniesieniu do faktów 2U. Gdyby dopuścić odwracanie dowolnych reguł 2U, doszłoby do naruszenia wymogu zależności dodatniej monotonicznej między przesłanką a konkluzją reguły. Z dwóch reguł: $\mathrm{grf}(p_{r1})$: $P \rightarrow C$: $\mathrm{irf}(p_{c1})$ i $\mathrm{grf}(p_{r2})$: $P \rightarrow \neg C$: $\mathrm{irf}(p_{c2})$, gdzie P oznacza przesłankę różną od true, przynajmniej jedna nie spełnia tego wymogu. W wypadku przesłanki równej true, reguła staje się faktem, a zależność między przesłanką a konkluzją nabiera charakteru trwałego: przesłanka true będzie zawsze spełniona w 100%.

6.4.3. Propagacja niepewności w procesie wnioskowania regresywnego

Choć wnioskowanie regresywne przebiega w systemie $\boldsymbol{R\_2U}$ według ogólnie przyjętych zasad (patrz podrozdz. 4.3.2), to używane w tym trybie funkcje propagacji niepewności mają charakter oryginalny. W dalszym ciągu podano definicje i krótkie opisy tych funkcji. Należy przypomnieć, że w procesie wnioskowania regresywnego nie powstają żadne nowe fakty rzeczywiste, czy wirtualne. Z tego powodu, nie trzeba tutaj dowodzić spełnienia jakichkolwiek ograniczeń, tak jak to miało miejsce w wypadku funkcji używanych w trybie wnioskowania progresywnego.

Wsteczna propagacja niepewności hipotezy – przez regułę – do nowej hipotezy. Podstawową czynnością wykonywaną w procesie wnioskowania wstecz jest reformułowanie hipotezy wejściowej. Jeśli to postępowanie prowadzi ostatecznie do aksjomatu/aksjomatów znajdujących się w bazie danych, to następuje weryfikacja hipotezy, w przeciwnym razie – jej falsyfikacja. Jeśli na pewnym etapie tego procesu pytanie ma postać Q_c: $(\mathrm{grf}(p_{rc}) : C : \mathrm{irf}(p_{cc}) \leftarrow \mathrm{true}, rel_c)$, a w bazie wiedzy znajduje się reguła R w postaci: $\mathrm{grf}(p_{rr}) : P_1 \wedge P_2 \wedge \ldots \wedge P_n \rightarrow C : \mathrm{irf}(p_{cr})$ taka, że: $p_{cr} \neq 0$, $p_{cr} \geq p_{cc}$ i $p_{rr}\, rel_c\, p_{rc}$, to można to pytanie przekształcić do nowej postaci Q_v: $(\mathrm{grf}(p_{rv}) : P_1, P_2, \ldots, P_n : \mathrm{irf}(p_{cv}) \leftarrow \mathrm{true}, rel_v)$, gdzie rel_v oznacza operator wyznaczony przy użyciu funkcji f_{50}, a p_{cv} i p_{rv} oznaczają współczynniki wiarygodności obliczone przy użyciu funkcji, odpowiednio f_{51} i f_{52}, w postaci:

$$\mathrm{f}_{50}: \{\leq, \geq\} \rightarrow \{\leq, \geq\} ;$$

$$\mathrm{f}_{51}, \mathrm{f}_{52}: ([0;1] \times [0;1])^2 \rightarrow [0;1] ;$$

$$\mathrm{f}_{50}(\leq) = \leq ; \mathrm{f}_{50}(\geq) = \geq ;$$

$$\mathrm{f}_{51}(p_{cc}, p_{rc}, p_{cr}, p_{rr}) = \frac{p_{cc}}{p_{cr}} ;$$

$$\mathrm{f}_{52}(p_{cc}, p_{rc}, p_{cr}, p_{rr}) = p_{rc} .$$

Kierunek wnioskowania regresywnego, prowadzącego od hipotezy do aksjomatów, jest odwrotny do kierunku wnioskowania progresywnego. Stąd, funkcja f_{51} wykonuje działanie odwrotne do jej progresywnego odpowiednika f_{21}, a funkcja f_{52} – działanie odwrotne do jej odpowiednika f_{22}. Założenia odnośnie do postaci reguły R są konieczne dla zagwarantowania poprawności definicji f_{51} i f_{52}.

Wsteczna propagacja niepewności hipotezy złożonej do hipotez składowych. Reguła użyta na pewnym etapie weryfikacji hipotezy wejściowej może mieć przesłankę złożoną, sformułowaną w postaci koniunkcyjnej. Aby określić wymagany stopień pewności hipotez składowych, zawartych w pytaniach Q_i: $(\mathrm{grf}(p_{ri}) : P_i \leftarrow \mathrm{true} : \mathrm{irf}(p_{ci}), rel_i)$ $(1 \leq i \leq n)$, na podstawie wymaganego stopnia pewności hipotezy złożonej, zawartej w pytaniu Q_v: $(\mathrm{grf}(p_{rv}) : P_1, P_2, \ldots, P_n : \mathrm{irf}(p_{cv}) \leftarrow \mathrm{true}, rel_v)$, należy zastosować poniższe funkcje f_{60i}, f_{6i1} i f_{6i2} $(1 \leq i \leq n)$:

$$\mathrm{f}_{60i}: \{\leq, \geq\} \rightarrow \{\leq, \geq\} ;$$

$$\mathrm{f}_{6i1}, \mathrm{f}_{6i2}: [0;1] \times [0;1] \rightarrow [0;1] ;$$

$\mathrm{f}_{60i}(\leq) = \leq$; $\mathrm{f}_{60i}(\geq) = \geq$;

$\mathrm{f}_{6i1}(p_{cv}, p_{rv}) = p_{cv}$;

$\mathrm{f}_{6i2}(p_{cv}, p_{rv}) = p_{rv}$.

Z definicji tych wynika, że zarówno współczynniki irf, jak i współczynniki grf hipotez składowych wymagających weryfikacji dla udowodnienia pewnej hipotezy złożonej są takie same jak w tej hipotezie złożonej. Tym samym, każda z funkcji f_{6i1} i f_{6i2} $(1 \leq i \leq n)$ jest w istocie funkcją jednoargumentową.

Wsteczna propagacja niepewności pytania do pytania przeciwnego. Podobnie jak w procesie wnioskowania progresywnego zachodzi potrzeba odwracania hipotez, tak podczas wnioskowania regresywnego zachodzi potrzeba odwracania pytań. Proces formułowania pytania Q': $(\mathrm{grf}(p'_r) : P' \leftarrow \mathrm{true} : \mathrm{irf}(p'_c), op')$, przeciwnego do pytania Q: $(\mathrm{grf}(p_r) : P \leftarrow \mathrm{true} : \mathrm{irf}(p_c), op)$, $P' = \neg P$, odbywa się przy użyciu następujących funkcji propagacji niepewności f_{70}, f_{71} i f_{72}:

$\mathrm{f}_{70}: \{\leq, \geq\} \rightarrow \{\leq, \geq\}$;

$\mathrm{f}_{71}, \mathrm{f}_{72}: [0; 1] \times [0; 1] \rightarrow [0; 1]$;

$\mathrm{f}_{70}(\leq) = \geq$; $\mathrm{f}_{70}(\geq) = \leq$;

$\mathrm{f}_{71}(p_c, p_r) = 1 - p_c$;

$\mathrm{f}_{72}(p_c, p_r) = p_r$.

Spośród przytoczonych funkcji, szczególną uwagę zwraca f_{70}: jest to jedyna funkcja wstecznej propagacji niepewności, która dokonuje zmiany operatora relacyjnego na przeciwny. Podobnie jak poprzednio, funkcje f_{71} i f_{72} są w istocie funkcjami jednoargumentowymi.

6.5 Przykłady wnioskowania

Zamieszczone poniżej przykłady ilustrują sposób działania silnika wnioskującego w systemie R_2U. W wyliczeniach wartości współczynników irf i grf zastosowano zaokrąglenia arytmetyczne do dwóch miejsc po przecinku.

Przykład 6.1. W bazie wiedzy systemu $\boldsymbol{R\_2U}$ znajdują się następujące reguły $R_1 - R_5$, reprezentujące wiedzę z dziedziny o umownej nazwie Studenci_Politechniki_Poznańskiej, na temat zależności występujących pomiędzy:

– wydziałem i trybem studiów (Wydz_Rodz) studenta Politechniki Poznańskiej (PP) a kierunkiem jego studiów na PP (Kier) – R_1, R_2;
– kierunkiem studiów, rokiem studiów (Rok) i płcią (Plec) studenta PP – a posiadanym przez niego sprzętem komputerowym (Sprzet) – R_3, R_4, R_5:

R_1: $\mathrm{grf}(0.75) : \mathrm{Wydz\_Rodz} = \{\mathrm{WInf\_st}, \mathrm{WEl\_st}\} \oplus$
$\rightarrow \mathrm{Kier} = \{\mathrm{Inf}\} \odot : \mathrm{irf}(0.55)$

R_2: grf(0.9) : Wydz_Rodz = {WInf_st} $\odot$
→ Kier = {Inf} $\odot$: irf(0.8)

R_3: grf(0.7) : Kier = {Inf} $\odot$
→ Sprzet = {K_st, Lap} $\odot$: irf(0.6)

R_4: grf(0.9) : Kier = {Inf} $\odot$,
Rok = ¬{1,2} $\oplus$,
→ Sprzet = {K_st, Lap} $\odot$: irf(0.8)

R_5: grf(0.85) : Kier = {Inf} $\odot$,
Rok = ¬{1,2} $\oplus$,
Plec = ¬{K} $\odot$
→ Sprzet = {K_st, Lap} $\odot$: irf(0.85)

gdzie literał WInf_st oznacza Wydział_Informatyki_studia_stacjonarne; WEl_st – Wydział_Elektryczny_studia_stacjonarne; Inf – Informatykę; K_st – komputer_stacjonarny; Lap – laptop; 1, 2 – odpowiednio pierwszy i drugi rok studiów; K – kobietę. Użyte w regułach współczynniki wiarygodności mają jedynie charakter poglądowy.

Konstrukcję zbioru reguł $R_1 - R_5$ należy uznać za poprawną, ponieważ:

- nie istnieje żadna przesłanka świadcząca o nieprawidłowej budowie którejkolwiek z reguł $R_1 - R_5$,
- nie ma w zbiorze takiej pary reguł, które pozostawałyby we wzajemnej sprzeczności,
- nie ma w zbiorze reguły subsumowanej przez inną regułę z tego zbioru.

Jeśli w opisanej sytuacji, przy stanie bazy danych:

F_1^*: grf(1.0) : true
→ Wydz_Rodz = {WInf_st} $\odot$: irf(1.0)

F_2^*: grf(0.95) : true
→ Rok = {1,2} $\oplus$: irf(0.0)

przy parametrach progowych $\tau_1 = 0.7$ i $\tau_2 = 0.7$ oraz stosunku wag hipotez wielokrotnych $g = 1.8$ zostanie zainicjowane wnioskowanie w trybie progresywnym, to zawartość bazy danych będzie się zmieniać jak następuje:

Krok1:

F_1^*: grf(1.0) : true
→ Wydz_Rodz = {WInf_st} $\odot$: irf(1.0)

F_2^*: grf(0.95) : true
→ Rok = {1,2} $\oplus$: irf(0.0)

F_3: grf(0.9) : true
→ Kier = {Inf} $\odot$: irf(0.8)

Ten stan bazy danych jest wynikiem wykonania reguły R_2, wybranej z dwóch zakwalifikowanych do agendy reguł R_1 i R_2. Obie reguły spełniały wymagania progowe τ_1 ($\min\{0.75, 1.0\} \geq 0.7$, $\min\{0.9, 1.0\} \geq 0.7$) i τ_1 ($1.0 \geq 0.7$).

Wybór reguły R_2 wynikał z relacji pomiędzy współczynnikami grf obu reguł: $0.9 \geq 0.75$. Wykonanie reguły R_2 wiązało się z użyciem funkcji propagacji f_{21} i f_{22}. Wartość p_{c3} współczynnika irf nowo uzyskanego faktu F_3 została obliczona z zależności: $p_{c3} = 0.8 \cdot 1.0 = 0.8$, a wartość p_{r3} współczynnika grf tego faktu – z zależności: $p_{c3} = \min\{0.9, 1.0\} = 0.9 = 0.9$.

W zapisie faktów F_1^* i F_2^* użyto symbolu $*$ na oznaczenie tego, że są one aksjomatami i nie podlegają żadnym modyfikacjom w całym procesie wnioskowania. Odmienny charakter ma fakt F_3, który stanie się w Kroku3 podmiotem działania funkcji propagacji niepewności f_{31} i f_{32}. Krok 2:

F_1^*: grf(1.0) : true
$\rightarrow$ Wydz_Rodz = {WInf_st} $\odot$: irf(1.0)

F_2^*: grf(0.95) : true
$\rightarrow$ Rok = {1,2} $\oplus$: irf(0.0)

F_3: grf(0.9) : true
$\rightarrow$ Kier = {Inf} $\odot$: irf(0.8)

F_4: grf(0.9) : true
$\rightarrow$ Sprzet = {K_st, Lap} $\odot$: irf(0.64)

Nowo umieszczony w bazie danych fakt F_4 znalazł się tam w wyniku wykonania reguły R_4, wybranej z agendy spośród reguł R_1, R_3 i R_4. Kwalifikacja R_4 do agendy odbyła się po uprzednim odwróceniu faktu F_2^* do postaci $F_2^{*\prime}$: grf(0.95) : true $\rightarrow$ Rok = $\neg$\{1,2\} $\oplus$: irf(1.0) (z zastosowaniem funkcji propagacji f_{41} i f_{42}), a następnie skonstruowaniu faktu wirtualnego F_v: grf(0.9) : true $\rightarrow$ (Kier = {Inf} $\odot$ $\wedge$ Rok = $\neg$\{1,2\} $\oplus$) : irf(0.8) (z zastosowaniem funkcji propagacji f_{11} i f_{12}). Wartości p_{c4} i p_{r4} współczynników, odpowiednio irf i grf faktu F_4 wyznaczono ze wzorów: $p_{c4} = 0.8 \cdot 0.8 = 0.64$ i $p_{r4} = \min\{0.9, 0.9\} = 0.9$.

Krok3:

F_1^*: grf(1.0) : true
$\rightarrow$ Wydz_Rodz = {WInf_st} $\odot$: irf(1.0)

F_2^*: grf(0.95) : true
$\rightarrow$ Rok = {1,2} $\oplus$: irf(0.0)

F_3 : rf(0.9) : true
$\rightarrow$ Kier = {Inf} $\odot$: irf(0.71)

F_4 : grf(0.9) : true
$\rightarrow$ Sprzet = {K_st, Lap} $\odot$: irf(0.64)

W kolejnym kroku wnioskowania nastąpiło wykonanie reguły R_1. Reguła ta spowodowała wygenerowanie faktu cząstkowego F_{3_2}: $\mathrm{grf}(0.75) : \mathrm{true} \rightarrow \mathrm{Kier} = \{\mathrm{Inf}\} \odot : \mathrm{irf}(0.55)$, a następnie zastąpienie dotychczasowego faktu F_3 – nowym (zbiorczym) faktem F_3, w postaci: $\mathrm{grf}(0.9) : \mathrm{true} \rightarrow \mathrm{Kier} = \{\mathrm{Inf}\} \odot : \mathrm{irf}(0.71)$. Wartości p_{c3} i p_{r3} współczynników, odpowiednio irf i grf zmodyfikowanego faktu F_3 zostały obliczone przy użyciu funkcji propagacji f_{31} i f_{32}, jako: $p_{c3} = \left(1 - \frac{1}{1.8+1}\right) \cdot 0.8 + \frac{1}{1.8+1} \cdot 0.55 \cong 0.71$ oraz $p_{r3} = \max\{0.9, 0.75\} = 0.9$.

Krok 4:

F_1^*: $\mathrm{grf}(1.0) :$ true
$\rightarrow \mathrm{Wydz\_Rodz} = \{\mathrm{WInf\_st}\} \odot : \mathrm{irf}(1.0)$

F_2^*: $\mathrm{grf}(0.95) :$ true
$\rightarrow \mathrm{Rok} = \{1,2\} \oplus : \mathrm{irf}(0.0)$

F_3: $\mathrm{grf}(0.9) :$ true
$\rightarrow \mathrm{Kier} = \{\mathrm{Inf}\} \odot : \mathrm{irf}(0.71)$

F_4: $\mathrm{grf}(0.9) :$ true
$\rightarrow \mathrm{Sprzet} = \{\mathrm{K\_st}, \mathrm{Lap}\} \odot : \mathrm{irf}(0.57)$

Jako ostatnia w procesie wnioskowania została wykonana reguła R_3. Także i ona spowodowała wygenerowanie faktu cząstkowego – F_{4_2}, w postaci: $\mathrm{grf}(0.7) : \mathrm{true} \rightarrow \mathrm{Sprzet} = \{\mathrm{K\_st}, \mathrm{Lap}\} \odot : \mathrm{irf}(0.43)$, a po nim nowego (zbiorczego) faktu F_4: cktóry zastąpił w bazie danych fakt F_4 z Kroku3. Współczynniki tego nowego faktu zostały wyznaczone w rezultacie obliczeń: $p_{c4} = \left(1 - \frac{1}{1.8+1}\right) \cdot 0.64 + + \frac{1}{1.8+1} \cdot 0.43 \cong 0.57$ oraz $p_{r4} = \max\{0.9, 07\} = 0.9$.

Wszystkie reguły kwalifikowane w poszczególnych krokach wnioskowania do agendy spełniały, co łatwo sprawdzić, wymagania progowe τ_1 i τ_2. Przy obowiązującym w systemie $\boldsymbol{R\_2U}$ założeniu otwartego świata OWA, reguła R_5 nie znalazła się w agendzie ze względu na brak w bazie danych faktu z konkluzją (Plec = {K} ⊙). □

Przykład 6.2. Jeśli baza danych będzie miała identyczną zawartość początkową jak w przykładzie 6.1, to – przy tej samej zawartości bazy wiedzy (R_1 – R_5), tych samych wartościach progowych $\tau_1 = 0.7$ i $\tau_2 = 0.7$ oraz nowym, radykalnie zmniejszonym stosunku wag hipotez wielokrotnych $g = 1.2$ – proces wnioskowania będzie miał następujący przebieg:

– Krok1 i Krok2 zakończą się identycznym rezultatem, jak w trakcie poprzedniego wnioskowania;

– Krok3 spowoduje zmianę stanu bazy danych na poniższy, różny od osiągniętego w trakcie poprzedniego wnioskowania:

Krok3:

F_1^*: grf(1.0) : true
$\rightarrow$ Wydz_Rodz = {WInf_st} $\odot$: irf(1.0)

F_2^*: grf(0.95) : true
$\rightarrow$ Rok = {1,2} $\oplus$: irf(0.0)

F_3: grf(0.9) : true
$\rightarrow$ Kier = {Inf} $\odot$: irf(0.69)

F_4: grf(0.9) : true
$\rightarrow$ Sprzet = {K_st, Lap} $\odot$: irf(0.64)

gdzie wartości p_{c3} i p_{r3} współczynników, odpowiednio irf i grf zmodyfikowanego faktu F_3 wynikają ze wzorów: $p_{c3} = \left(1 - \frac{1}{1.2+1}\right) \cdot 0.8 + \frac{1}{1.2+1} \cdot$ $\cdot\, 0.55 \cong 0.69$ oraz $p_{r3} = \max\{0.9, 0.75\} = 0.9$;

– proces wnioskowania dobiegnie końca ze względu na brak reguł kwalifikujących się do agendy. W nowych warunkach reguła R_3 nie spełnia wymogu $\tau_2 \geq 0.7$.

Pobieżna analiza odpowiedzi uzyskanych w procesach wnioskowania z przykładów 6.1 i 6.2 prowadzi do wniosku, że przebiegiem wnioskowania progresywnego w systemie $\boldsymbol{R\_2U}$ można sterować za pomocą każdego z parametrów τ_1, τ_2 i g. W szczególności, sama zmiana wartości g może wpływać na skrócenie/wydłużenie łańcuchów wnioskowania.

W ogólności, wraz z przebiegiem procesu wnioskowania progresywnego postępuje spadek wielkości współczynników (grf i irf) nowo generowanych faktów 2U. Spadek ten powoduje, że spełnienie wymagań progowych τ_1 i τ_2 staje się coraz trudniejsze, a w końcu niemożliwe, dając efekt w postaci pustej agendy. Z drugiej strony, przy dużych wartościach g, wpływ kolejno uzyskiwanych faktów cząstkowych na postać faktu zbiorczego może być tak nikły, że nie warto go brać pod uwagę. W takich okolicznościach można by rozważyć nowe kryterium zakończenia procesu wnioskowania progresywnego. □

Przykład 6.3. Jeśli baza wiedzy będzie miała identyczną zawartość jak w przykładzie 6.1 (reguły $R_1 - R_5$), to przy nowej początkowej zawartości bazy danych:

F_1^*: grf(1.0) : true
$\rightarrow$ Wydz_Rodz = {WEl_st} $\odot$: irf(1.0)

F_2^*: grf(0.95) : true
$\rightarrow$ Rok = {1,2} $\oplus$: irf(0.0)

oraz niezmienionych wartościach progowych $\tau_1 = 0.7$, $\tau_2 = 0.7$ i stosunku wag hipotez wielokrotnych $g = 1.8$, proces wnioskowania progresywnego zakończy się na Kroku1, w którym zostanie wygenerowany nowy fakt F_3, w postaci:

$$F_3: \mathrm{grf}(0.75): \ \mathrm{true} \\ \rightarrow \ \mathrm{Kier} = \{\mathrm{Inf}\} \odot : \mathrm{irf}(0.55)$$

Powyższy fakt jest wynikiem wykonania reguły R_1, jedynej znajdującej się w agendzie. W nowym stanie bazy danych żadna z reguł nie spełni wymogów kwalifikacji do agendy. Główną przeszkodę na tej drodze stanowi niska wiarygodność irf faktu F_3.

Podsumowując, o ile dla studenta Wydziału_Informatyki_studiów_stacjonarnych roku trzeciego lub wyższego dało się wygenerować hipotezę na temat posiadania przez niego komputera stacjonarnego i laptopa, o tyle nie można uzyskać analogicznej hipotezy dla studenta Wydziału_Elektrycznego_studiów_stacjonarnych. Ta różnica jest w pełni uzasadniona: wiedza na temat posiadanego sprzętu komputerowego dotyczy tylko studentów kierunku Informatyka (reguły $R_3 - R_5$), a ci są w mniejszości wśród studentów Wydziału_Elektrycznego_studiów_stacjonarnych (uśredniona wartość 0.55 współczynnika irf w regule R_1 odnosi się do studentów obu rozważanych wydziałów – Elektrycznego i Informatyki; to uśrednienie znajduje wyraz w stosunkowo niskiej wiarygodności grf tej reguły).

Chcąc koniecznie uzyskać hipotezę dotyczącą sprzętu komputerowego studenta Wydziału_Elektrycznego_studiów_stacjonarnych, należy obniżyć wymagania progowe, np. do poziomów $\tau_1 = 0.6$, $\tau_2 = 0.5$. W takiej sytuacji, przy niezmienionym poziomie $g = 1.8$, system wygeneruje w odpowiedzi następujące fakty:

$$F_3: \mathrm{grf}(0.75): \ \mathrm{true} \\ \rightarrow \mathrm{Kier} = \{\mathrm{Inf}\} \odot : \mathrm{irf}(0.55)$$

$$F_4: \mathrm{grf}(0.75): \ \mathrm{true} \\ \rightarrow \mathrm{Sprzet} = \{\mathrm{K\_st}, \mathrm{Lap}\} \odot : \mathrm{irf}(0.40)$$

Jak widać, jest miedzy nimi żądana hipoteza (F_4); jej wiarygodność zewnętrzna ($p_{r4} = 0.75$) jest jednak znacznie niższa od tej, z jaką wygenerowano analogiczną hipotezę dla studenta Wydziału_Informatyki_studiów_stacjonarnych. □

Przykład 6.4. Gdyby w bazie wiedzy systemu ***R_2U*** umieścić reguły $R_1 - R_4$ w postaci identycznej jak w przykładzie 6.1 oraz regułę R_5 w nowej postaci:

$$R_5: \mathrm{grf}(0.85): \mathrm{Kier} = \{\mathrm{Inf}\} \odot, \\ \mathrm{Rok} = \neg\{1,2\} \oplus, \\ \mathrm{Plec} = \neg\{\mathrm{K}\} \odot \\ \rightarrow \mathrm{Sprzet} = \{\mathrm{K\_st}, \mathrm{Lap}\} \odot : \mathrm{irf}(0.75)$$

ze zmienioną wartością współczynnika irf w stosunku do pierwowzoru, to wówczas jedną z reguł R_4 i R_5 należałoby uznać za zredagowaną niepoprawnie (niespełnienie warunku zależności dodatniej monotonicznej między przesłanką a konkluzją reguły). □

Przykład 6.5. Gdyby w bazie wiedzy systemu $\boldsymbol{R\_2U}$ umieścić reguły $R_1 - R_5$ w postaci identycznej jak w przykładzie 6.1 i, dodatkowo, następującą regułę R_{3_1}:

$$R_{3\_1} : \mathrm{grf}(0.75) \ : \ \mathrm{Kier} = \{\mathrm{Inf}\} \odot$$
$$\rightarrow \mathrm{Sprzet} = \{\mathrm{K\_st}\} \odot : \mathrm{irf}(0.55)$$

to wówczas wystąpiłaby sprzeczność między regułami R_3 i R_{3_1} ($\mathrm{Sprzet} = \{\mathrm{K\_st}\} \odot \ \leq_{\mathrm{AS}} \mathrm{Sprzet} = \{\mathrm{K\_st}, \mathrm{Lap}\} \odot$ oraz $0.55 < 0.6$). Tę sprzeczność należałoby rozwiązać poprzez usunięcie z bazy przynajmniej jednej z tych reguł (np. R_3, ze względu na niższą wartość współczynnika grf). □

Przykład 6.6. Gdyby w bazie wiedzy systemu $\boldsymbol{R\_2U}$ umieścić reguły $R_1 - R_4$ w postaci identycznej jak w przykładzie 6.1 oraz regułę R_5 w nowej postaci:

$$R_5 \colon \mathrm{grf}(0.85) : \mathrm{Kier} = \{\mathrm{Inf}\} \odot ,$$
$$\mathrm{Rok} = \neg\{1,2\} \oplus ,$$
$$\mathrm{Plec} = \neg\{\mathrm{K}\} \odot$$
$$\rightarrow \mathrm{Sprzet} = \{\mathrm{K\_st}, \mathrm{Lap}\} \odot : \mathrm{irf}(0.75)$$

ze zmienioną wartością współczynnika irf w stosunku do pierwowzoru, to wówczas ta regułą, subsumowana w nowej sytuacji przez regułę R_4, stałaby się nadmiarowa. □

Przykład 6.7. Jeśli w bazie wiedzy systemu $\boldsymbol{R\_2U}$ zostaną umieszczone reguły $R_1 - R_1$ w postaci identycznej jak w przykładzie 6.1, to przy nowej początkowej zawartości bazy danych:

$$F_1^* \colon \mathrm{grf}(1.0) : \ \mathrm{true}$$
$$\rightarrow \mathrm{Wydz\_Rodz} = \{\mathrm{WInf\_st}\} \odot : \mathrm{irf}(1.0)$$

$$F_2^* \colon \mathrm{grf}(0.95) : \mathrm{true}$$
$$\rightarrow \mathrm{Rok} = \{1,2\} \oplus : \mathrm{irf}(0.0)$$

$$F_5^* \colon \mathrm{grf}(1.0) : \ \mathrm{true}$$
$$\rightarrow \mathrm{Plec} = \{\mathrm{K}\} \odot : \mathrm{irf}(0.1) ,$$

oraz niezmienionych wartościach progowych $\tau_1 = 0.7$, $\tau_2 = 0.7$ i stosunku wag hipotez wielokrotnych $g = 1.8$, proces wnioskowania będzie miał następujący przebieg:

– Krok1 i Krok2 zakończą się podobnym rezultatem, jak w trakcie wnioskowania z przykładu 6.1 (w bazie danych będzie się znajdował dodatkowy fakt F_5^*);
– w Kroku3 zostanie wykonana dodatkowa reguła R_5, która spowoduje wygenerowanie faktu cząstkowego F_{4_2}: $\mathrm{grf}(0.85)\colon \mathrm{true} \to \mathrm{Sprzet} = \{\mathrm{K\_st}, \mathrm{Lap}\} \odot : \mathrm{irf}(0.68)$, a następnie zastąpienie dotychczasowego faktu F_4 – nowym (zbiorczym) faktem F_4, w postaci: $\mathrm{grf}(0.9)\colon \mathrm{true} \to \mathrm{Sprzet} = \{\mathrm{K\_st}, \mathrm{Lap}\} \odot : \mathrm{irf}(0.65)$, w którym wartości p_{c4} i p_{r4} współczynników zmodyfikowanego faktu F_4 obliczono ze wzorów: $p_{c4} = \left(1 - \frac{1}{1.8+1}\right) \cdot 0.64 + + \frac{1}{1.8+1} \cdot 0.68 \cong 0.65$ oraz $p_{r4} = \max\{0.9, 0.85\} = 0.9$;
– w Kroku4 nastąpi wykonanie reguły R_1; w wyniku tego wykonania dotychczasowy fakt F_3 zostanie zastąpiony nowym, w postaci: $\mathrm{grf}(0.9) : \mathrm{true} \to \mathrm{Kier} = \{\mathrm{Inf}\} \odot : \mathrm{irf}(0.71)$, tak jak to miało miejsce w Kroku3 wnioskowania z przykładu 6.1;
– w Kroku5 nastąpi wykonanie reguły R_3; w wyniku zostanie wygenerowany fakt cząstkowy F_{4_3}: $\mathrm{grf}(0.7) : \mathrm{true} \to \mathrm{Sprzet} = \{\mathrm{K\_st}, \mathrm{Lap}\} \odot : \mathrm{irf}(0.43)$, a następnie dotychczasowy fakt c zostanie zastąpiony nowym (zbiorczym) faktem F_4, w postaci: $(0.9) : \mathrm{true} \to \mathrm{Sprzet} = \{\mathrm{K\_st}, \mathrm{Lap}\} \odot : \mathrm{irf}(0.61)$; wartości p_{c4} i p_{r4} współczynników tego zmodyfikowanego faktu F_4 zostały obliczone ze wzorów: $p_{c4} = 1 - \frac{\frac{1}{1.8+1}}{\frac{1}{1.8+1}+1.8} \cdot 0.65 + \frac{\frac{1}{1.8+1}}{\frac{1}{1.8+1}+1.8} \cdot 0.43 \cong 0.61$ oraz $p_{r4} = \max\{0.9, 0.7\} = 0.9$.

Reasumując, po zakończeniu procesu wnioskowania baza danych będzie miała następującą zawartość:

F_1^*: $\mathrm{grf}(1.0) :\ \mathrm{true}$
$\quad \to \mathrm{Wydz\_Rodz} = \{\mathrm{WInf\_st}\} \odot : \mathrm{irf}(1.0)$

F_2^*: $\mathrm{grf}(0.95) : \mathrm{true}$
$\quad \to \mathrm{Rok} = \{1{,}2\} \oplus : \mathrm{irf}(0.0)$

F_5^*: $\mathrm{grf}(1.0) :\ \mathrm{true}$
$\quad \to \mathrm{Plec} = \{\mathrm{K}\} \odot : \mathrm{irf}(0.1)$

F_3: $\mathrm{grf}(0.9) :\ \mathrm{true}$
$\quad \to \mathrm{Kier} = \{\mathrm{Inf}\} \odot : \mathrm{irf}(0.71)$

F_4: $\mathrm{grf}(0.9) :\ \mathrm{true}$
$\quad \to \mathrm{Sprzet} = \{\mathrm{K\_st}, \mathrm{Lap}\} \odot : \mathrm{irf}(0.61)$

Znajdujący się w tej bazie fakt zbiorczy F_4 jest wynikiem trzykrotnej generacji konkluzji: $\mathrm{Sprzet} = \{\mathrm{K\_st}, \mathrm{Lap}\} \odot$, ze współczynnikami grf, kolejno: 0.9, 0.85 i 0.7 oraz współczynnikami irf, kolejno: 0.64, 0.68 i 0.43. Wpływ trzech ostatnich wartości na wartość końcową współczynnika irf był zróżnicowany i zależny od wag konkluzji cząstkowych, te zaś – od wartości stałej proporcji $g = 1.8$. Poszczególne wagi wyniosły:

$$w_1 = 1 \cdot \left(1 - \frac{1}{1.8+1}\right) \cdot \left(1 - \frac{\frac{1}{1.8+1}}{\frac{1}{1.8+1}+1.8}\right) = \frac{324}{604} \cong 0.536;$$

$$w_2 = \frac{1}{1.8+1} \cdot \left(1 - \frac{\frac{1}{1.8+1}}{\frac{1}{1.8+1}+1.8}\right) = \frac{180}{604} \cong 0.298;$$

$$w_3 = \frac{\frac{1}{1.8+1}}{\frac{1}{1.8+1}+1.8} = \frac{100}{604} \cong 0.166.$$

Oczywiście, zachodzą zależności: $\frac{w_1}{w_2} = \frac{w_2}{w_3} = 1.8$ oraz $w_1 + w_2 + w_3 = 1$. □

Przykład 6.8. Jeśli w bazie wiedzy systemu $\boldsymbol{R\_2U}$ zostaną umieszczone reguły $R_1 - R_5$ w postaci identycznej jak w przykładzie 6.1 i, dodatkowo, reguła R_6, w postaci:

$$R_6 : \mathrm{grf}(0.7) \ : \ \mathrm{Wydz\_Rodz} = \{\mathrm{WInf\_st}\} \odot$$
$$\rightarrow \mathrm{Sprzet} = \{\mathrm{Lap}\} \odot : \mathrm{irf}(0.5) \ ,$$

to przy tej samej zawartości początkowej bazy danych (fakty F_1^* i F_2^*), niezmienionych wartościach progowych $\tau_1 = 0.7$, $\tau_2 = 0.7$ i stosunku wag hipotez wielokrotnych $g = 1.8$, proces wnioskowania będzie miał następujący przebieg:

- Krok1, Krok2 i Krok3 zakończą się identycznym rezultatem, jak w trakcie wnioskowania z przykładu 6.1;
- w Kroku4 do agendy zostaną zakwalifikowane reguły R_3 i R_6; wartość współczynnika wiarygodności grf obu reguł jest identyczna (0.7); w tej sytuacji, konflikt w agendzie będzie rozstrzygnięty na podstawie dodatkowego kryterium; jeśli jest nim prządek tekstowy reguł w bazie wiedzy, to ten konflikt będzie rozstrzygnięty na korzyść reguły R_3 poprzedzającej w zapisie regułę R_6;
- proces wnioskowania będzie kontynuowany, i w Kroku5 zostanie wykonana reguła R_6; w wyniku tego wykonania zostanie wygenerowany nowy fakt F_5,w postaci:

$$F_5 : \mathrm{grf}(0.7) \ : \ \mathrm{true}$$
$$\rightarrow \mathrm{Sprzet} = \{\mathrm{Lap}\} \odot : \mathrm{irf}(0.5),$$

który będzie pozostawał w sprzeczności ze znajdującym się tam faktem F_4:

$$F_4 : \mathrm{grf}(0.9) \ : \ \mathrm{true}$$
$$\rightarrow \mathrm{Sprzet} = \{\mathrm{K\_st, Lap}\} \odot : \mathrm{irf}(0.57).$$

Powyższa sprzeczność powinna zostać wykryta i wyeliminowana przez podsystem utrzymywania wiarygodności. Można to uczynić, między innymi, poprzez usunięcie z bazy danych faktu F_5, jako tego, który cechuje się mniejszą wiarygodnością $(0.7 < 0.9)$. □

Przykład 6.9. W bazie wiedzy systemu ***R_2U*** znajdują się reguły $R_1 - R_5$, a w bazie danych – fakty F_1^*, F_2^* i F_5^* w postaci identycznej jak w przykładzie 6.7. W opisanej sytuacji, uruchomienie procesu wnioskowania regresywnego dla pytania Q_1, w postaci:

$$Q_1\text{: (grf(0.85) : Sprzet = \{K\_st, Lap\}} \odot \text{ : irf(0.75)} \leftarrow \text{ true, } \geq)$$

spowoduje wykonanie następujących czynności:

Krok1:

$$\begin{aligned} Q_2\text{: (grf(0.85) : } & (\text{Kier} = \{\text{Inf}\} \odot), \\ & (\text{Rok} = \neg\{1,2\} \oplus) : \text{irf}(0.94) \leftarrow \text{ true, } \geq) \end{aligned}$$

Poszukiwanie w bazie danych faktu z konkluzją nie słabszą niż konkluzja $\text{Sprzet} = \{\text{K\_st, Lap}\} \odot$ kończy się porażką. W tej sytuacji jest inicjowany proces przeszukiwania bazy wiedzy pod kątem obecności w niej reguł z wymaganą konkluzją. W porządku tekstowym pierwsza taka regułą jest R_3:

$$\begin{aligned} R_3\text{: grf(0.7) } & : \text{Kier} = \{\text{Inf}\} \odot \\ & \rightarrow \text{Sprzet} = \{\text{K\_st, Lap}\} \odot : \text{irf}(0.6). \end{aligned}$$

Niestety, jej współczynniki wiarygodności nie spełniają obligatoryjnego warunku: $(0.6 \neq 0) \wedge (0.6 \geq 0.75) \wedge (0.7 \geq 0.85)$ (patrz podrozdz. 6.4.3). Kolejna reguła z wymagana konkluzją to R_4:

$$\begin{aligned} R_4\text{: grf(0.9) } & : \text{Kier} = \{\text{Inf}\} \odot, \\ & \quad \text{Rok} = \neg\{1,2\} \oplus, \\ & \rightarrow \text{Sprzet} = \{\text{K\_st, Lap}\} \odot : \text{irf}(0.8) \end{aligned}$$

W tym wypadku warunek dotyczący współczynników wiarygodności reguły $(0.8 \neq 0) \wedge (0.8 \geq 0.75) \wedge (0.9 \geq 0.85)$ jest spełniony, więc następuje generacja pytania pomocniczego Q_2, z hipotezą opartą na przesłance tej reguły. Generacja pytania Q_2 wiąże się z użyciem funkcji wstecznej propagacji niepewności f_{50}, f_{51} i f_{52}, odpowiadających za:

- ustawienie operatora pytania na $\geq$ – operator identyczny jak w pytaniu Q_1;
- obliczenie wartości p_{c2} współczynnika irf ze wzoru: $p_{c2} = \frac{0.75}{0.8} \cong 0.94$,
- ustawienie wartości p_{r2} współczynnika grf na: $p_{r2} = 0.85$.

Krok2:

$$\left\{ \begin{aligned} & Q_{3\_1}\text{: (grf(0.85) : } \text{Kier} = \{\text{Inf}\} \odot : \text{irf}(0.94) \leftarrow \text{ true}, \geq), \\ & Q_{3\_2}\text{: (grf(0.85) : } \text{Rok} = \neg\{1,2\} \oplus : \text{irf}(0.94) \leftarrow \text{ true}, \geq) \end{aligned} \right\}$$

Pytanie Q_2 jest transformowane do zbioru pytań pomocniczych $\{Q_{3\_1}, Q_{3\_2}\}$, z których każde reprezentuje jedną hipotezę składową hipotezy złożonej pochodzącej z Q_2. Niniejsza transformacja jest wykonywana przy wykorzystaniu funkcji wstecznej propagacji niepewności f_{60i}, f_{6ij} $(1 \leq i, j \leq 2)$, odpowiadających za:

- $\mathrm{f}_{601}, \mathrm{f}_{602}$ – ustawienie operatora pytań Q_{3_1} i Q_{3_2} na $\geq$ – operator identyczny jak w pytaniu Q_2;
- $\mathrm{f}_{611}, \mathrm{f}_{612}$ – ustawienie wartości p_{c3_1} i p_{r3_1} współczynników irf i grf pytania Q_{3_1} na identyczne jak w pytaniu Q_2;
- $\mathrm{f}_{621}, \mathrm{f}_{622}$ – ustawienie wartości p_{c3_2} i p_{r3_2} współczynników irf i grf pytania Q_{3_2} na identyczne jak w pytaniu Q_2.

Krok3:

$$
\begin{aligned}
Q_4\text{: }(\mathrm{grf}(0.85) : &(\mathrm{Kier} = \{\mathrm{Inf}\} \odot)\, ,\\
&(\mathrm{Rok} = \neg\{1{,}2\} \oplus)\, ,\\
&(\mathrm{Plec} = \neg\{\mathrm{K}\} \odot) : \mathrm{irf}(0.88) \leftarrow \mathrm{true},\ \geq)
\end{aligned}
$$

Poszukiwanie w bazie danych faktu z konkluzją nie słabszą niż $\mathrm{Kier} = \{\mathrm{Inf}\} \odot$ kończy się porażką. Niestety, porażką kończy się także poszukiwanie w bazie wiedzy odpowiedniej reguły z taką konkluzją (żadna z reguł R_1 i R_2 nie spełnia warunku dotyczącego współczynników wiarygodności). W tej sytuacji następuje w prowadzonym procesie abdukcyjnym nawrót do najbliższego punktu wnioskowania, w którym można dokonać wyboru alternatywnej reguły. Jest nim punkt wcześniejszego wyboru reguły R_4. Zamiast niej, silnik wnioskujący rozpatruje w tym kroku regułę R_5:

$$
\begin{aligned}
R_5\text{: }\mathrm{grf}(0.85) : &\ \mathrm{Kier} = \{\mathrm{Inf}\} \odot\, ,\\
&\ \mathrm{Rok} = \neg\{1{,}2\} \oplus\, ,\\
&\ \mathrm{Plec} = \neg\{\mathrm{K}\} \odot\\
\rightarrow &\ \mathrm{Sprzet} = \{\mathrm{K\_st}, \mathrm{Lap}\} \odot : \mathrm{irf}(0.85)
\end{aligned}
$$

Niniejsza reguła spełnia warunek dotyczący współczynników wiarygodności, więc ma miejsce generacja nowego pytania pomocniczego Q_4. Procedura wyznaczania operatora pytania Q_4 i współczynników wiarygodności irf i grf jego hipotezy ma podobny przebieg jak w Kroku1.

Krok4

$$
\left\{
\begin{aligned}
&Q_{5\_1}\text{: }(\mathrm{grf}(0.85) :\ \mathrm{Kier} = \{\mathrm{Inf}\} \odot : \mathrm{irf}(0.88) \leftarrow \mathrm{true}, \geq),\\
&Q_{5\_2}\text{: }(\mathrm{grf}(0.85) :\ \mathrm{Rok} = \neg\{1{,}2\} \oplus : \mathrm{irf}(0.88) \leftarrow \mathrm{true}, \geq),\\
&Q_{5\_3}\text{: }(\mathrm{grf}(0.85) :\ \mathrm{Plec} = \neg\{\mathrm{K}\} \odot : \mathrm{irf}(0.88) \leftarrow \mathrm{true}, \geq)
\end{aligned}
\right\}
$$

Ze względu na złożoną postać hipotezy zawartej w pytaniu Q_4, następuje w tym punkcie transformacja Q_4 do zbioru pytań pomocniczych $\{Q_{5\_1}, Q_{5\_2}, Q_{5\_3}\}$. Ma ona analogiczny przebieg jak w Kroku2.

Krok5

Niestety, zarówno poszukiwanie w bazie danych odpowiedniego faktu z konkluzją nie słabszą niż Kier = $\{\text{Inf}\}\odot$, jak i poszukiwanie w bazie wiedzy odpowiedniej reguły z taką konkluzją kończy się porażką. W tej sytuacji, przy braku możliwości wykonania kolejnego nawrotu, proces wnioskowania kończy się udzieleniem odpowiedzi false na pytanie wyjściowe Q_1. □

Przykład 6.10. W bazie wiedzy systemu $\boldsymbol{R\_2U}$ znajdują się następujące reguły $R_1 - R_5$, w których zaszyta jest wiedza na temat zależności pomiędzy:

– wynikami osiągniętymi przez studenta na ćwiczeniach laboratoryjnych (Lab_SI), kolokwium (Kol_SI) i teście egzaminacyjnym (Test_SI) ze sztucznej inteligencji (SI) oraz
– poziomem obecności i aktywności studenta na wykładach z SI (Wykl_SI) a poziomem jego wiedzy z SI, wyrażonym oceną z egzaminu z SI (Egz_SI):

$$\begin{aligned} R_1\colon \mathrm{grf}(1.0)\ :\ & \mathrm{Test\_SI} = \{\mathrm{z}\}\odot \\ & \rightarrow \mathrm{Egz\_SI} = \{+\}\odot : \mathrm{irf}(0.75) \end{aligned}$$

$$\begin{aligned} R_2\colon \mathrm{grf}(0.9)\ :\ & \mathrm{Egz\_SI} = \{+\}\odot, \\ & \mathrm{Lab\_SI} = \neg\{\mathrm{z}\}\odot \\ & \rightarrow \mathrm{Ocena\_posr} = \{\mathrm{sciaga}\}\odot : \mathrm{irf}(0.9) \end{aligned}$$

$$\begin{aligned} R_3\colon \mathrm{grf}(0.7)\ :\ & \mathrm{Ocena\_posr} = \{\mathrm{sciaga}\}\odot, \\ & \mathrm{Wykl\_SI} = \neg\{\mathrm{obecn}\}\odot \\ & \rightarrow \mathrm{Egz\_SI} = \neg\{+\}\odot : \mathrm{irf}(0.95) \end{aligned}$$

$$\begin{aligned} R_4\colon \mathrm{grf}(0.9)\ :\ & \mathrm{Egz\_SI} = \neg\{+\}\odot, \\ & \mathrm{Wykl\_SI} = \{\mathrm{aktywn}\}\odot \\ & \rightarrow \mathrm{Ocena\_posr} = \{\mathrm{watpl}\}\odot : \mathrm{irf}(1.0) \end{aligned}$$

$$\begin{aligned} R_5\colon \mathrm{grf}(0.7)\ :\ & \mathrm{Ocena\_posr} = \{\mathrm{watpl}\}\odot, \\ & \mathrm{Kol\_SI} = \neg\{\mathrm{z}\}\odot \\ & \rightarrow \mathrm{Sprzet} = \{\mathrm{K\_st}, \mathrm{Lap}\}\odot : \mathrm{irf}(0.85)\ , \end{aligned}$$

gdzie literał z oznacza fakt zaliczenia kolokwium/ćwiczeń laboratoryjnych/testu egzaminacyjnego, $+$ – fakt zdania egzaminu, sciaga – przekonanie egzaminatora o użyciu przez studenta ściągi na teście egzaminacyjnym, watpl – wątpliwość egzaminatora odnośnie do poziomu wiedzy studenta, obecn – częstą obecność na wykładach z SI, aktywn – dużą aktywność na wykładach z SI.

Jeśli w powyższym systemie, przy stanie bazy danych:

$$\begin{aligned} F_1^*\colon \mathrm{grf}(1.0)\ :\ & \mathrm{true} \\ & \rightarrow \mathrm{Test\_SI} = \{\mathrm{z}\}\odot : \mathrm{irf}(1.00) \end{aligned}$$

$$\begin{aligned} F_2^*\colon \mathrm{grf}(1.0)\ :\ & \mathrm{true} \\ & \rightarrow \mathrm{Kol\_SI} = \{\mathrm{z}\}\odot : \mathrm{irf}(0.4) \end{aligned}$$

F_3^*: grf(1.0) : true
→ Lab_SI = {z} ⊙ : irf(0.1)

F_4^*: grf(1.0) : true
→ Wykl_SI = {obecn} ⊙ : irf(0.2)

F_5^*: grf(1.0) : true
→ Wykl_SI = {aktywn} ⊙ : irf(0.6)

parametrach progowych $\tau_1 = 0.65$, $\tau_2 = 0.5$ oraz stosunku wag hipotez wielokrotnych $g = 0.5$ zostanie zainicjowane wnioskowanie w trybie progresywnym, to w kolejnych krokach wnioskowania baza danych będzie zawierać następujące fakty prócz aksjomatycznych $F_1^* - F_5^*$:

Krok1

F_6: grf(1.0) : true
→ Egz_SI = {+} ⊙ : irf(0.75)

(nowy fakt F_6 – wynik wykonania reguły R_1 z użyciem funkcji propagacji niepewności f_{21} i f_{22})

Krok2

F_6: grf(1.0) : true
→ Egz_SI = {+} ⊙ : irf(0.75)

F_7: grf(0.9) : true
→ Ocena_posr = {sciaga} ⊙ with irf(0.68)

(nowy fakt F_7 - wynik użycia funkcji propagacji niepewności f_{41}, f_{42}, f_{11} i f_{12} oraz wykonania reguły R_2 z użyciem funkcji propagacji niepewności f_{21} i f_{22})

Krok3

F_6: grf(1.0) : true
→ Egz_SI = {+} ⊙ : irf(0.48)

F_7: grf(0.9) : true
→ Ocena_posr = {sciaga} ⊙ with irf(0.68)

(nowa postać faktu F_6 - wynik użycia funkcji propagacji niepewności f_{41}, f_{42}, f_{11} i f_{12} oraz wykonania reguły R_3 z użyciem funkcji propagacji niepewności f_{41}, f_{42}, f_{31} i f_{32}; ze względu na obniżenie wartości współczynnika irf hipotezy poniżej wymaganego $\tau_2 = 0.5$, reguła R_2 nie spełnia chwilowo drugiego z wymaganych warunków kwalifikacji do agendy)

Krok4

F_6: grf(1.0) : true
→ Egz_SI = {+} ⊙ : irf(0.48)

F_7: grf(0.9) : true
$\rightarrow$ Ocena_posr = {sciaga} $\odot$ with irf(0.68)

F_8: grf(0.9) : true
$\rightarrow$ Ocena_posr = {watpl} $\odot$ with irf(0.52)

(nowy fakt F_8 - wynik użycia funkcji propagacji niepewności f_{41}, f_{42}, f_{11} i f_{12} oraz wykonania reguły R_5 z użyciem funkcji propagacji niepewności f_{41}, f_{42}, f_{31} i f_{32}; ze względu na podwyższenie wartości współczynnika irf hipotezy powyżej wymaganego progu $\tau_2 = 0.5$, reguła R_2 odzyska w kolejnym kroku kwalifikację do agendy)

Krok6

F_6: grf(1.0) : true
$\rightarrow$ Egz_SI = {+} $\odot$: irf(0.54)

F_7: grf(0.9) : true
$\rightarrow$ Ocena_posr = {sciaga} $\odot$ with irf(0.55)

F_8: grf(0.9) : true
$\rightarrow$ Ocena_posr = {watpl} $\odot$ with irf(0.52)

(nowa postać faktu F_7 - wynik użycia funkcji propagacji niepewności f_{41}, f_{42}, f_{11} i f_{12} oraz wykonania reguły R_2 z użyciem funkcji propagacji niepewności f_{31} i f_{32}). □

Przykład 6.10 ma jedynie znaczenie ilustracyjne. Warto jednak odnotować specyficzny charakter użytej w nim bazy wiedzy oraz specyficzny stosunek wag hipotez wielokrotnych $g = 0.5$ zastosowany we wnioskowaniu. Specyfika bazy wiedzy polega na występowaniu cykli: formuła atomowa Egz_SI = {+} $\odot$ pełni w regułach zarówno rolę przesłanki (R_2 i R_4), jak i konkluzji (R_1, R_3 i R_5). Taka baza wiedzy może być wykorzystana do programowania działań układów ze sprzężeniem zwrotnym. Stosunek $g < 1$ oznacza, że waga generowanych hipotez wielokrotnych będzie rosła wraz z postępem procesu wnioskowania. Przy g dążącym do 0, $g \rightarrow 0$, bieżąco wygenerowana hipoteza wielokrotna stanie się hipotezą zbiorczą. Taka wartość stosunku g może być przydatna w systemach $\boldsymbol{R\_2U}$ przeznaczonych do sterowania procesami dynamicznymi.

Rozdział 7

Indukcja reguł 2U

7.1 Wstęp

Siła ekspresji reguł 2U ma dwa niezależne źródła. Pierwszym z nich są współczynniki wiarygodności irf i grf, gwarantujące możliwość wyrażania niepewności drugiego rzędu. Drugie źródło to uogólniona postać formuł atomowych, dopuszczających stosowanie wartości zbiorowych (w miejsce wartości pojedynczych) oraz operatorów relacyjnych działających na tych wartościach. Potrzeba posługiwania się niepewnością drugiego rzędu została uzasadniona w podrozdziale 3.9. Czy równie celowe jest rozszerzenie postaci formuł atomowych? Także i na to pytanie należy odpowiedzieć twierdząco. Wystarczy wskazać na przydatność takich formuł w medycznych systemach wspomagania decyzji, czy systemach monitorowania i nadzoru. Przykładowo, przy ocenie postępowania leczniczego podejmowanego w ataku astmy oskrzelowej niezbędna jest znajomość nie jednej, lecz wszystkich istotnych chorób współtowarzyszących, w szczególności – cukrzycy i schorzeń kardiologicznych. Z kolei, sterowanie pracą linii produkcyjnych wymaga ciągłego testowania parametrów technologicznych pod kątem ich przynależności do wymaganych zakresów wartości. I w jednym, i w drugim wypadku do reprezentacji danych potrzebne są zbiory wartości.

Aby móc w pełni skorzystać z możliwości oferowanych przez reguły 2U, należy zapewnić efektywność i wydajność projektowanego systemu $\boldsymbol{R\_2U}$. Przedstawione w podrozdziale 5.2 rozważania na temat jakości systemu regułowego z niepewnością mają charakter ogólny i abstrahują od szczegółowej metody reprezentacji wiedzy w systemie. Zaproponowana tam metryka jakości 5.1 ma więc odniesienie także do systemu $\boldsymbol{R\_2U}$. Z tego wynika, że przy konstrukcji bazy wiedzy systemu $\boldsymbol{R\_2U}$ należy zmierzać do pozyskania możliwie największego, wewnętrznie niesprzecznego i nieredundantnego zbioru reguł 2U cechujących się wysoką wiarygodnością grf. W tym celu trzeba proces indukcji reguł oprzeć na dużym zbiorze danych trenujących. Zważywszy fakt, że ten proces ma być zautomatyzowany, należy dodatkowo zapewnić jednorodność powyższego zbioru.

W ogólności, powyższe wymagania spełniają dane atrybutowe i uniwersalny format reprezentacji tych danych. Jest nim format danych zbiorczych, w którym można zapisywać także dane indywidualne. W większości wypadków, dziedzinowe dane atrybutowe pochodzą z różnych źródeł, a ich zapis oparty jest na różnych schematach danych. Jeśli jednak każdy z tych schematów da się sprowadzić przy zachowaniu semantyki atrybutów do pewnego wskazanego schematu referencyjnego, to całemu zbiorowi danych dziedzinowych można nadać jednorodną postać.

7.2 Dane atrybutowe zbiorcze

7.2.1 Składnia i semantyka danych

Niech $\mathcal{D}_{\mathcal{U}}$, $\mathcal{T}_{\mathcal{DU}}$, $C_{\mathcal{TDU}} = \{\mathrm{C_{TDu1}}, \mathrm{C_{TDu2}}, \dots, \mathrm{C_{TDu}}_n\}$, $\mathbf{A_S} = \{\mathrm{A_{s1}}, \mathrm{A_{s2}}, \dots, \mathrm{A}_{sn}\}$ oznaczają, odpowiednio: dziedzinę problemową, wybraną konceptualizację dziedziny, skończony zbiór kategorii użytych w tej konceptualizacji i odpowiadający mu skończony zbiór atrybutów zbiorowych, tak jak w podrozdziale 6.2.

Definicja 7.1. Schematem danych atrybutowych (indywidualnych) ze zbiorami kwalifikowanymi w dziedzinie $\mathcal{D}_{\mathcal{U}}$, zgodnym z konceptualizacją $\mathcal{T}_{\mathcal{DU}}$, nazywamy dowolny niepusty podzbiór $\mathbf{S_S} = \{\mathrm{A}_{si1}, \mathrm{A}_{si2}, \dots, \mathrm{A}_{sim}\}$ zbioru atrybutów $\mathbf{A_S}$.

Definicja 7.2. Daną atrybutową (indywidualną) ze zbiorami kwalifikowanymi, zbudowaną według schematu $\mathbf{S_S} = \{\mathrm{A}_{si1}, \mathrm{A}_{si2}, \dots, \mathrm{A}_{sim}\}$, nazywa się zbiór par uporządkowanych w postaci:

$$\{(\mathrm{A}_{si1}, neg_{i1}vq_{i1}), (\mathrm{A}_{si2}, neg_{i2}vq_{i2}) \dots, (\mathrm{A}_{sim}, neg_{im}vq_{im})\} \tag{7.1}$$

w którym vq_{ik} $(1 \leq k \leq m)$ oznacza dowolny kwalifikowany zbiór wartości – przykładów kategorii $C_{\mathrm{TDu}ik}$, a neg_{ik} $(1 \leq k \leq m)$ – opcjonalny operator negacji ($neg_{ik} \in \{\varepsilon, \neg\}$). Do zapisu powyższego zbioru stosuje się uproszczoną notację krotkową $\langle \mathrm{A}_{si1}: neg_{i1}vq_{i1};\ \mathrm{A}_{si2}: neg_{i2}vq_{i2};\ \dots; \mathrm{A}_{sim}: neg_{im}vq_{im}\rangle$ z dowolnym porządkiem wyszczególniania par $(\mathrm{A}_{sik}, neg_{ik}vq_{ik})$.

Formalny opis danych atrybutowych ze zbiorami kwalifikowanymi jest możliwy na gruncie logiki atrybutowej ALSV(FD) (ang. *Attributive Logic with Set Values*) [105]. Tak jak w przypadku logik atrybutowych AAL i VAAL, alfabet tej logiki jest sumą rozłącznych zbiorów $\mathbf{O}$, $\mathbf{A_S}$ i $\mathbf{V}$ zawierających, odpowiednio: nazwy obiektów, nazwy atrybutów i wartości atrybutów. Jeśli przyjąć, że zbiór nazw atrybutów ma postać $\mathbf{A_S} = \{\mathrm{A_{s1}}, \mathrm{A_{s2}}, \dots, \mathrm{A}_{sn}\}$, a $\mathbf{V} = \mathbf{V_1} \cup \mathbf{V_2} \cup \dots \cup \mathbf{V}_n$ jest sumą dziedzin wartościowania poszczególnych atrybutów ze zbioru $\mathbf{A_S}$, to w logice ALSV(FD) każdy atrybut A_{si} należy postrzegać jako funkcję: $\mathrm{A}_{si}: \mathbf{O} \to \mathbf{2}^{\mathbf{V}_i} \times \{=, \in, \subseteq, \supseteq, \sim\}$, stanowiącą odpowiednik bazodanowego operatora projekcji.

Formułą atomową logiki ALSV(FD) jest każde i tylko takie wyrażenie, które ma postać: $\mathrm{A}_{si}(O)\ rel\ v_i$ lub $\neg(\mathrm{A}_{si}(O)\ rel\ v_i)$, gdzie $\mathrm{A}_{si} \in \mathbf{A}$, $O \in \mathbf{O}$, $v_i \in \mathbf{2}^{\mathbf{V}_i}$, a $rel \in \{=,\ \in, \subseteq, \supseteq, \sim\}$. Formuły złożone logiki ALSV(FD) powstają przez zastosowanie do formuł atomowych spójników: koniunkcji $\wedge$, alternatywy $\vee$, implikacji $\Rightarrow$ i równoważności $\Leftrightarrow$ oraz kwantyfikatorów: ogólnego $\forall$ i szczegółowego $\exists$.

Danej atrybutowej 7.1 odpowiada na gruncie logiki ALSV(FD) formuła:

$$neg_{i1}(\mathrm{A}_{si1}(O)\ rel_{i1}\ v_{i1}) \wedge neg_{i2}(\mathrm{A}_{si2}(O)\ rel_{i2}\ v_{i2}) \wedge \ldots \wedge neg_{im}(\mathrm{A}_{sim}(O)\ rel_{im}\ v_{im}) \tag{7.2}$$

w której postać operatora rel_{ik} $(1 \leq k \leq m)$ zależy od postaci kwalifikatora zbioru kwalifikowanego vq_{ik}:

– jeśli kwalifikator zbioru jest równy $\oplus$, to operator w formule ma postać $\in$,
– jeśli kwalifikator zbioru jest równy $\odot$, to ma on postać $\supseteq$.

Z powyższego wynika, że do interpretacji danych atrybutowych ze zbiorami kwalifikowanymi wystarczą te formuły atomowe ALSV(FD), które są formułami klasycznymi w myśl definicji 6.2.

Definicja 7.3. Schematem danych atrybutowych zbiorczych w dziedzinie $\mathcal{D}_{\mathcal{U}}$, zgodnym z konceptualizacją $\mathcal{T}_{\mathcal{DU}}$, nazywamy dowolną parę uporządkowaną $(\mathbf{S_{S1}},\ \mathbf{S_{S2}})$ podzbiorów $\mathbf{S_{S1}}$ i $\mathbf{S_{S2}}$ zbioru atrybutów $\mathbf{A_S}$, takich że:

– $\mathbf{S_{S2}} \neq \emptyset$,
– $\mathbf{S_{S1}} \cap \mathbf{S_{S2}} = \emptyset$.

Definicja 7.4. Daną atrybutową zbiorczą zbudowaną według schematu $(\mathbf{S_{S1}},\ \mathbf{S_{S2}})$, gdzie $\mathbf{S_{S1}} = \{\mathrm{A}_{si1}, \ldots, \mathrm{A}_{sim1}\}$ i $\mathbf{S_{S2}} = \{\mathrm{A}_{sj1}, \ldots, \mathrm{A}_{sjm2}\}$, nazywa się zbiór par uporządkowanych w postaci:

$$\left\{ \begin{array}{l} (\mathrm{A}_{si1},\ neg_{i1}vq_{i1}|d_i),\ \ldots\ ,\ (\mathrm{A}_{sim1}, neg_{im1}vq_{im1}|d_i), \\ \left(\mathrm{A}_{sj1}, neg_{j1}vq_{j1}|d_{j1}\right), \ldots, \left(\mathrm{A}_{sjm2}, neg_{jm2}vq_{jm2}|d_{jm2}\right) \end{array} \right\}, \tag{7.3}$$

w którym vq_{ik} $(1 \leq k \leq m_1)$ oraz vq_{jl} $(1 \leq l \leq m_2)$ oznaczają dowolne kwalifikowane zbiory wartości – przykładów kategorii, odpowiednio, $\mathrm{C}_{\mathrm{TDu}ik}$ oraz $\mathrm{C}_{\mathrm{TDu}jl}$, neg_{ik} $(1 \leq k \leq m_1)$ oraz neg_{jl} $(1 \leq l \leq m_2)$ – opcjonalne operatory negacji $(neg_{ik}, neg_{jl} \in \{\varepsilon, \neg\})$, a $d_i \in \mathbf{N}$ i d_{jl} $(1 \leq l \leq m_2) \in (\mathbf{N} \cup \{0\})$ – liczności odnoszące się do zbiorów, odpowiednio vq_{ik} $(1 \leq k \leq m_1)$ i vq_{jl} $(1 \leq l \leq m_2)$. Powyższe liczności muszą spełniać warunek: $\forall d_{jl}(1 \leq l \leq m_2)\left(d_{jl} \leq d_i\right)$.

W dalszym ciągu, do zapisu powyższego zbioru stosuje się uproszczoną notację krotkową:

$$< \mathrm{A}_{si1}{:}\, neg_{i1}vq_{i1}|d_i; \ldots;\ \mathrm{A}_{sim1}{:}\, neg_{im1}vq_{im1}|d_i; \\ \mathrm{A}_{sj1}{:}\, neg_{j1}vq_{j1}\big|d_{j1} \ldots;\ \mathrm{A}_{sjm2}{:}\, neg_{jm2}vq_{jm2}\big|d_{jm2}\ >, \tag{7.4}$$

z dowolnym porządkiem wyszczególniania par w pierwszej części krotki (pary z indeksem i, w postaci $\mathrm{A}_{sik}: neg_{ik} vq_{ik} | d_i$) oraz dowolnym porządkiem wyszczególniania par w drugiej części krotki (pary z indeksem j, w postaci $\mathrm{A}_{sjl}: neg_{jl} vq_{jl} | d_{jl}$). Atrybuty z indeksem i nazywają się kluczowymi, a atrybuty z indeksem j – niekluczowymi. Dla zapewnienia przejrzystości zapisu, niezależnie od postaci kwalifikowanych zbiorów wartości, wszystkie pary uporządkowane są specyfikowane jawnie.

Do określenia sekwencji w postaci $neg\ vq|d$, gdzie neg oznacza potencjalny operator negacji, $neg \in \{\varepsilon, \neg\}$, vq – zbiór kwalifikowany, a d – liczność odnoszącą się do tego zbioru, $d \in (\mathbf{N} \cup \{0\})$, używa się w dalszym ciągu terminu „zbiór kwalifikowany z licznością".

Dana atrybutowa zbiorcza w postaci 7.3 „mieści" w sobie d_i danych atrybutowych indywidualnych ze zbiorami kwalifikowanymi, w postaci:

$$< \mathrm{A}_{si1}: neg_{i1k} vq_{i1k}; \dots; \mathrm{A}_{sim1}: neg_{im1k} vq_{im1k}; \\ \mathrm{A}_{sj1}: neg_{j1k} vq_{j1k}; \dots; \mathrm{A}_{sjm2}: neg_{jm2k} vq_{jm2k} > \tag{7.5}$$

gdzie $k = 1, \dots, d_i$, spełniających zależności:

– $\forall(1 \leq l \leq m_1)$ $(\mathrm{A}_{sil} \blacksquare_{il} vq_{il}) \leq_{\mathrm{AS}} (\mathrm{A}_{sil} \blacksquare_{ilk} vq_{ilk})$, gdzie $\blacksquare_{il}$ $(\blacksquare_{ilk})$ oznacza symbol relacyjny $=$, gdy neg_{il} (neg_{ilk}) jest operatorem pustym ε, lub symbol relacyjny $\neq$, gdy neg_{il} (neg_{ilk}) jest operatorem negacji ¬ ;
– $\forall(1 \leq l \leq m_2)$ $\left(vq_{jl} =_{\mathrm{id}} vq_{jlk}\right)$, gdzie $=_{\mathrm{id}}$ jest symbolem relacji identyczności,
– $\forall(1 \leq l \leq m_2)\left(\sum_{k=1}^{d_i}\left(neg_{jl} =_{\mathrm{id}} neg_{jlk}\ ?\ 1 : 0\right) = d_{jl}\right)$, gdzie $\left(neg_{jl} =_{\mathrm{id}} neg_{jlk}\ ?\ 1 : 0\right)$ oznacza operator warunkowy dający w wyniku:
  – 1, gdy wartością wyrażenia relacyjnego $neg_{jl} =_{\mathrm{id}} neg_{jlk}$ jest true,
  – 0, w przeciwnym wypadku.

Podsumowując, daną atrybutową zbiorczą można uzyskać przez złączenie takich danych (indywidualnych) ze zbiorami kwalifikowanymi, które są: zbudowane według tego samego schematu, parami do siebie „podobne" na odpowiadających sobie atrybutach kluczowych (podobieństwo definiowane za pomocą relacji $\leq_{\mathrm{AS}}$), oraz identyczne lub różniące się tylko występowaniem operatora negacji na odpowiadających sobie atrybutach niekluczowych.

Dane atrybutowe zbiorcze uzyskuje się w wyniku rozmaitych badań statystycznych, prowadzonych na wyselekcjonowanych grupach podobnych obiektów (zwłaszcza – osób) w celu zweryfikowania/sfalsyfikowania pewnych ich właściwości, stanów lub zachowań. Przykładem są szeroko rozpowszechnione badania kliniczne, których celem jest zbadanie skuteczności określonych terapii farmakologicznych w leczeniu wskazanych chorób lub zaburzeń. Dobór obiektów do tych badań odbywa się w sposób rygorystyczny: każda cecha kluczowa obiektu, reprezentowana atrybutem kluczowym danej zbiorczej, musi mieć ustaloną, z góry narzuconą wartość, np. określona płeć pacjenta, wiek pochodzący z zadanego przedziału wiekowego, te same choroby główne i towarzyszące. Atrybuty niekluczowe reprezentują cechy potencjalnie różnicujące. Dla każdej z nich określa się pewną charakterystyczną wartość, której występowanie jest przedmiotem badania klinicznego. Liczność odpowiedniego kwalifikowanego zbioru wartości (d_{jl}, $1 \leq l \leq m_2$, dla zbiorów wartości wewnątrz danej w postaci 7.3) określa liczbę tych spośród wszystkich poddanych badaniu obiektów (d_i dla danej w postaci 7.3), które osiągnęły tę charakterystyczną wartość. Przykłady danych atrybutowych zbiorczych uzyskanych w wyniku badań klinicznych można znaleźć m.in. w pracy [80].

Wiarygodne dane atrybutowe zbiorcze mają wyjątkową wartość. Jeśli taka dana reprezentuje odpowiednio dużą liczbę danych indywidualnych, to na jej podstawie można zaprojektować regułę 2U/zbiór reguł 2U cechujących się odpowiednią wiarygodnością. Jeśli natomiast liczba tych danych indywidualnych jest zbyt mała, to można najpierw podjąć próbę integracji danej atrybutowej zbiorczej z innymi podobnymi do niej danymi, a następnie proces indukcji reguł 2U przeprowadzić w odniesieniu do danej zintegrowanej. Powyższy proces integracji danych atrybutowych zbiorczych opiera się na wielokrotnym wykonywaniu dla nich operacji sumy lub iloczynu, zdefiniowanych w podrozdziale 7.2.2.

7.2.2 Algebra danych

Danym atrybutowym zbiorczym można nadać interpretację algebraiczną. Formalną definicję algebry takich danych (def. 7.11) poprzedza szereg definicji pomocniczych i lematów.

Definicja 7.5. Niech vq_i oraz vq_j oznaczają dwa zbiory kwalifikowane zbudowane w oparciu o dziedzinę **V**, w postaci, odpowiednio: $vq_i = \{v_{i1}, v_{i2}, \dots, v_{iki}\}qlr_i$ i $vq_j = \{v_{j1}, v_{j2}, \dots, v_{jkj}\}qlr_j$, gdzie v_{il} $(1 \leq l \leq k_i)$, v_{jl} $(1 \leq l \leq k_j) \in \mathbf{V}$, qlr_i, $qlr_j \in \{\oplus, \odot\}$. Zbiór kwalifikowany vq_i jest podzbiorem zbioru kwalifikowanego vq_j w dziedzinie **V**, $vq_i \subseteq_{\mathrm{q},\mathbf{V}} vq_j$, wtedy i tylko wtedy, gdy:

- $qlr_i = qlr_j = \oplus$ i $\{v_{j1}, v_{j2}, \dots, v_{jkj}\} \subseteq \{v_{i1}, v_{i2}, \dots, v_{iki}\}$ lub
- $qlr_i = qlr_j = \odot$ i $\{v_{i1}, v_{i2}, \dots, v_{iki}\} \subseteq \{v_{j1}, v_{j2}, \dots, v_{jkj}\}$ lub
- $qlr_i = \oplus$ i $qlr_j = \odot$ i $k_i = k_j = 1$ i $v_{i1} = v_{j1}$ lub
- $\exists_{vq_k}(vq_i \subseteq_{\mathrm{q},\mathbf{V}} vq_k) \wedge (vq_k \subseteq_{\mathrm{q},\mathbf{V}} vq_j)$,

gdzie $\subseteq$ oznacza klasyczny teoriomnogościowy operator zawierania się zbiorów.

Ponadto, zanegowany zbiór kwalifikowany $\neg vq_i$ jest podzbiorem zanegowanego zbioru kwalifikowanego $\neg vq_j$ w dziedzinie **V**, $\neg vq_i \subseteq_{\mathrm{q},\mathbf{V}} \neg vq_j$, wtedy i tylko wtedy, gdy:

– $qlr_i = qlr_j = \oplus$ i $\{v_{i1}, v_{i2}, \ldots, v_{iki}\} \subseteq \{v_{j1}, v_{j2}, \ldots, v_{jkj}\}$ lub
– $qlr_i = qlr_j = \odot$ i $\{v_{j1}, v_{j2}, \ldots, v_{jkj}\} \subseteq \{v_{i1}, v_{i2}, \ldots, v_{iki}\}$ lub
– $qlr_i = \odot$ i $qlr_j = \oplus$ i $k_i = k_j = 1$ i $v_{i1} = v_{j1}$ lub
– $\exists_{\neg vq_k}(\neg vq_i \subseteq_{\mathrm{q},\mathbf{V}} \neg vq_k) \wedge (\neg vq_k \subseteq_{\mathrm{q},\mathbf{V}} \neg vq_j)$.

Relacja bycia podzbiorem $\subseteq_{\mathrm{q},\mathbf{V}}$ nie zachodzi pomiędzy żadnymi zbiorem kwalifikowanym i zanegowanym zbiorem kwalifikowanym.

Definicja 7.6. Niech $\mathbf{Q_V}$ oznacza zbiór wszystkich kwalifikowanych zbiorów z licznością dających się zbudować w oparciu o dziedzinę $\mathbf{V}$, a $neg_i vq_i|d_i$ i $neg_j vq_j|d_j$ oznaczają dwa dowolne zbiory kwalifikowane z licznością pochodzące z $\mathbf{Q_V}$. Zbiór z licznością $neg_i vq_i|d_i$ jest podzbiorem zbioru z licznością $neg_j vq_j|d_j$ w dziedzinie $\mathbf{V}$, $neg_i vq_i|d_i \subseteq_{\mathrm{qd},\mathbf{V}} neg_j vq_j|d_j$, wtedy i tylko wtedy, gdy spełniona jest zależność:

– $neg_i vq_i \subseteq_{\mathrm{q},\mathbf{V}} neg_j vq_j$.

Liczności zbiorów kwalifikowanych nie mają więc wpływu na zachodzenie relacji $\subseteq_{\mathrm{qd},\mathbf{V}}$. Tym samym, dwa dowolne zbiory z licznością $neg\, vq|d_i$ oraz $neg\, vq|d_j$ są nierozróżnialne (tożsame) względem relacji $\subseteq_{\mathrm{qd},\mathbf{V}}$. Z definicji 7.5 i 7.6 i twierdzenia 6.1 wynika oczywisty lemat 7.1.

Lemat 7.1. Relacja $\subseteq_{\mathrm{qd},\mathbf{V}}$ jest relacją porządku częściowego.

Definicja 7.7. Niech $neg_i vq_i$ i $neg_j vq_j$ oznaczają dwa zanegowane bądź niezanegowane zbiory kwalifikowane zbudowane w oparciu o dziedzinę $\mathbf{V}$, w postaci, odpowiednio $neg_i\{v_{i1}, \ldots, v_{ili}\}qlr_i$ oraz $neg_j\{v_{j1}, \ldots, v_{jlj}\}qlr_j$, gdzie $neg_i, neg_j \in \{\varepsilon, \neg\}$, $v_{i1}, \ldots, v_{ili}, v_{j1}, \ldots, v_{jlj} \in \mathbf{V}$, qlr_i, $qlr_j \in \{\oplus, \odot\}$. Sumą $neg_i vq_i \cup_{\mathrm{q},\mathbf{V}} neg_j vq_j$ i iloczynem $neg_i vq_i \cap_{\mathrm{q},\mathbf{V}} neg_j vq_j$ tych zbiorów nazywa się zanegowane bądź niezanegowane zbiory kwalifikowane $neg_k \mathbf{V_k} qlr_k$ i $neg_l \mathbf{V_l} qlr_l$ w postaci, odpowiednio:

$neg_i vq_i \cup_{\mathrm{q},\mathbf{V}} neg_j vq_j = neg_k \mathbf{V_k} qlr_k$, gdzie

– $\mathbf{V_k} = \{v_{i1}, \ldots, v_{ili}\} \cap \{v_{j1}, \ldots, v_{jlj}\}$, $qlr_k = \oplus$, $neg_k = \varepsilon$,
jeśli $qlr_i = qlr_j = \oplus$ i $neg_i = neg_j = \varepsilon$,
– $\mathbf{V_k} = \{v_{i1}, \ldots, v_{ili}\} \cup \{v_{j1}, \ldots, v_{jlj}\}$, $qlr_k = \odot$, $neg_k = \varepsilon$,
jeśli $qlr_i = qlr_j = \odot$ i $neg_i = neg_j = \varepsilon$,
– $\mathbf{V_k} = \{v_{i1}, \ldots, v_{ili}\} \cup \{v_{j1}, \ldots, v_{jlj}\}$, $qlr_k = \oplus$, $neg_k = \neg$,
jeśli $qlr_i = qlr_j = \oplus$ i $neg_i = neg_j = \neg$,
– $\mathbf{V_k} = \{v_{i1}, \ldots, v_{ili}\} \cap \{v_{j1}, \ldots, v_{jlj}\}$, $qlr_k = \odot$, $neg_k = \neg$,
jeśli $qlr_i = qlr_j = \odot$ i $neg_i = neg_j = \neg$,
– $\mathbf{V_k} = \{\,\}$, $qlr_k = \oplus$, $neg_k = \varepsilon$,
jeśli $qlr_i \neq qlr_j$ lub $neg_i \neq neg_j$;

$neg_i vq_i \cap_{\mathrm{q,V}} neg_j vq_j = neg_l \mathbf{V}_l qlr_l$, gdzie

- $\mathbf{V}_l = \{v_{i1}, \dots, v_{ili}\} \cup \{v_{j1}, \dots, v_{jlj}\}$, $qlr_l = \oplus$, $neg_l = \varepsilon$,
  jeśli $qlr_i = qlr_j = \oplus$ i $neg_i = neg_j = \varepsilon$,
- $\mathbf{V}_l = \{v_{i1}, \dots, v_{ili}\} \cap \{v_{j1}, \dots, v_{jlj}\}$, $qlr_l = \odot$, $neg_l = \varepsilon$,
  jeśli $qlr_i = qlr_j = \odot$ i $neg_i = neg_j = \varepsilon$,
- $\mathbf{V}_l = \{v_{i1}, \dots, v_{ili}\} \cap \{v_{j1}, \dots, v_{jlj}\}$, $qlr_l = \oplus$, $neg_l = \neg$,
  jeśli $qlr_i = qlr_j = \oplus$ i $neg_i = neg_j = \neg$,
- $\mathbf{V}_l = \{v_{i1}, \dots, v_{ili}\} \cup \{v_{j1}, \dots, v_{jlj}\}$, $qlr_l = \odot$, $neg_l = \neg$,
  jeśli $qlr_i = qlr_j = \odot$ i $neg_i = neg_j = \neg$,
- $\mathbf{V}_l = \{\}$, $qlr_l = \odot$, $neg_l = \varepsilon$,
  jeśli $qlr_i \neq qlr_j$ lub $neg_i \neq neg_j$,

gdzie $\cup$ i $\cap$ oznaczają klasyczne operatory teoriomnogościowe.

Definicja 7.8. Niech $\mathbf{Q_V}$ oznacza zbiór wszystkich kwalifikowanych zbiorów z licznością dających się zbudować w oparciu o dziedzinę **V**, a $neg_i vq_i|d_i$ i $neg_j vq_j|d_j$ oznaczają dwa dowolne zbiory kwalifikowane z licznością pochodzące z $\mathbf{Q_V}$. Na zbiorze $\mathbf{Q_V}$ definiuje się operacje sumy $\cup_{\mathrm{qd},\mathbf{V}}$ i iloczynu $\cap_{\mathrm{qd},\mathbf{V}}$, stanowiące rozszerzenia określonych wyżej (def. 7.7) operacji $\cup_{\mathrm{q},\mathbf{V}}$ i $\cap_{\mathrm{q},\mathbf{V}}$:

- $neg_i vq_i|d_i \cup_{\mathrm{qd,V}} neg_j vq_j|d_j = neg_k vq_k|d_k$, gdzie
  $neg_k vq_k = neg_i vq_i \cup_{\mathrm{q,V}} neg_j vq_j$, $d_k = d_i + d_j$;
- $neg_i vq_i|d_i \cap_{\mathrm{qd,V}} neg_j vq_j|d_j = neg_l vq_l|d_l$, gdzie
  $neg_l vq_l = neg_i vq_i \cap_{\mathrm{q,V}} neg_j vq_j$, $d_k = d_i + d_j$.

Z definicji 7.7 i 7.8 oraz z faktu nierozróżnialności (tożsamości) dwóch zdefiniowanych w dziedzinie **V** zbiorów z licznością $neg\, vq|d_i$ oraz $neg\, vq|d_j$ wynika następujący lemat 7.2.

Lemat 7.2. Operacje sumy $\cup_{\mathrm{qd},\mathbf{V}}$ i iloczynu $\cap_{\mathrm{qd},\mathbf{V}}$ spełniają następujące zależności:

$$\forall\big(neg_i vq_i\big|d_i,\ neg_j vq_j\big|d_j \in \mathbf{Q_V}\big)$$
$$\Big(neg_i vq_i\big|d_i \cup_{\mathrm{qd},\mathbf{V}} neg_j vq_j\big|d_j = \sup\big(neg_i vq_i|d_i,\ neg_j vq_j|d_j\big)\Big)$$
$$\forall\big(neg_i vq_i\big|d_i,\ neg_j vq_j\big|d_j \in \mathbf{Q_V}\big)$$
$$\Big(neg_i vq_i\big|d_i \cap_{\mathrm{qd},\mathbf{V}} neg_j vq_j\big|d_j = \inf\big(neg_i vq_i|d_i,\ neg_j vq_j|d_j\big)\Big).$$

Prostą konsekwencją lematów 7.1 i 7.2 jest twierdzenie 7.1.

Twierdzenie 7.1. Algebra $\mathcal{A}_\mathbf{V} = \big(\mathbf{Q_V},\ \subseteq_{\mathrm{qd},\mathbf{V}},\ \cup_{\mathrm{qd},\mathbf{V}},\ \cap_{\mathrm{qd},\mathbf{V}}\big)$ jest kratą.

Relację porządku częściowego $\subseteq_{\mathrm{qd},\mathbf{V}}$, operację sumy $\cup_{\mathrm{qd},\mathbf{V}}$ i operację iloczynu $\cap_{\mathrm{qd},\mathbf{V}}$ określone na zbiorze kwalifikowanych zbiorów z licznością $\mathbf{Q}_{\mathbf{V}}$ można rozszerzyć na zbiór danych atrybutowych zbiorczych zbudowanych według schematu $(\mathbf{S}_{\mathbf{S1}}, \mathbf{S}_{\mathbf{S2}})$.

Definicja 7.9. Niech $\mathbf{E}_{\mathbf{S1,S2,V}}$ oznacza zbiór danych atrybutowych zbiorczych dających się zbudować według schematu $(\mathbf{S}_{\mathbf{S1}}, \mathbf{S}_{\mathbf{S2}})$ w oparciu o dziedzinę $\mathbf{V}$, a e_f oraz e_g oznaczają dwie dane z tego zbioru, w postaci:

$$\begin{aligned} e_f = <\ & \mathrm{A}_{si1}: neg_{fi1} vq_{fi1} \big| d_{fi}; \ldots; \mathrm{A}_{sim1}: neg_{fim1} vq_{fim1} \big| d_{fi}; \\ & \mathrm{A}_{sj1}: neg_{fj1} vq_{fj1} | d_{fj1} \ldots; \mathrm{A}_{sjm2}: neg_{fjm2} vq_{fjm2} | d_{fjm2} > \end{aligned} \tag{7.6}$$

$$\begin{aligned} e_g = <\ & \mathrm{A}_{si1}: neg_{gi1} vq_{gi1} \big| d_{gi}; \ldots; \mathrm{A}_{sim1}: neg_{gim1} vq_{gim1} \big| d_{gi}; \\ & \mathrm{A}_{sj1}: neg_{gj1} vq_{gj1} | d_{gj1} \ldots; \mathrm{A}_{sjm2}: neg_{gjm2} vq_{gjm2} | d_{gjm2} > \end{aligned} \tag{7.7}$$

Dana atrybutowa zbiorcza e_f subsumuje daną atrybutową zbiorczą e_g, $e_f \subseteq_{\mathbf{S1,S2,V}} e_g$, wtedy i tylko wtedy, gdy spełnione są warunki:

– $\forall(1 \le k \le m_1) \left(neg_{fik} vq_{fik} | d_{fi} \subseteq_{\mathrm{qd},\mathbf{V}} neg_{gik} vq_{gik} | d_{gi}\right)$,

– $\forall(1 \le l \le m_2) \left(neg_{gjl} vq_{gjl} | d_{gjl} \subseteq_{\mathrm{qd},\mathbf{V}} neg_{fjl} vq_{fjl} | d_{fjl}\right)$.

Warunki te oznaczają, że relacja subsumpcji zachodzi między parą takich danych atrybutowych zbiorczych, które są:

– podobne w sensie relacji $\subseteq_{\mathrm{qd},\mathbf{V}}$ na wszystkich atrybutach kluczowych,
– podobne w sensie relacji $\supseteq_{\mathrm{qd},\mathbf{V}}$ na wszystkich atrybutach niekluczowych.

Z definicji 7.4 i 7.9 oraz z lematu 7.1 wynika natychmiast lemat 7.3.

Lemat 7.3. Relacja $\subseteq_{\mathbf{S1,S2,V}}$ jest relacją porządku częściowego.

Definicja 7.10. Niech $\mathbf{E}_{\mathbf{S1,S2,V}}$ oznacza zbiór danych atrybutowych zbiorczych dających się zbudować według schematu $(\mathbf{S}_{\mathbf{S1}}, \mathbf{S}_{\mathbf{S2}})$ w oparciu o dziedzinę $\mathbf{V}$, a e_f oraz e_g oznaczają dwie dane z tego zbioru, w postaci 7.6 i 7.7. Sumą $\cup_{\mathbf{S1,S2,V}}$ i iloczynem $\cap_{\mathbf{S1,S2,V}}$ danych atrybutowych zbiorczych nazywa się operacje, które są zdefiniowane jak następuje:

$$\begin{aligned} e_f \cup_{\mathbf{S1,S2,V}} e_g = <\ & \mathrm{A}_{si1}: neg_{pi1} vq_{pi1} \big| d_{pi}; \ldots; \mathrm{A}_{sim1}: neg_{pim1} vq_{pim1} \big| d_{pi}; \\ & \mathrm{A}_{sj1}: neg_{pj1} vq_{pj1} | d_{pj1} \ldots; \mathrm{A}_{sjm2}: neg_{pjm2} vq_{pjm2} | d_{pjm2} > \end{aligned}$$

$$\begin{aligned} e_f \cap_{\mathbf{S1,S2,V}} e_g = <\ & \mathrm{A}_{si1}: neg_{ri1} vq_{ri1} | d_{ri}; \ldots; \mathrm{A}_{sim1}: neg_{rim1} vq_{rim1} | d_{ri}; \\ & \mathrm{A}_{sj1}: neg_{rj1} vq_{rj1} | d_{rj1} \ldots; \mathrm{A}_{sjm2}: neg_{rjm2} vq_{rjm2} | d_{rjm2} >, \end{aligned}$$

gdzie

$\forall(1 \le k \le m_1)\ neg_{pik} vq_{pik} \big| d_{pi} = neg_{fi1} vq_{fik} \big| d_{fi} \cap_{\mathrm{qd},\mathbf{V}} neg_{gi1} vq_{gik} \big| d_{gi}$,

$\forall(1 \le l \le m_2)\ neg_{pjl} vq_{pjl} \big| d_{pjl} = neg_{fjl} vq_{fjl} \big| d_{fjl} \cup_{\mathrm{qd},\mathbf{V}} neg_{gjl} vq_{gjl} \big| d_{gjl}$,

$\forall(1 \leq k \leq m_1)\ neg_{rik} vq_{pik}|d_{pi} = neg_{fi1} vq_{fik}|d_{fi}\ \cup_{\mathrm{qd},\mathbf{V}}\ neg_{gi1} vq_{gik}|d_{gi}$,
$\forall(1 \leq l \leq m_2)\ neg_{rjl} vq_{pjl}|d_{pjl} = neg_{fjl} vq_{fjl}|d_{fjl}\ \cap_{\mathrm{qd},\mathbf{V}}\ neg_{gjl} vq_{gjl}|d_{gjl}$.

Z definicji 7.10 i lematu 7.3 wynika następujący lemat 7.4.

Lemat 7.4. Niech $\mathbf{E_{S1,S2,V}}$ oznacza zbiór danych atrybutowych zbiorczych dających się zbudować według schematu $(\mathbf{S_{S1}}, \mathbf{S_{S2}})$ w oparciu o dziedzinę **V**. Operacje sumy $\cup_{\mathbf{S1,S2,V}}$ i iloczynu $\cap_{\mathbf{S1,S2,V}}$ spełniają następujące zależności:

$$\forall(e_f, e_g \in \mathbf{E_{S1,S2,V}})\ \left(e_f \cup_{\mathbf{S1,S2,V}} e_g = \sup(e_f, e_g)\right)$$
$$\forall(e_f, e_g \in \mathbf{E_{S1,S2,V}})\ \left(e_f \cap_{\mathbf{S1,S2,V}} e_g = \inf(e_f, e_g)\right).$$

Definicja 7.11. Algebrą danych atrybutowych zbiorczych zbudowanych według schematu $(\mathbf{S_{S1}}, \mathbf{S_{S2}})$ w oparciu o dziedzinę **V** nazywa się algebrę:

$$\mathcal{A}_{\mathrm{S1,S2,V}} = \left(\mathbf{E_{S1,S2,V}}, \subseteq_{\mathbf{S1,S2,V}}, \cup_{\mathbf{S1,S2,V}}, \cap_{\mathbf{S1,S2,V}}\right) \tag{7.8}$$

Konsekwencją lematów 7.3 i 7.4 jest twierdzenie 7.2.

Twierdzenie 7.2. Algebra $\mathcal{A}_{\mathrm{S1,S2,V}} = \left(\mathbf{E_{S1,S2,V}}, \subseteq_{\mathbf{S1,S2,V}}, \cup_{\mathbf{S1,S2,V}}, \cap_{\mathbf{S1,S2,V}}\right)$ jest kratą.

7.3 Algorytmy indukcji i selekcji reguł

Indukcyjne uczenie reguł 2U odbywa się cyklicznie, w dwóch etapach. Pierwszy z nich polega na łączeniu danych atrybutowych zbiorczych według pewnego klucza. W wyniku powyższego łączenia, zwanego dalej integracją danych atrybutowych zbiorczych, powstaje dana o charakterze wirtualnym. W drugim etapie, na podstawie tej danej, odbywa się właściwy proces generacji reguły 2U. Proces ten obejmuje konstrukcję przesłanki i konkluzji reguły oraz obliczanie jej współczynników wiarygodności. Zaproponowany algorytm indukcji reguł 2U zasadza się na wykorzystaniu podobnych pomysłów, co algorytm Apriori [2].

Integracja danych atrybutowych zbiorczych. Idea integrowania danych ma na celu konstruowanie takich (wirtualnych) danych zbiorczych, które będą „mieścić w sobie" największą możliwą liczbę danych atrybutowych indywidualnych. W oczywisty sposób, powyższe dane indywidualne muszą być do siebie podobne w sensie formalnym. Przy odpowiednim poziomie „licznościowym", dana zbiorcza może być użyta do bezpośredniej generacji reguły z niepewnością.

Niech $\mathbf{E_{S1,S2,V}}$ oznacza zbiór danych atrybutowych zbiorczych dających się zbudować według schematu $(\mathbf{S_{S1}}, \mathbf{S_{S2}})$ w oparciu o dziedzinę **V**, a e – dowolną daną z tego zbioru. W dalszym ciągu zakłada się, że e jest w postaci:

$$< \mathrm{A}_{si1}: neg_{i1} vq_{i1} | d_i; \ldots; \mathrm{A}_{sim1}: neg_{im1} vq_{im1} | d_i;$$
$$\mathrm{A}_{sj1}: neg_{j1} vq_{j1} \big| d_{j1} \ldots; \mathrm{A}_{sjm2}: neg_{jm2} vq_{jm2} \big| d_{jm2} > \qquad (7.9)$$

gdzie poszczególne elementy zachowują swoje oznaczenia z definicji 7.3.

Ze względu na interpretację danej atrybutowej zbiorczej (str. 120), daną e można by zastąpić, przy zachowaniu semantyki zbioru $\mathbf{E_{S1,S2,V}}$, następującym zbiorem danych atrybutowych zbiorczych z jednym atrybutem niekluczowym:

$$\left\{\begin{array}{l} < \mathrm{A}_{si1}: neg_{i1} vq_{i1} | d_i; \ldots; \mathrm{A}_{sim1}: neg_{im1} vq_{im1} | d_i;\ \mathrm{A}_{sj1}: neg_{j1} vq_{j1} | d_{j1} >, \\ < \mathrm{A}_{si1}: neg_{i1} vq_{i1} | d_i; \ldots; \mathrm{A}_{sim1}: neg_{im1} vq_{im1} | d_i;\ \mathrm{A}_{sj2}: neg_{j2} vq_{j2} | d_{j2} >, \\ \qquad\qquad\qquad \ldots \\ < \mathrm{A}_{si1}: neg_{i1} vq_{i1} | d_i; \ldots; \mathrm{A}_{sim1}: neg_{im1} vq_{im1} | d_i; \\ \qquad\qquad\qquad\qquad \mathrm{A}_{sjm2}: neg_{jm2} vq_{jm2} | d_{jm2} > \end{array}\right\} \qquad (7.10)$$

Zgodnie z definicją danej atrybutowej zbiorczej (def. 7.3), liczność przypisana do zbiorów kwalifikowanych powiązanych w krotce z atrybutami kluczowymi jest identyczna i maksymalna w całej krotce, a (niewiększe od nich) liczności przypisane do zbiorów kwalifikowanych powiązanych z atrybutami niekluczowymi mogą być różne. Dla danej atrybutowej zbiorczej w postaci 7.9 spełniony jest zatem warunek: $\forall d_{jl}(1 \le l \le m_2)\big(d_{jl} \le d_i\big)$. Użyta w warunku relacja arytmetyczna ma charakter nieostry; może się zdarzyć, że dla jednego lub więcej atrybutów niekluczowych, liczność przypisana do powiązanego z tym atrybutem zbioru będzie równa liczności maksymalnej. Niech $\mathbf{A_{Si}} = \{\mathrm{A}_{sjc1}, \mathrm{A}_{sjc2}, \ldots, \mathrm{A}_{sjcz}\} \subseteq \{\mathrm{A}_{sj1}, \mathrm{A}_{sj2}, \ldots, \mathrm{A}_{sjm2}\}$ oznacza niepusty zbiór takich atrybutów, $\mathbf{A_{Si}} \ne \emptyset$. Bez konsekwencji dla poprawności semantycznej danej atrybutowej zbiorczej, każdy z atrybutów ze zbioru $\mathbf{A_{Si}}$ można przesunąć wraz z odpowiadającym mu zbiorem kwalifikowanym na pozycję atrybutu kluczowego danej. Liczba l_c możliwych kombinacji takich przesunięć wynosi:

– $l_c = \big(2^{|\mathbf{A_{Si}}|} - 1\big)$, gdy $\{\mathrm{A}_{sjc1}, \mathrm{A}_{sjc2}, \ldots, \mathrm{A}_{sjcz}\} \subset \{\mathrm{A}_{sj1}, \mathrm{A}_{sj2}, \ldots, \mathrm{A}_{sjm2}\}$,
– $l_c = \big(2^{|\mathbf{A_{Si}}|} - 2\big)$, gdy $\{\mathrm{A}_{sjc1}, \mathrm{A}_{sjc2}, \ldots, \mathrm{A}_{sjcz}\} = \{\mathrm{A}_{sj1}, \mathrm{A}_{sj2}, \ldots, \mathrm{A}_{sjm2}\}$ (zbiór atrybutów niekluczowych nie może być zbiorem pustym).

Za pomocą takich przesunięć można powiększyć zbiór danych atrybutowych zbiorczych z jednym atrybutem niekluczowym dających się uzyskać z danej wyjściowej w postaci 7.9.

Opisany proces „rozkładu” zostanie zrealizowany dla wszystkich dostępnych danych ze zbioru $\mathbf{E_{S1,S2,V}}$. W dalszej kolejności, każda uzyskana w ten sposób dana atrybutowa zbiorcza z jednym atrybutem niekluczowym będzie integrowana ze wszystkimi innymi danymi podobnymi do niej w sensie relacji subsumpcji szczegółowej (patrz def. 7.14).

Definicja 7.12. Schematem danych atrybutowych zbiorczych z jednym atrybutem niekluczowym w dziedzinie $\mathcal{D}_{\mathcal{U}}$, zgodnym z konceptualizacją $\mathcal{T}_{\mathcal{DU}}$, nazywamy dowolną parę uporządkowaną $(\mathbf{S_{S1}},\ \mathrm{A}_{sj})$, zbudowaną z podzbioru $\mathbf{S_{S1}}$ zbioru atrybutów $\mathbf{A_S}$ i atrybutu $\mathrm{A}_{sj} \in \mathbf{A_S}$ spełniającego warunek:

– $\mathbf{S_{S1}} \cap \{\mathrm{A}_{sj}\} = \emptyset$.

Definicja 7.13. Daną atrybutową zbiorczą z jednym atrybutem niekluczowym zbudowaną według schematu $(\mathbf{S_{S1}},\ \mathrm{A}_{sj})$ w oparciu o dziedzinę $\mathbf{V}$, gdzie $\mathbf{S_{S1}} = \{\mathrm{A}_{si1}, \dots, \mathrm{A}_{sim1}\}$, nazywa się zbiór par uporządkowanych w postaci:

$$\begin{Bmatrix} (\mathrm{A}_{si1},\ neg_{i1}vq_{i1}|d_i),\ \dots,\ (\mathrm{A}_{sim1}, neg_{im1}vq_{im1}|d_i), \\ (\mathrm{A}_{sj}, neg_j vq_j|d_j) \end{Bmatrix} \tag{7.11}$$

w którym neg_{ik}, vq_{ik} $(1 \le k \le m_1)$ oraz d_i zachowują oznaczenia z definicji 7.4, vq_j oznacza dowolny kwalifikowany zbiór wartości – przykładów kategorii C_j, neg_j – opcjonalny operator negacji, $neg_j \in \{\varepsilon, \neg\}$), a $d_j \in (\mathbf{N} \cup \{\mathbf{0}\})$ – liczność zbioru vq_j, spełniającą wymóg: $d_j \le d_i$.

W dalszym ciągu, do zapisu danych atrybutowych zbiorczych z jednym atrybutem niekluczowym używa się uproszczonej notacji krotkowej (7.4).

Definicja 7.14. Niech $\mathbf{E}_{\mathbf{S1},\mathrm{A}sj,\mathbf{V}}$ oznacza zbiór danych atrybutowych zbiorczych z jednym atrybutem niekluczowym, dających się zbudować według schematu $(\mathbf{S_{S1}}, \mathrm{A}_{sj})$ (def. 7.13) w oparciu o dziedzinę $\mathbf{V}$, a $e1_f$ oraz $e1_g$ oznaczają dwie dane z tego zbioru, w postaci:

$$\begin{aligned} e1_f = <\ & \mathrm{A}_{si1}: neg_{fi1}vq_{fi1}|d_{fi}; \dots;\ \mathrm{A}_{sim1}: neg_{fim1}vq_{fim1}|d_{fi}; \\ & \mathrm{A}_{sj}: neg_{fj}vq_{fj}|d_{fj} > \end{aligned} \tag{7.12}$$

$$\begin{aligned} e1_g = <\ & \mathrm{A}_{si1}: neg_{gi1}vq_{gi1}|d_{gi}; \dots;\ \mathrm{A}_{sim1}: neg_{gim1}vq_{gim1}|d_{gi}; \\ & \mathrm{A}_{sj}: neg_{gj}vq_{gj}|d_{gj} > \end{aligned} \tag{7.13}$$

Dana atrybutowa zbiorcza z jednym atrybutem niekluczowym $e1_f$ subsumuje szczegółowo daną atrybutową zbiorczą z jednym atrybutem niekluczowym $e1_g$, $e1_f \sqsubseteq_{\mathbf{S1},\mathrm{A}sj,\mathbf{V}} e1_g$, wtedy i tylko wtedy, gdy spełnione są warunki:

– $\forall (1 \le k \le m_1)\ (neg_{fik}vq_{fik}|d_{fi} \sqsubseteq_{\mathrm{qd},\mathbf{V}} neg_{gk}vq_{gik}|d_{gi})$,

– $neg_{fj}vq_{fj}|d_{fj} \sqsubseteq_{\mathrm{qd},\mathbf{V}} neg_{gj}vq_{gj}|d_{gj}$ i
$neg_{gj}vq_{gj}|d_{gj} \sqsubseteq_{\mathrm{qd},\mathbf{V}} neg_{fj}vq_{fj}|d_{fj}$.

Z definicji 7.5 i 7.6 wynika, że drugi z wymienionych warunków jest równoważny żądaniu: $neg_{fj} = neg_{gj}$ i $vq_{fj} = vq_{gj}$. Relacja $\sqsubseteq_{\mathbf{S1},\mathrm{A}sj,\mathbf{V}}$ jest w oczywisty sposób podzbiorem relacji $\sqsubseteq_{\mathbf{S1},\{\mathrm{A}sj\},\mathbf{V}}$.

Z definicji 7.14 i 7.13 oraz z lematu 7.1 wynika natychmiast lemat 7.5.

Lemat 7.5. Relacja $\sqsubseteq_{\mathbf{S1},\mathrm{A}sj,\mathbf{V}}$ jest relacją porządku częściowego.

Definicja 7.15. Niech $(\mathbf{S_{S1}}, \mathbf{S_{S2}})$ oznacza schemat danych atrybutowych zbiorczych, a $(\mathbf{S_{S1\_k}}, \mathrm{A}_{sj})$, gdzie $\mathbf{S1} \subseteq \mathbf{S1\_k} \subset (\mathbf{S1} \cup \mathbf{S2})$ i $\mathrm{A}_{sj} \in (\mathbf{S_{S2}} \backslash \mathbf{S_{1\_\mathit{k}}})$ – pewien pochodny względem $(\mathbf{S_{S1}}, \mathbf{S_{S2}})$ schemat danych atrybutowych zbiorczych z jednym atrybutem niekluczowym.

Niech $\mathbf{E}_{\mathbf{S1\_k}, \mathrm{A}sj, \mathbf{V}}$ oznacza zbiór wszystkich danych z jednym atrybutem niekluczowym dających się zbudować według schematu $(\mathbf{S_{S1\_k}}, \mathrm{A}_{sj})$ w oparciu o dziedzinę $\mathbf{V}$, $e1_f$ – pewną daną ze zbioru $\mathbf{E}_{\mathbf{S1\_k}, \mathrm{A}sj, \mathbf{V}}$, a $\{e1_{g1}, e1_{g2}, \ldots, e1_{gh}\} \subseteq \mathbf{E}_{\mathbf{S1\_k}, \mathrm{A}sj, \mathbf{V}}$ – podzbiór tych wszystkich elementów ze zbioru $\mathbf{E}_{\mathbf{S1\_k}, \mathrm{A}sj, \mathbf{V}}$, które są integrowalne z $e1_f$, tzn. spełniają warunek: $e1_f \subseteq_{\mathbf{S1\_k}, \mathrm{A}sj, \mathbf{V}} e1_{gi}$, gdzie $1 \leq i \leq h$.

Daną wirtualną $e1_{f\_vir}(e1_f, \mathbf{E}_{\mathbf{S1\_k}, \mathrm{A}sj, \mathbf{V}})$ oblicza się jako sumę:

$$e1_{f_{vir}}(e1_f, \mathbf{E}_{\mathbf{S1_k}, \mathrm{A}sj, \mathbf{V}}) =$$

$$= \left(\left(\left(e1_f \ \cup_{\mathbf{S1}, \{\mathrm{A}sj\}, \mathbf{V}} \ e1_{g1} \right) \cup_{\mathbf{S1}, \{\mathrm{A}sj\}, \mathbf{V}} \right) \ \ldots \cup_{\mathbf{S1}, \{\mathrm{A}sj\}, \mathbf{V}} \ e1_{gh} \right)$$

Generacja reguły 2U. Uzyskana w wyniku integracji dana wirtualna $e1_{f\_vir}(e1_f, \mathbf{E}_{\mathbf{S1\_k}, \mathrm{A}sj, \mathbf{V}})$ zostanie przekształcona do postaci reguły 2U opartej na schemacie $\mathbf{S_S} = \mathbf{S_{S1\_k}} \cup \mathbf{S}_{\{\mathrm{A}_{sj}\}}$ i zwanej dalej w skrócie $R_{e1f\_\mathrm{A}_{sj}}$. Powyższe przekształcenie jest realizowane według następującego przepisu:

$$e1_{f\_vir}(e1_f, \mathbf{E}_{\mathbf{S1\_k}, \mathrm{A}sj, \mathbf{V}}) =$$
$$= < \mathrm{A}_{si1}: neg_{f_v i1} vq_{f_v i1} | d_{f_v i}; \ldots ; \mathrm{A}_{simk}: neg_{f_v imk} vq_{f_v imk} | d_{f_v i};$$
$$\mathrm{A}_{sj}: neg_{f_v j} vq_{f_v j} | d_{f_v j} >$$

$\Rightarrow$

$$\mathrm{grf}(p_r) : \mathrm{A}_{si1} \ \blacksquare_{f_v i1} \ vq_{f_v i1}, \ldots, \mathrm{A}_{simk} \ \blacksquare_{f_v imk} \ vq_{f_v imk} \rightarrow \mathrm{A}_{sj} \ \blacksquare_{f_v j} \ vq_{f_v j} : \mathrm{irf}(p_c)$$

gdzie $\blacksquare_{f_v il}$ $(1 \leq l \leq m_k)/\blacksquare_{f_v j}$ oznacza operator $=$, gdy $neg_{f_v il}/neg_{f_v j}$ jest symbolem pustym (ε) oraz operator $\neq$, gdy $neg_{f_v il}/neg_{f_v j}$ jest symbolem negacji (¬), a p_c i p_r oznaczają wartości współczynników wiarygodności, odpowiednio, wewnętrznego irf i zewnętrznego grf.

Algorytm obliczania wartości współczynników wiarygodności. Obliczanie współczynnika irf będzie się odbywać na etapie transformacji danej wirtualnej, na podstawie następującego wzoru:

$$p_c = \frac{d_{f_v j}}{d_{f_v i}} \tag{7.14}$$

Znacznie trudniej jest zaproponować poprawną metodę wyznaczania współczynnika grf. Brak uznanych metod algorytmicznych zachęca do rozwiązania tego problemu przy użyciu technik heurystycznych. Jedną z takich propozycji przedłożono w pracy [80]. Rozwiązanie opiera się na ocenie dwóch właściwości reguły 2U: wagi i precyzji. Waga reguły zależy od mocy wirtualnej danej zbiorczej, mierzonej liczbą danych indywidualnych reprezentowanych tą daną zbiorczą, i od wyrazistości reguły, mierzonej odległością współczynnika p_c od punktów krańcowych zakresu $\langle 0;1\rangle$. Można ją więc wyznaczyć na etapie transformacji danej wirtualnej, podobnie jak sam współczynnik p_c. Z kolei, precyzja reguły zależy od stopnia wzajemnego podobieństwa danych zbiorczych, z których zintegrowano wirtualną daną zbiorczą: im mniejsze zróżnicowanie tych danych, tym większa precyzja reguły. Ten stopień należy oszacować w procesie integracji danych – po jego zakończeniu dane potrzebne do obliczeń stają się niedostępne.

W rozwiązaniu zaproponowanym w [80], wpływ obu wymienionych właściwości na współczynnik grf przyjęto wyrażać następującym wzorem:

$$p_r = \min\left(\mathrm{wg}\left(R_{e1f\_\mathrm{A}_{sj}}\right), \mathrm{acc}\left(R_{e1f\_\mathrm{A}_{sj}}\right)\right) \tag{7.15}$$

gdzie wg i acc są nazwami funkcji wyznaczających, odpowiednio wagę i precyzję reguły, i przyjmujących wartości z przedziału $(0;1\rangle$. Po uzyskaniu doświadczeń związanych z użyciem reguł 2U do wnioskowania w pewnej konkretnej dziedzinie, wzór ten będzie można zmodyfikować lub zastąpić innym, gwarantującym lepszą efektywność i wydajność systemu $\boldsymbol{R\_2U}$ dla tej dziedziny.

Selekcja reguł. Celem konstrukcji reguły 2U jest jej wykorzystanie w systemie regułowym do prowadzenia wnioskowań. Nowo wygenerowaną regułę umieszcza się w bazie wiedzy tego systemu. Zanim to jednak nastąpi, należy:
- wykluczyć zachodzenie warunków poddających w wątpliwość dodatni monotoniczny charakter zależności reprezentowanej regułą (twierdz. 6.2),
- wykluczyć zachodzenie sprzeczności pomiędzy badaną regułą i dowolną z reguł znajdujących się w bazie wiedzy (def. 6.7),
- wykluczyć istnienie w bazie wiedzy reguły subsumującej badaną regułę (def. 6.8).

Wszystkie wymienione zadania mają charakter globalny. Z tego powodu zostaną wykonane dopiero po wygenerowaniu wszystkich reguł. Najtrudniejsze z zadań, czyli badanie reguły pod kątem dodatniej monotoniczności, będzie prowadzone dwuetapowo. Pierwszy etap będzie polegał na wyselekcjonowaniu takiego, możliwie największego, podzbioru zbioru wszystkich reguł, który gwarantuje niezachodzenie zależności 6.11 i 6.12. W ramach tego etapu, zostaną wyodrębnione wszystkie takie pary zbiorów kwalifikowanych (vq_i, vq_j), które spełniają następujące warunki:
- każdy ze zbiorów kwalifikowanych vq_i i vq_j występuje w przynajmniej jednej regule, po stronie jej przesłanki lub konkluzji,
- żadna z reguł nie zawiera (ani po stronie przesłanki, ani konkluzji) zbioru kwalifikowanego vq_k takiego, że $vq_k \subseteq_{\mathrm{q},\mathbf{V}} vq_i$ oraz $vq_k \neq vq_i$,

– żadna z reguł nie zawiera (ani stronie przesłanki, ani konkluzji) zbioru kwalifikowanego vq_k takiego, że $vq_k \subseteq_{\mathrm{q,V}} vq_j$ oraz $vq_k \neq vq_j$.

Następnie dla każdej pary zostaną zliczone reguły, w których po stronie, odpowiednio przesłanki i konkluzji reguły lub odwrotnie:

(a) występują w postaci niezanegowanej zbiory kwalifikowane, odpowiednio vq_i' i vq_j' takie, że $vq_i \subseteq_{\mathrm{q,V}} vq_i'$ oraz $vq_j \subseteq_{\mathrm{q,V}} vq_j'$ (liczba o symbolu $n\_vq_{i+}\_vq_{j+}$),

(b) występują w postaci zanegowanej zbiory kwalifikowane, odpowiednio $\neg vq_i'$ i $\neg vq_j'$ takie, że $vq_i \subseteq_{\mathrm{q,V}} vq_i'$ oraz $vq_j \subseteq_{\mathrm{q,V}} vq_j'$ (liczba o symbolu $n\_vq_{i-}\_vq_{j-}$),

(c) występują, odpowiednio: w postaci niezanegowanej zbiór kwalifikowany vq_i' i w postaci zanegowanej zbiór kwalifikowany $\neg vq_j'$ takie, że $vq_i \subseteq_{\mathrm{q,V}} vq_i'$ oraz $vq_j \subseteq_{\mathrm{q,V}} vq_j'$ (liczba o symbolu $n\_vq_{i+}\_vq_{j-}$),

(d) występują, odpowiednio: w postaci zanegowanej zbiór kwalifikowany $\neg vq_i'$ i w postaci niezanegowanej zbiór kwalifikowany vq_j' takie, że $vq_i \subseteq_{\mathrm{q,V}} vq_i'$ oraz $vq_j \subseteq_{\mathrm{q,V}} vq_j'$ (liczba o symbolu $n\_vq_{i-}\_vq_{j+}$).

W kolejności, nastąpi sprawdzenie relacji pomiędzy liczbami $\left(n\_vq_{i+}\_vq_{j+} + n\_vq_{i-}\_vq_{j-}\right)$ i $\left(n\_vq_{i+}\_vq_{j-} + n\_vq_{i-}\_vq_{j+}\right)$. Możliwe są następujące przypadki:

– $\left(n\_vq_{i+}\_vq_{j+} + n\_vq_{i-}\_vq_{j-}\right) \gg \left(n\_vq_{i+}\_vq_{j-} + n\_vq_{i-}\_vq_{j+}\right)$ – za poprawny należy uznać w regule układ elementów (a) i (b), za niepoprawny – układ elementów (c) i (d),

– $\left(n\_vq_{i+}\_vq_{j+} + n\_vq_{i-}\_vq_{j-}\right) \ll \left(n\_vq_{i+}\_vq_{j-} + n\_vq_{i-}\_vq_{j+}\right)$ – za poprawny należy uznać w regule układ elementów (c) i (d), za niepoprawny – układ elementów (a) i (b),

– $\left(n\_vq_{i+}\_vq_{j+} + n\_vq_{i-}\_vq_{j-}\right) \approx \left(n\_vq_{i+}\_vq_{j-} + n\_vq_{i-}\_vq_{j+}\right)$ – poprawność poszczególnych układów elementów jest trudna do zweryfikowania.

Niech $R_{e1f\_\mathrm{A}_{sj}}$ oznacza wygenerowaną regułę 2U w postaci:

$$\mathrm{grf}(p_r) : \mathrm{A}_{si1} \blacksquare_{f_v i1}\, vq_{f_v i1}, \ldots, \mathrm{A}_{simk} \blacksquare_{f_v imk}\, vq_{f_v imk} \rightarrow \mathrm{A}_{sj} \blacksquare_{f_v j}\, vq_{f_v j} : \mathrm{irf}(p_c)$$

Jeśli każdy z układów $\left(\blacksquare_{f_v il} vq_{f_v il}, \blacksquare_{f_v j}\, vq_{f_v j}\right)$ $(1 \leq l \leq m_k)$ jest poprawny, to regułę $R_{e1f\_\mathrm{A}_{sj}}$ należy dołączyć do podzbioru w niezmienionej postaci. Jeśli każdy z wymienionych układów jest niepoprawny, to regułę należy dołączyć do podzbioru w zmodyfikowanej postaci $R_{e1f\_\mathrm{A}_{sj}}{}'$:

$$\mathrm{grf}(p_r) : \mathrm{A}_{si1} \blacksquare_{f_v i1}\, vq_{f_v i1}, \ldots, \mathrm{A}_{simk} \blacksquare_{f_v imk}\, vq_{f_v imk} \rightarrow \mathrm{A}_{sj} \blacksquare_{f_v j}{}'\, vq_{f_v j} : \mathrm{irf}(p_c{}')$$

gdzie $\blacksquare_{f_v j}{}'$ oznacza operator odwrotny do $\blacksquare_{f_v j}$, a $p_c{}' = 1 - p_c$.

Jeśli są pomiędzy rozważanymi układami takie, których poprawności nie da się określić, to regułę $R_{e1f\_A_{sj}}$ należy pominąć przy konstrukcji podzbioru.

Drugi etap badania reguł pod kątem dodatniej monotoniczności będzie polegał na wyselekcjonowaniu z uzyskanego zbioru tych spośród reguł, które gwarantują niezachodzenie zależności 6.13 i 6.14. W tym celu, każda para reguł spełniających którąkolwiek z tych zależności zostanie oznaczona jako „podejrzana". Oznaczenia będę nanoszone wielokrotnie, przy każdym stwierdzeniu zależności. Następnie reguła/reguły z największą liczbą oznaczeń zostaną usunięte ze zbioru. Proces oznaczania i usuwania kolejnych reguł będzie powtarzany cyklicznie do momentu, dopóki liczba oznaczeń nie spadnie do zera. W uzyskanym w ten sposób zbiorze żadna para reguł nie będzie pozostawać w sprzeczności (patrz str. 92).

W ostatnim etapie finalny zbiór reguł będzie testowany pod kątem występowania reguł nadmiarowych, czyli subsumowanych przez inne reguły ze zbioru (def. 6.8). Wszystkie wykryte reguły nadmiarowe zostaną z niego bezwarunkowo usunięte.

Zaprezentowany algorytm selekcji reguł do bazy wiedzy systemu $\boldsymbol{R\_2U}$ daje gwarancję niesprzeczności zaprojektowanej bazy wiedzy i niewystępowania reguł nadmiarowych w tej bazie. Ponadto, algorytm zapewnia, że zależności reprezentowane przez wyselekcjonowane reguły będą – z wysokim prawdopodobieństwem – dodatnie monotoniczne.

Opisana rygorystyczna selekcja reguł będzie prowadzić do istotnego zmniejszenia liczności wyjściowego zbioru reguł. W wypadku ewidencji mocno skonfliktowanej wewnętrznie, może to doprowadzić do uzyskania małej, a przez to słabej bazy wiedzy. Alternatywnie, można podjąć próbę zastosowania innych, heurystycznych metod selekcji reguł. W procesie selekcji można się kierować, między innymi, kryterium zewnętrznego współczynnika wiarygodności grf.

7.4 Implementacja procesu projektowania bazy wiedzy

Jak wynika z wcześniejszych rozważań, proces projektowania bazy wiedzy systemu $\boldsymbol{R\_2U}$ będzie przebiegał w dwóch krokach. Pierwszy z nich polega na cyklicznym wykonywaniu integracji danych atrybutowych zbiorczych i generacji reguł 2U. Każdy cykl tego procesu jest sterowany postacią jednej z dostępnych danych atrybutowych zbiorczych z jednym atrybutem niekluczowym. Drugi krok procesu to „przycinanie" zbioru wygenerowanych reguł 2U. Ma ono na celu eliminację ze zbioru tych reguł, które naruszają pewne istotne wymagania jakościowe stawiane bazie wiedzy systemu $\boldsymbol{R\_2U}$.

Powyższy proces projektowania bazy wiedzy zaimplementowano w języku C#, przy wykorzystaniu kolekcji standardowych. Kod 7.1 prezentuje kluczowy fragment tej implementacji, wzorowany na programie zamieszczonym w [78].

```
s.Clear(KB);
foreach (d_i in D)  //D - wektor danych atrybutowych zbiorczych
  { s.Remove(D, d_i);
    K_i = d.key(d_i); U_i = d.nonkey(d_i); C_i = d.common(d_i);
    foreach (E_ij in s.PowerSet(U_i))
      if (s.IsSubset(E_ij, C_i))
         { s.Subtract(U_i, E_ij); s.Sum(K_i, E_ij);
           foreach (f_ijk in U_i)
             { d_vir = d_i;
               foreach (d_h in D)
                 { K_h = d.key(d_h); U_h = d.nonkey(d_h); C_h = d.common(d_h);
                   flag1 = true;
                   foreach (A_ig in K_i)
                     if (!(s.Contains(C_h, A_ig) &&
                           d.preserves(d_i, d_h, A_ig)))
                        { flag1 = false; break; }
                   if (!flag1)
                      continue;
                   if (!(s.Contains(U_h, f_ijk) &&
                         s.Contains(C_h, f_ijk) &&
                         d.preserves(d_i, d_h, f_ijk)) &&
                       !(s.Contains(U_h, f_ijk) &&
                         !s.Contains(C_h, f_ijk) &&
                         d.preserves(d_i, d_h, f_ijk) &&
                         d.preserves(d_h, d_i, f_ijk)))
                      continue;
                   d.integrate(d_vir, d_h);
                 }
               R_ijk = d.generate(d_vir, d_i, K_i, f_ijk);
               s.Add(KB, R_ijk);
             }
           s.Sum(U_i, E_ij);
         }
    s.Add(D, d_i);
  }
kb.Prune1(KB);
kb.Prune2(KB);
foreach (R_s in KB)
  { s.Remove(KB, R_s); flag2 = true;
    foreach (R_t in KB)
      if (r.subsumes(R_t, R_s))
         { flag2 = false; break; }
    if (flag2)
       Add(KB, R_s);
  }
```

Kod 7.1. Implementacja procesu projektowania bazy wiedzy systemu ***R_2U***

Użyte w kodzie funkcje: `Clear`, `Add`, `Remove`, `Contains`, `IsSubset`, `Sum`, `Subtract` i `PowerSet` mają charakter generyczny i reprezentują operacje na zbiorach, odpowiednio: czyszczenia zbioru; dodawania elementu do zbioru; usuwania elementu ze zbioru; weryfikacji przynależności elementu do zbioru; weryfikacji zawierania się zbiorów; sumowania zbiorów; odejmowania zbiorów; tworzenia zbioru potęgowego dla wskazanego zbioru. Z kolei funkcje: `key`, `non_key`, `common`, `preserves`, `integrate` i `generate` odnoszą się do danych atrybutowych zbiorczych, implementowanych w postaci słowników. Funkcje te służą do, odpowiednio: wyodrębnienia zbioru atrybutów kluczowych; wyodrębnienia zbioru atrybutów niekluczowych; wyodrębnienia atrybutów, do których przypisano zbiory kwalifikowane z maksymalną licznością (w obrębie danej); testowania zachodzenia relacji $\subseteq_{\mathrm{q},\mathbf{V}}$ pomiędzy prostymi/zanegowanymi zbiorami kwalifikowanymi przypisanymi temu samemu atrybutowi w obrębie dwóch różnych danych atrybutowych; złączenia wskazanej danej atrybutowej z daną atrybutową wyjściową; generacji reguły 2U ze wskazanej danej atrybutowej wirtualnej. Przedmiotem działania funkcji `subsumes` są reguły 2U. Funkcja ta służy do weryfikacji faktu subsumowania jednej reguły 2U przez drugą. Złożone operacje `Prune1` i `Prune2` odpowiadają za: wstępne badanie zależności dodatniej monotonicznej reguł ze wskazanej bazy wiedzy systemu $\boldsymbol{R\_2U}$; zaawansowane badanie zależności dodatniej monotonicznej (i niesprzeczności reguł) ze wskazanej bazy wiedzy systemu $\boldsymbol{R\_2U}$.

7.5 Przykłady indukcji i selekcji reguł

Zamieszczone w niniejszym podrozdziale przykłady ilustrują kolejno: rozkład danej atrybutowej zbiorczej na zbiór danych atrybutowych z jednym atrybutem niekluczowym, integrację danych z jednym atrybutem niekluczowym, transformację danej atrybutowej wirtualnej do postaci reguły 2U, projektowanie bazy wiedzy systemu $\boldsymbol{R\_2U}$, poprzez: wstępną selekcję reguł 2U, zaawansowaną selekcję reguł 2U, usuwanie reguł nadmiarowych.

Przykład 7.1. Niech $(\mathbf{S_{S1}}, \mathbf{S_{S2}})$ oznacza schemat danych atrybutowych zbiorczych w postaci $(\{\mathrm{Wydz\_Rodz}, \mathrm{Kier}\}, \{\mathrm{plec}, \mathrm{rok}, \mathrm{sprzet}\})$, a e_1 – daną atrybutową zbiorczą zbudowaną według schematu $(\mathbf{S_{S1}}, \mathbf{S_{S2}})$, w oparciu o dziedzinę o umownej nazwie Studenci_Politechniki_Poznanskiej (w skrócie – SPP), z odpowiadającym jej zbiorem wartości $\mathbf{V_{SPP}}$. Dana e_1 ma postać:

$$
\begin{aligned}
e_1 = <\ & \mathrm{Wydz\_Rodz} : \{\mathrm{WInf\_st}, \mathrm{WInf\_nst}\} \oplus |32; \\
& \mathrm{Kier} : \{\mathrm{Inf}, \mathrm{AiR}\} \oplus |32; \\
& \mathrm{plec} : \{\mathrm{M}\} \odot |32; \\
& \mathrm{rok} : \{3\} \odot |32; \\
& \mathrm{sprzet} : \{\mathrm{K\_st}, \mathrm{Lap}\} \odot |19 >
\end{aligned}
$$

gdzie poszczególne atrybuty i wartości zachowują swoje oznaczenia z rozdziału 6 (jedyna modyfikacja polega na zmianie pierwszych liter w nazwach atrybutów niekluczowych z wielkich na małe).

Powyższą daną można zastąpić następującym, semantycznie równoważnym zbiorem danych atrybutowych zbiorczych z jednym atrybutem niekluczowym:

$$\begin{aligned}
\{ <\ & \text{Wydz_Rodz} : \{\text{WInf_st}, \text{WInf_nst}\} \oplus |32; && (e1_a)\\
& \text{Kier} : \{\text{Inf}, \text{AiR}\} \oplus |32; &&\\
& \text{plec} : \{\text{M}\} \odot |32 >, &&\\
<\ & \text{Wydz_Rodz} : \{\text{WInf_st}, \text{WInf_nst}\} \oplus |32; && (e1_b)\\
& \text{Kier} : \{\text{Inf}, \text{AiR}\} \oplus |32; &&\\
& \text{rok} : \{3\} \odot |32 >, &&\\
<\ & \text{Wydz_Rodz} : \{\text{WInf_st}, \text{WInf_nst}\} \oplus |32; && (e1_c)\\
& \text{Kier} : \{\text{Inf}, \text{AiR}\} \oplus |32; &&\\
& \text{sprzet} : \{\text{K_st}, \text{Lap}\} \odot |19 > &&\\
<\ & \text{Wydz_Rodz} : \{\text{WInf_st}, \text{WInf_nst}\} \oplus |32; && (e1_d)\\
& \text{Kier} : \{\text{Inf}, \text{AiR}\} \oplus |32; &&\\
& \text{Plec} : \{\text{M}\} \odot |32; &&\\
& \text{rok} : \{3\} \odot |32 >, &&\\
<\ & \text{Wydz_Rodz} : \{\text{WInf_st}, \text{WInf_nst}\} \oplus |32; && (e1_e)\\
& \text{Kier} : \{\text{Inf}, \text{AiR}\} \oplus |32; &&\\
& \text{Plec} : \{\text{M}\} \odot |32; &&\\
& \text{sprzet} : \{\text{K_st}, \text{Lap}\} \odot |19 >, &&\\
<\ & \text{Wydz_Rodz} : \{\text{WInf_st}, \text{WInf_nst}\} \oplus |32; && (e1_f)\\
& \text{Kier} : \{\text{Inf}, \text{AiR}\} \oplus |32; &&\\
& \text{Rok} : \{3\} \odot |32; &&\\
& \text{plec} : \{\text{M}\} \odot |32 >, &&\\
<\ & \text{Wydz_Rodz} : \{\text{WInf_st}, \text{WInf_nst}\} \oplus |32; && (e1_g)\\
& \text{Kier} : \{\text{Inf}, \text{AiR}\} \oplus |32; &&\\
& \text{Rok} : \{3\} \odot |32; &&\\
& \text{sprzet} : \{\text{K_st}, \text{Lap}\} \odot |19 >, &&\\
<\ & \text{Wydz_Rodz} : \{\text{WInf_st}, \text{WInf_nst}\} \oplus |32; && (e1_h)\\
& \text{Kier} : \{\text{Inf}, \text{AiR}\} \oplus |32; &&\\
& \text{Plec} : \{\text{M}\} \odot |32; &&\\
& \text{Rok} : \{3\} \odot |32; &&\\
& \text{sprzet} : \{\text{K_st}, \text{Lap}\} \odot |19 > \} \;. \;\square &&
\end{aligned}$$

Przykład 7.2. Niech e_1 oznacza daną atrybutową zbiorczą w postaci identycznej jak w przykładzie 7.1, a e_2 i e_3 – dwie nowe dane atrybutowe zbiorcze, zbudowane według tego samego schematu $(\mathbf{S_{S1}}, \mathbf{S_{S2}})$, w oparciu o tę samą dziedzinę wartości $\mathbf{V_{SPP}}$, w postaci:

$$
\begin{aligned}
e_2 = <\ & \text{Wydz\_Rodz} : \{\text{WInf\_st}\} \odot |45; \\
& \text{Kier} : \{\text{Inf}\} \odot |45; \\
& \text{plec} : \{\text{M}\} \odot |41; \\
& \text{rok} : \{2{,}3\} \oplus |39; \\
& \text{sprzet} : \{\text{Lap}\} \odot |43 >
\end{aligned}
$$

$$
\begin{aligned}
e_3 = <\ & \text{Wydz\_Rodz} : \{\text{WInf\_st}, \text{WInf\_nst}\} \oplus |71; \\
& \text{Kier} : \{\text{Inf}, \text{AiR}\} \oplus |71; \\
& \text{plec} : \{\text{M}, \text{K}\} \oplus |71; \\
& \text{rok} : \neg\{3\} \odot |36; \\
& \text{sprzet} : \{\text{K\_st}, \text{Lap}\} \odot |68 >
\end{aligned}
$$

Integracja danych atrybutowych zbiorczych z jednym atrybutem niekluczowym uzyskanych przez rozkład e_1 (dane $e1_a - e1_h$) z danymi atrybutowymi zbiorczymi z jednym atrybutem niekluczowym uzyskanymi przez analogiczny rozkład e_2 i e_3 da w wyniku następujący zbiór danych atrybutowych wirtualnych:

$$
\begin{aligned}
\{ <\ & \text{Wydz\_Rodz} : \{\text{WInf\_st}, \text{WInf\_nst}\} \oplus |77; && (e1_{vir\_a}) \\
& \text{Kier} : \{\text{Inf}, \text{AiR}\} \oplus |77; \\
& \text{plec} : \{\text{M}\} \odot |73 >, \\
<\ & \text{Wydz\_Rodz} : \{\text{WInf\_st}, \text{WInf\_nst}\} \oplus |32; && (e1_{vir\_b}) \\
& \text{Kier} : \{\text{Inf}, \text{AiR}\} \oplus |32 \\
& \text{rok} : \{3\} \odot |32 >, \\
<\ & \text{Wydz\_Rodz} : \{\text{WInf\_st}, \text{WInf\_nst}\} \oplus |103; && (e1_{vir\_c}) \\
& \text{Kier} : \{\text{Inf}, \text{AiR}\} \oplus |103; \\
& \text{sprzet} : \{\text{K\_st}, \text{Lap}\} \odot |87 >, \\
<\ & \text{Wydz\_Rodz} : \{\text{WInf\_st}, \text{WInf\_nst}\} \oplus |32; && (e1_{vir\_d}) \\
& \text{Kier} : \{\text{Inf}, \text{AiR}\} \oplus |32; \\
& \text{Plec} : \{\text{M}\} \odot |32; \\
& \text{rok} : \{3\} \odot |32 >, \\
<\ & \text{Wydz\_Rodz} : \{\text{WInf\_st}, \text{WInf\_nst}\} \oplus |32; && (e1_{vir\_e}) \\
& \text{Kier} : \{\text{Inf}, \text{AiR}\} \oplus |32; \\
& \text{Plec} : \{\text{M}\} \odot |32; \\
& \text{sprzet} : \{\text{K\_st}, \text{Lap}\} \odot |19 >, \\
<\ & \text{Wydz\_Rodz} : \{\text{WInf\_st}, \text{WInf\_nst}\} \oplus |32; && (e1_{vir\_f}) \\
& \text{Kier} : \{\text{Inf}, \text{AiR}\} \oplus |32; \\
& \text{Rok} : \{3\} \odot |32; \\
& \text{plec} : \{\text{M}\} \odot |32 >,
\end{aligned}
$$

$< \text{Wydz\_Rodz} : \{\text{WInf\_st}, \text{WInf\_nst}\} \oplus |32;$ $(e1_{vir\_g})$
$\text{Kier} : \{\text{Inf}, \text{AiR}\} \oplus |32;$
$\text{Rok} : \{3\} \odot |32;$
$\text{sprzet} : \{\text{K\_st}, \text{Lap}\} \odot |19 >,$

$< \text{Wydz\_Rodz} : \{\text{WInf\_st}, \text{WInf\_nst}\} \oplus |32;$ $(e1_{vir\_h})$
$\text{Kier} : \{\text{Inf}, \text{AiR}\} \oplus |32;$
$\text{Plec} : \{\text{M}\} \odot |32;$
$\text{Rok} : \{3\} \odot |32;$
$\text{sprzet} : \{\text{K\_st}, \text{Lap}\} \odot |19 > \}$.

Przykładowo, dana wirtualna $e1_{vir\_a}$ jest wynikiem integracji danej $e1_a$ z następującą daną $e1_i$, uzyskaną przez rozkład danej atrybutowej zbiorczej e_2:

$< \text{Wydz\_Rodz} : \{\text{WInf\_st}\} \odot |45;$ $(e1_i)$
$\text{Kier} : \{\text{Inf}\} \odot |45;$
$\text{plec} : \{\text{M}\} \odot |41 \ >$

Dana $e1_i$ jest zbudowana według tego schematu co $e1_a$ $(\mathbf{S_{S1}}, \text{plec}) = (\{\text{Wydz\_Rodz}, \text{Kier}\}, \text{plec}))$ i jest podobna do niej w sensie relacji $\subseteq_{\mathbf{S1},\text{plec},\mathbf{V_{SPP}}}$ ze względu na spełnienie warunku:

$$(\{\text{WInf\_st}, \text{WInf\_nst}\} \oplus |32 \subseteq_{\text{qd},\mathbf{V_{SPP}}} \{\text{WInf\_st}\} \odot |45) \wedge$$
$$\left(\{\text{Inf}, \text{AiR}\} \oplus |32 \subseteq_{\text{qd},\mathbf{V_{SPP}}} \{\text{Inf}\} \odot |45\right) \wedge$$
$$\left(\{\text{M}\} \odot |32 \subseteq_{\text{qd},\mathbf{V_{SPP}}} \{\text{M}\} \odot |41\right) \wedge \left(\{\text{M}\} \odot |41 \subseteq_{\text{qd},\mathbf{V_{SPP}}} \{\text{M}\} \odot |32\right)$$

Wymagane podobieństwo nie występuje natomiast pomiędzy daną $e1_a$ a następującą daną $e1_l$, uzyskaną przez rozkład danej atrybutowej zbiorczej e_3 i zbudowaną według tego samego schematu $(\mathbf{S_{S1}}, \text{plec})$:

$< \text{Wydz\_Rodz} : \{\text{WInf\_st}, \text{WInf\_nst}\} \odot |71;$ (e_l)
$\text{Kier} : \{\text{Inf}, \text{AiR}\} \odot |71;$
$\text{plec} : \{\text{M}, \text{K}\} \oplus |71 \ >$

Ze względu na niespełnienie warunku:

$$\{\text{M}\} \odot |32 \subseteq_{\text{qd},\mathbf{V_{SPP}}} \{\text{M}, \text{K}\} \oplus |41 ,$$

zachodzi ostatecznie:

$$e1_{vir\_a} = e1_a \cup_{\mathbf{S1},\{\text{plec}\},\mathbf{V_{SPP}}} e1_i =$$
$$=< \text{Wydz\_Rodz} : (\{\text{WInf\_st}, \text{WInf\_nst}\} \oplus) \cap_{\text{qd},\mathbf{V_{SPP}}} (\{\text{WInf\_st}\} \odot)|(32 + 45);$$
$$\text{Kier} : (\{\text{Inf}, \text{AiR}\} \oplus) \cap_{\text{qd},\mathbf{V_{SPP}}} (\{\text{Inf}\} \odot)(32 + 45);$$
$$\text{plec} : (\{\text{M}\} \odot) \cup_{\text{q},\mathbf{V_{SPP}}} (\{\text{M}\} \odot)|(32 + 41) > .$$

Analogiczne procesy integracji dla danych atrybutowych zbiorczych z jednym atrybutem niekluczowym uzyskanych przez rozkład e_2 i e_3 dadzą w wyniku 8 kolejnych danych atrybutowych wirtualnych, zwanych umownie $e1_{vir\_i}$ – $e1_{vir\_p}$. □

Przykład 7.3. Niech $e1_{vir\_a}$ oznacza daną atrybutową wirtualną w postaci identycznej jak w przykładzie 7.2:

$$
\begin{aligned}
< & \text{Wydz\_Rodz} : \{\text{WInf\_st}, \text{WInf\_nst}\} \oplus |77; \\
& \text{Kier} : \{\text{Inf}, \text{AiR}\} \oplus |77; \\
& \text{plec} : \{\text{M}\} \odot |73 > .
\end{aligned}
$$

Przy użyciu algorytmu zdefiniowanego w podrozdziale 7.3, daną tę można przekształcić do postaci następującej reguły 2U, zwanej dalej R_{e1vir_a}:

$$
\begin{aligned}
& \text{grf}(p_r) : \\
& \quad \text{Wydz\_Rodz} = \{\text{WInf\_st}, \text{WInf\_nst}\} \oplus, \\
& \quad \text{Kier} = \{\text{Inf}, \text{AiR}\} \oplus \\
\rightarrow & \quad \text{plec} = \{\text{M}\} \odot : \text{irf}(p_c)
\end{aligned}
$$

gdzie $p_c = \frac{73}{77} \approx 0.95$, a $p_r = \min\left(\text{wg}\left(R_{e1vir\_a}\right), \text{acc}\left(R_{e1vir\_a}\right)\right) =$

$$= \min\left(\min\left(1 - 2 \cdot 1.96 \cdot \sqrt{\frac{0.95 \cdot 0.05}{77}}, 0.95\right), \frac{2 \cdot \frac{1 \cdot 32 + \frac{45}{2}}{77} + 1}{3}\right) \approx \min(0.90, 0.81) =$$

$= 0.81$. W obliczeniach współczynnika p_r wzięto pod uwagę:

- przy oznaczaniu wartości wagi wg – współczynnik $p_c' = \min(p_c, 0.95)$ w miejsce współczynnika p_c ,
- przy oznaczaniu wartości precyzji acc – względną precyzję formuł zawartych w przesłance i konkluzji reguły [80].

W wyniku analogicznego procesu transformacji danych wirtualnych $e1_{vir\_b}$ - $e1_{vir\_p}$, zostaną wygenerowane reguły $R_{e1vir\_b} - R_{e1vir\_p}$. □

Przykład 7.4. W zbiorze wygenerowanych 16 reguł 2U znajdują się, między innymi, reguła R_{e1vir_b}, w postaci:

$$
\begin{aligned}
& \text{grf}(0.85) : \\
& \quad \text{Wydz\_Rodz} = \{\text{WInf\_st}, \text{WInf\_nst}\} \oplus, \\
& \quad \text{Kier} = \{\text{Inf}, \text{AiR}\} \oplus \\
\rightarrow & \quad \text{rok} = \{3\} \odot : \text{irf}(1.0)
\end{aligned}
$$

oraz reguła R_{e1vir_m}, w postaci:

$$
\begin{aligned}
& \text{grf}(0.77) : \\
& \quad \text{Wydz\_Rodz} = \{\text{WInf\_st}, \text{WInf\_nst}\} \oplus, \\
& \quad \text{Kier} = \{\text{Inf}, \text{AiR}\} \oplus \\
\rightarrow & \quad \text{rok} = \neg\{3\} \odot : \text{with irf}(0.51)
\end{aligned}
$$

Wstępne badanie zbioru wszystkich reguł pod kątem dodatniej monotoniczności prowadzi do konkluzji, że:

$$n\_\{\mathrm{WInf\_st}, \mathrm{WInf\_nst}\} \oplus_{+} \_\{3\} \odot_{+} + n\_\{\mathrm{WInf\_st}, \mathrm{WInf\_nst}\} \oplus_{-} \_\{3\} \odot_{-} =$$
$$n\_\{\mathrm{WInf\_st}, \mathrm{WInf\_nst}\} \oplus_{+} \_\{3\} \odot_{-} + n\_\{\mathrm{WInf\_st}, \mathrm{WInf\_nst}\} \oplus_{-} \_\{3\} \odot_{+} =$$
$$= 2$$

W takiej sytuacji, nie jest możliwe określenie charakteru zależności pomiędzy elementami $\{\mathrm{WInf\_st}, \mathrm{WInf\_nst}\} \oplus$ i $\{3\} \odot$, a tym samym – określenie poprawności reguł R_{e1vir_b} i R_{1evir_m}. Z tego powodu, obie wymienione reguły zostaną usunięte ze zbioru (z tego samego powodu, zostaną usunięte także reguły R_{e1vir_d} i R_{e1vir_o}). □

Przykład 7.5. W kolejności, nastąpi proces zaawansowanego badania reguł pod kątem dodatniej monotoniczności oraz niesprzeczności. Badaniu zostanie poddana każda para reguł 2U pochodzących z 12-elementowego zbioru uzyskanego w poprzednim etapie projektowania bazy wiedzy systemu $\boldsymbol{R\_2U}$. W zbiorze tym znajdują się, między innymi:

– reguła R_{e1vir_c}, w postaci:

$$\begin{aligned}&\mathrm{grf}(0.86):\\&\quad \mathrm{Wydz\_Rodz} = \{\mathrm{WInf\_st}, \mathrm{WInf\_nst}\} \oplus,\\&\quad \mathrm{Kier} = \{\mathrm{Inf}, \mathrm{AiR}\} \oplus\\&\to\ \mathrm{sprzet} : \{\mathrm{K\_st}, \mathrm{Lap}\} \odot : \mathrm{irf}(0.84)\ ,\end{aligned}$$

– reguła R_{e1vir_e}, w postaci:

$$\begin{aligned}&\mathrm{grf}(0.66):\\&\quad \mathrm{Wydz\_Rodz} = \{\mathrm{WInf\_st}, \mathrm{WInf\_nst}\} \oplus,\\&\quad \mathrm{Kier} = \{\mathrm{Inf}, \mathrm{AiR}\} \oplus,\\&\quad \mathrm{Plec} = \{\mathrm{M}\} \odot\\&\to\ \mathrm{sprzet} : \{\mathrm{K\_st}, \mathrm{Lap}\} \odot : \mathrm{irf}(0.59)\ ,\end{aligned}$$

– reguła R_{e1vir_g}, w postaci:

$$\begin{aligned}&\mathrm{grf}(0.66):\\&\quad \mathrm{Wydz\_Rodz} = \{\mathrm{WInf\_st}, \mathrm{WInf\_nst}\} \oplus,\\&\quad \mathrm{Kier} = \{\mathrm{Inf}, \mathrm{AiR}\} \oplus,\\&\quad \mathrm{Rok} = \{3\} \odot\\&\to\ \mathrm{sprzet} : \{\mathrm{K\_st}, \mathrm{Lap}\} \odot : \mathrm{irf}(0.59)\ ,\end{aligned}$$

– reguła R_{e1vir_h}, w postaci:

$$\begin{aligned}&\mathrm{grf}(0.66):\\&\quad \mathrm{Wydz\_Rodz} = \{\mathrm{WInf\_st}, \mathrm{WInf\_nst}\} \oplus,\\&\quad \mathrm{Kier} = \{\mathrm{Inf}, \mathrm{AiR}\} \oplus,\\&\quad \mathrm{Plec} = \{\mathrm{M}\} \odot,\\&\quad \mathrm{Rok} = \{3\} \odot\\&\to\ \mathrm{sprzet} : \{\mathrm{K\_st}, \mathrm{Lap}\} \odot : \mathrm{irf}(0.59)\ .\end{aligned}$$

Łatwo zauważyć, że dla każdej pary reguł: $(R_{e1vir\_c}, R_{e1vir\_e})$, $(R_{e1vir\_c}, R_{e1vir\_g})$ i $(R_{e1vir\_c}, R_{1evir\_h})$ zachodzi warunek 6.13, świadczący o tym, że przynajmniej jedna z reguł w parze jest zbudowana nieprawidłowo. Na podstawie tego spostrzeżenia, przytoczone reguły zostaną oznaczone etykietą „podejrzana”: reguła R_{e1vir_c} – trzykrotnie, a reguły R_{e1vir_e}, R_{e1vir_g} i R_{e1vir_h} – jednokrotnie.

Dla całego, 12-elementowego zbioru reguł zostanie wyodrębnionych 9 par reguł, dla których zachodzi zależność 6.13 lub 6.14. Równocześnie, żadna para reguł nie będzie pozostawać w sprzeczności (def. 6.7). W rezultacie, po przeprowadzeniu pełnego badania, 4 reguły (R_{e1vir_c}, R_{e1vir_e}, R_{e1vir_h}, R_{e1vir_n}) zostaną oznaczone etykietą „podejrzana” trzykrotnie, 2 reguły (R_{e1vir_g}, R_{e1vir_p}) – dwukrotnie, oraz 2 reguły (R_{e1vir_a}, R_{e1vir_i}) – jednokrotnie.

Po usunięciu ze zbioru reguły R_{e1vir_c}, w zbiorze pozostanie 6 par reguł spełniających zależność 6.13 lub 6.14. Dalsza eliminacja „podejrzanych” reguł, w kolejności R_{e1vir_n} (identycznej z regułą R_{e1vir_c}) i R_{e1vir_p}, doprowadzi do uzyskania zbioru 9 reguł, z których dwie (R_{e1vir_a}, R_{e1vir_i}) nadal będą „podejrzane”. Ponieważ obie te reguły są oznaczone etykietą „podejrzana” jednokrotnie, ostatnia operacja usunięcia reguły odbędzie się na podstawie dodatkowego kryterium – zewnętrznego współczynnika wiarygodności grf. Ze względu na zależność: $\mathrm{grf}(R_{e1vir\_a}) = 0.81 < 0.83 = \mathrm{grf}(R_{e1vir\_i})$, ze zbioru zostanie usunięta reguła R_{e1vir_a}. □

Przykład 7.6. Ostatni etap projektowania bazy wiedzy systemu $\boldsymbol{R\_2U}$ polega na poszukiwaniu i eliminacji reguł nadmiarowych. W bieżącym, 8-elementowym zbiorze reguł, jedyną nadmiarową będzie reguła R_{e1vir_h}, w postaci:

$$\begin{aligned}
&\mathrm{grf}(0.66) : \\
&\quad \mathrm{Wydz_Rodz} = \{\mathrm{WInf_st}, \mathrm{WInf_nst}\} \oplus, \\
&\quad \mathrm{Kier} = \{\mathrm{Inf}, \mathrm{AiR}\} \oplus, \\
&\quad \mathrm{Plec} = \{\mathrm{M}\} \odot, \\
&\quad \mathrm{Rok} = \{3\} \odot \\
&\rightarrow \quad \mathrm{sprzet} : \{\mathrm{K_st}, \mathrm{Lap}\} \odot : \mathrm{irf}(0.59)\ .
\end{aligned}$$

Jest ona subsumowana zarówno przez regułę R_{e1vir_e}:

$$\begin{aligned}
&\mathrm{grf}(0.66) : \\
&\quad \mathrm{Wydz_Rodz} = \{\mathrm{WInf_st}, \mathrm{WInf_nst}\} \oplus, \\
&\quad \mathrm{Kier} = \{\mathrm{Inf}, \mathrm{AiR}\} \oplus, \\
&\quad \mathrm{Plec} = \{\mathrm{M}\} \odot \\
&\rightarrow \quad \mathrm{sprzet} : \{\mathrm{K_st}, \mathrm{Lap}\} \odot : \mathrm{irf}(0.59)\ ,
\end{aligned}$$

jak i regułę R_{e1vir_g}:

$$\begin{aligned}
&\mathrm{grf}(0.66): \\
&\quad \mathrm{Wydz_Rodz} = \{\mathrm{WInf_st}, \mathrm{WInf_nst}\} \oplus, \\
&\quad \mathrm{Kier} = \{\mathrm{Inf}, \mathrm{AiR}\} \oplus, \\
&\quad \mathrm{Rok} = \{3\} \odot \\
&\rightarrow \quad \mathrm{sprzet}: \{\mathrm{K_st}, \mathrm{Lap}\} \odot : \mathrm{irf}(0.59)\ .
\end{aligned}$$

Po usunięciu ze zbioru reguły R_{e1vir_h}, do bazy wiedzy $\mathrm{KB}_r(\boldsymbol{R\_2U})$ projektowanego systemu $\boldsymbol{R\_2U}$ kandydują ostatecznie reguły:

$$\begin{aligned}
&\mathrm{grf}(0.66): && (R_{e1vir_e}) \\
&\quad \mathrm{Wydz_Rodz} = \{\mathrm{WInf_st}, \mathrm{WInf_nst}\} \oplus, \\
&\quad \mathrm{Kier} = \{\mathrm{Inf}, \mathrm{AiR}\} \oplus, \\
&\quad \mathrm{Plec} = \{\mathrm{M}\} \odot \\
&\rightarrow \quad \mathrm{sprzet}: \{\mathrm{K_st}, \mathrm{Lap}\} \odot : \mathrm{irf}(0.59)\ ,
\end{aligned}$$

$$\begin{aligned}
&\mathrm{grf}(0.85): && (R_{e1vir_f}) \\
&\quad \mathrm{Wydz_Rodz} = \{\mathrm{WInf_st}, \mathrm{WInf_nst}\} \oplus, \\
&\quad \mathrm{Kier} = \{\mathrm{Inf}, \mathrm{AiR}\} \oplus, \\
&\quad \mathrm{Rok} = \{3\} \odot \\
&\rightarrow \quad \mathrm{plec} = \{\mathrm{M}\} \odot : \mathrm{irf}(1.00)\ ,
\end{aligned}$$

$$\begin{aligned}
&\mathrm{grf}(0.66): && (R_{e1vir_g}) \\
&\quad \mathrm{Wydz_Rodz} = \{\mathrm{WInf_st}, \mathrm{WInf_nst}\} \oplus, \\
&\quad \mathrm{Kier} = \{\mathrm{Inf}, \mathrm{AiR}\} \oplus, \\
&\quad \mathrm{Rok} = \{3\} \odot \\
&\rightarrow \quad \mathrm{sprzet}: \{\mathrm{K_st}, \mathrm{Lap}\} \odot : \mathrm{irf}(0.59)\ ,
\end{aligned}$$

$$\begin{aligned}
&\mathrm{grf}(0.83): && (R_{e1vir_i}) \\
&\quad \mathrm{Wydz_Rodz} = \{\mathrm{WInf_st}\} \odot, \\
&\quad \mathrm{Kier} = \{\mathrm{Inf}\} \odot \\
&\rightarrow \quad \mathrm{plec} = \{\mathrm{M}\} \odot \ \text{ with } \mathrm{irf}(0.91)\ ,
\end{aligned}$$

$$\begin{aligned}
&\mathrm{grf}(0.80): && (R_{e1vir_j}) \\
&\quad \mathrm{Wydz_Rodz} = \{\mathrm{WInf_st}\} \odot, \\
&\quad \mathrm{Kier} = \{\mathrm{Inf}\} \odot \\
&\rightarrow \quad \mathrm{rok} = \{2{,}3\} \oplus : \mathrm{irf}(0.87)\ ,
\end{aligned}$$

$$\begin{aligned}
&\mathrm{grf}(0.89): && (R_{e1vir_k}) \\
&\quad \mathrm{Wydz_Rodz} = \{\mathrm{WInf_st}\} \odot, \\
&\quad \mathrm{Kier} = \{\mathrm{Inf}\} \odot \\
&\rightarrow \quad \mathrm{sprzet}: \{\mathrm{Lap}\} \odot : \mathrm{irf}(0.96)\ ,
\end{aligned}$$

$$\begin{aligned}
&\mathrm{grf}(0.90): && (R_{e1vir_l}) \\
&\quad \mathrm{Wydz_Rodz} = \{\mathrm{WInf_st}, \mathrm{WInf_nst}\} \oplus, \\
&\quad \mathrm{Kier} = \{\mathrm{Inf}, \mathrm{AiR}\} \oplus \\
&\rightarrow \quad \mathrm{plec} = \{\mathrm{M}, \mathrm{K}\} \oplus : \mathrm{irf}(1.00)\ \}
\end{aligned}$$

Pobieżna analiza przytoczonych reguł prowadzi do wniosku, że ostatnia z nich (R_{e1vir_l}) ma charakter nadmiarowy. Wynika on nie z istnienia w bazie wiedzy $\mathrm{KB}_r(\boldsymbol{R\_2U})$ reguły subsumującej regułę R_{e1vir_l}, lecz ze znajomości dziedziny wartościowania atrybutu Plec. Następujący fakt ma charakter aksjomatyczny:

$$
\begin{aligned}
&\mathrm{grf}(1.00) :\\
&\quad \mathrm{true}\\
&\rightarrow \quad \mathrm{plec} = \{\mathrm{M}, \mathrm{K}\} \oplus : \mathrm{irf}(1.00)\ \}
\end{aligned}
$$

W sposób oczywisty powyższy fakt subsumuje regułę R_{e1vir_l}. Faktu tego nie da się jednak wygenerować w procesie klasycznego uczenia indukcyjnego. Można to uczynić jedynie w oparciu o konceptualizację dziedziny (patrz rozdz. 8). □

Użyte w przykładach dane atrybutowe zbiorcze są danymi sztucznymi, skonstruowanymi dla celów ilustracyjnych. Choć uprawdopodobniono je tak, by mogły pochodzić z rzeczywistego zbioru danych, to jednak ich liczba nie uprawnia do uznania wyindukowanych reguł za realne, a zaproponowanej bazy wiedzy za pełną. W systemie z tak małą bazą wiedzy nie można by przeprowadzić żadnego ciekawego wnioskowania.

7.6 Ocena jakości systemu R_2U

Prezentację systemu $\boldsymbol{R\_2U}$ można podsumować przy użyciu metryki jakości zaproponowanej w podrozdziale 5.2. Przed przystąpieniem do podsumowania warto przypomnieć, że – zgodnie z definicją 6.5 – baza wiedzy dowolnego systemu $\boldsymbol{R\_2U}$ musi spełniać następujące wymagania jakościowe:

- wszystkie reguły w bazie wiedzy mają poprawną budowę (dodatni monotoniczny charakter zależności między przesłanką a konkluzją reguły) i „pochodzą” od odpowiednio dużego zbioru danych rzeczywistych,
- baza wiedzy jest wolna od sprzeczności i nadmiaru.

Spośród ośmiu zalecanych dychotomicznych kryteriów jakości, każdy system $\boldsymbol{R\_2U}$ spełnia bezwzględnie pięć pierwszych (str. 80, (a) – (e)). W celu spełnienia kryterium szóstego (f), system $\boldsymbol{R\_2U}$ można wyposażyć w (opcjonalnie działający) moduł do precyzyjnego objaśniania odpowiedzi uzyskiwanych na poszczególnych etapach procesu wnioskowania. Dla odmiany, spełnienie kryterium siódmego (g) może się okazać w pewnych obszarach aplikacyjnych trudne. Jeśli oczekiwania wydajnościowe względem systemu $\boldsymbol{R\_2U}$ są wysokie, to jedyny sposób ich realizacji może polegać na wyposażeniu silnika wnioskującego w mechanizm „przycinania” procesów wnioskowania. Jego użycie może jednak pociągać za sobą jawne lub niejawne obniżenie wiarygodności niektórych wygenerowanych konkluzji, a także niewygenerowanie innych, poprawnych konkluzji. Ostatnie, ósme na liście kryterium otwartości systemu na modyfikacje (h) należy w wypadku systemu $\boldsymbol{R\_2U}$ uznać za bardzo trudne do zrealizowania. Aktualizacja stanu bazy wiedzy na podstawie zmiany/przyrostu danych ewidencyjnych z pewnością nie może odbywać się „w locie", w toku normalnej pracy systemu.

W celu oszacowania średniej trafności odpowiedzi udzielanych przez system $\boldsymbol{R\_2U}$ w procesach wnioskowania należy przeprowadzić eksperyment, polegający na badaniu występowania w zbiorze odpowiedzi systemu – odpowiedzi oczekiwanych. Daną wejściową dla procesu wnioskowania uzyskuje się z wybranej danej rzeczywistej, przez zastąpienie losowej liczby wartości losowo wybranych atrybutów danej – wartością „nieznaną". Praktycznie oznacza to, że wśród faktów aksjomatycznych dla procesu wnioskowania nie wystąpią fakty z atrybutami o wartościach „nieznanych". Wnioskowanie będzie skuteczne z punktu widzenia atrybutu wyjściowo „nieznanego", jeśli po jego zakończeniu w zbiorze odpowiedzi systemu znajdzie się przynajmniej jedna dotycząca tego atrybutu. W celu uzyskania miarodajnych wyników, należy przeprowadzić odpowiednio dużą liczbę skutecznych procesów wnioskowania. Dobór danych rzeczywistych do prowadzenia eksperymentu powinien być zgodny z rozkładem prawdopodobieństwa występowania tych danych w dziedzinie aplikacji systemu.

Do oszacowania średniej trafności odpowiedzi systemu niezbędna jest znajomość pozycji badanej konkluzji na liście rankingowej wszystkich konkluzji związanych z tym samym atrybutem, co badana. W wypadku systemu $\boldsymbol{R\_2U}$ ta pozycja jest tym wyższa, im wyższy jest stopień wiarygodności konkluzji, w pierwszej kolejności – wewnętrznej irf, w drugiej – zewnętrznej grf.

Teoretyczną ocenę jakości bazy wiedzy systemu $\boldsymbol{R\_2U}$ przeprowadza się w oderwaniu od jakichkolwiek obliczeń. Konfrontacja tej oceny z oceną działania systemu w warunkach rzeczywistych może prowadzić do następujących wniosków i potencjalnych zmian:

- brak zgodności obu ocen, przy równoczesnej niskiej ocenie trafności odpowiedzi systemu stanowi podstawę do zmodyfikowania, najpierw – algorytmu wyznaczania wartości współczynników grf reguł, a w dalszej kolejności – algorytmu działania silnika wnioskującego systemu;
- brak zgodności obu ocen, przy równoczesnej wysokiej ocenie trafności odpowiedzi systemu oznacza, że wartości współczynników grf mogą być niedoszacowane; w takiej sytuacji można ewentualnie przeprowadzić ich przeskalowanie;
- zgodność obu ocen, przy równoczesnej niskiej ocenie trafności odpowiedzi systemu może oznaczać, że ewidencja na podstawie której wygenerowano reguły z bazy wiedzy systemu była niewielka i niereprezentatywna; w takiej sytuacji należy pomyśleć o poszerzeniu ewidencji (patrz podrozdz. 8.2 i 8.3) i ponownej indukcji reguł 2U;

– zgodność obu ocen, przy równoczesnej wysokiej ocenie trafności odpowiedzi systemu stanowi podstawę do uznania systemu $\boldsymbol{R\_2U}$ za system odpowiedniej jakości.