PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU nr 163
Ekonometria 30 2011

Jadwiga Sobieska-Karpińska
Uniwersytet Ekonomiczny we Wrocławiu

Marcin Hernes
Społeczna Wyższa Szkoła Przedsiębiorczości i Zarządzania w Łodzi

METODY CONSENSUSU W ODNIESIENIU DO NIEPEŁNYCH POKRYĆ UPORZĄDKOWANYCH

Streszczenie: W niniejszym artykule przedstawiono metody consensusu dla niepełnych pokryć uporządkowanych, które w wielu przypadkach stanowią strukturę informacji ekspertowej. Struktura ta znajduje zastosowanie wtedy, gdy ekspert sklasyfikuje elementy zbioru tak, że każdy element należy do jednej klasy lub do wielu klas. Wybór consensusu jest jedną z metod rozwiązywania konfliktów wiedzy w systemach rozproszonych. W artykule przedstawiono definicję niepełnego pokrycia uporządkowanego, przekształcenia niepełnego pokrycia uporządkowanego w inne niepełne pokrycie uporządkowane, definicję funkcji odległości między niepełnymi pokryciami uporządkowanymi, algorytmy consensusu według kryterium C_1 i C_2. Ponieważ algorytm consensusu według kryterium C_2 jest algorytmem heurystycznym, przedstawiono wyniki badania efektywności tego algorytmu.

Słowa kluczowe: systemy rozproszone, konflikty wiedzy w systemach, metody consensusu, struktury wiedzy ekspertowej, niepełne pokrycia uporządkowane.

1. Wstęp

Metody consensusu znajdują zastosowanie w rozwiązywaniu konfliktów w systemach rozproszonych. Konflikty takie mogą wystąpić np. w systemach ekspertowych czy też wieloagentowych w sytuacji, gdy eksperci lub agenty będą miały różne opinie na ten sam temat [Chantemargue i in. 1998]. Inaczej mówiąc, rozwiązania danego problemu przez poszczególnych ekspertów/agentów będą się różnić. Użytkownik systemu natomiast chce otrzymać tylko jedno rozwiązanie problemu, aby na jego podstawie podjąć odpowiednią decyzję. W takim przypadku możemy wykorzystać metody consensusu do uzgodnienia jednej, wspólnej wersji rozwiązania.

Dotychczas opracowano metody consensusu dla wielu struktur wiedzy eksperta (lub agenta), np. dla podziałów uporządkowanych, pokryć uporządkowanych, pokryć i podziałów hierarchicznych, niepełnych podziałów uporządkowanych hierarchicznych i ważonych niepełnych podziałów uporządkowanych [Sobieska-Karpińska, Hernes 2008; 2009; Hernes, Nguyen 2004]. Nie zostały natomiast określone metody consensusu w odniesieniu do niepełnych pokryć uporządkowanych, wobec tego w niniejszym artykule postanowiono te metody opracować.

2. Niepełne pokrycia uporządkowane

Niepełne pokrycia uporządkowane nie zostały dotychczas zdefiniowane, jednak w wielu przypadkach stanowią strukturę wiedzy eksperta czy też agenta. Z taką strukturą mamy do czynienia wtedy, gdy ekspert lub agent sklasyfikuje elementy zbioru tak, że każdy element należy do jednej klasy lub do wielu klas. Przykładem może być klasyfikacja owoców ze względu na kolor. Na przykład część z nich klasyfikujemy jako czerwone, część jako zielone, natomiast niektóre mogą być zarówno czerwone, jak i zielone. Niepełne pokrycie uporządkowane definiowane jest następująco:

Definicja 1
K-klasowym niepełnym pokryciem uporządkowanym skończonego zbioru $X = \{x_1, \ldots, x_N\}$ nazywamy dowolny ciąg $P = \langle P_1, \ldots, P_K \rangle$, gdzie:

$$\bigcup_{i=1,2,\ldots,K} P_i = X. \tag{1}$$

Zbiór wszystkich *K*-klasowych niepełnych pokryć uporządkowanych zbioru *X* oznaczamy jako $NV_K(X)$.

Alternatywna definicja niepełnego pokrycia uporządkowanego przedstawia się następująco:

Definicja 2
Każde *K*-klasowe niepełne pokrycie uporządkowane może być reprezentowane za pomocą macierzy charakterystycznej:

$$P = \left[p_{i,j}\right] = \begin{bmatrix} p_{11} & . & . & . & p_{1N} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ p_{K1} & . & . & . & p_{KN} \end{bmatrix}, \tag{2}$$

gdzie: $p_{i,j} = \begin{cases} 1, & \text{jeśli } x_j \in P_i \\ 0, & \text{jeśli } x_j \notin P_i \end{cases}.$

Niepełne pokrycie uporządkowane ma następujące cechy:
– w niepełnym pokryciu uporządkowanym elementy zbioru *X* mogą się powtarzać,
– pokrycie to musi składać się ze wszystkich elementów zbioru *X*,

– ważna jest kolejność klas, kolejność elementów w klasach nie jest istotna,
– niepełne pokrycie uporządkowane może składać się z dowolnej liczby klas,
– dana klasa pokrycia może składać się z dowolnej liczby elementów zbioru *X*.

W odniesieniu do niepełnych pokryć uporządkowanych można zdefiniować następujące przekształcenia:

Definicja 3

Niech $P = \langle P_1, \ldots, P_k, \ldots, P_l, \ldots, P_K \rangle \in V_K(X)$ oraz $x_n \notin P_l$. Przekształcenie:

$$D_n^l(P) = \langle P_1, \ldots, P_k, \ldots, P_l \cup \{x_n\}, \ldots, P_K \rangle \quad (3)$$

nazywamy dodaniem elementu x_n do klasy P_l.

Definicja 4

Niech $P = \langle P_1, \ldots, P_k, \ldots, P_l, \ldots, P_K \rangle \in vV_K(X)$ oraz $x_n \in P_k$. Przekształcenie:

$$E_n^k(P) = \langle P_1, \ldots, P_k \setminus \{x_n\}, \ldots, P_l, \ldots, P_K \rangle \quad (4)$$

nazywamy eliminacją elementu x_n z klasy P_k.

Na podstawie tych przekształceń zdefiniowana zostanie funkcja odległości.

Definicja 5

Odległością μ (P,Q) pomiędzy niepełnymi pokryciami uporządkowanymi *P i Q* $(P,Q \in NU_K(X))$ nazywamy minimalną liczbę dodawań, eliminacji potrzebnych do przekształcenia niepełnego pokrycia uporządkowanego *P* w niepełne pokrycie uporządkowane Q.

Z definicji odległości wynika następujące twierdzenie, które w sposób alternatywny można wykorzystać do obliczania odległości.

Twierdzenie 1

$$\mu\ (P,Q) = \sum_{i=1}^{K} \sum_{j=1}^{N} \left| p_{ij} - q_{ij} \right|, \quad (5)$$

gdzie p_{ij} i q_{ij} są elementami macierzy *P*, *Q* będących reprezentacjami pokryć *P*, *Q* odpowiednio.

Wykorzystując odległość, można przejść do określenia metod consensusu.

3. Metody consensusu dla niepełnych pokryć uporządkowanych

Metody wyznaczania consensusu w odniesieniu do niepełnych pokryć uporządkowanych nie zostały dotychczas opracowane. Aby wyznaczyć ten consensus, można wykorzystać funkcje (kryteria) wyboru reprezentacji zbiorów oraz ich własności przedstawione w pracy [Daniłowicz, Nguyen 1992]. W niniejszym artykule jako kryteria wyznaczania consensusu przyjmiemy funkcję $C_{n,}$ gdzie $n = 1, 2,...,$ zdefiniowaną w pracy [Daniłowicz, Nguyen 1987] w następujący sposób:

$$C_n : 2^{NU_K(X)} \to NV_K(X) . \tag{6}$$

Niech dany będzie profil $Z \in 2^{NV_K(X)}$. Funkcja:

$$\omega_n(x,Z) = \sum_{i=1}^{M} \left[\omega\left(x, Z^i\right) \right]^n \tag{7}$$

dla $n = 1,2,....,$ $x \in X$, M – liczebność zbioru Z.

Definicja 6

$$x \in C_n(Z) \leftrightarrow \omega_n(x,Z) = \min_{y\ \in NU_K(X)} \omega_n(y,Z). \tag{8}$$

Ponieważ zostało udowodnione w [Nguyen 2002], że od pewnej wartości n consensus będzie miał taką samą postać, w niniejszym artykule zajmiemy się wyborem consensusu dla $n = 1$ oraz $n = 2$. Wiadomo, że dla $n = 1$ consensus jest bardzo podobny do jednego z elementów profilu, natomiast dla $n = 2$ odległości pomiędzy consensusem a elementami profilu mogą być większe, lecz bardziej równomierne. Decyzję, który consensus jest lepszy, podejmujemy na podstawie charakteru konkretnego zadania, które chcemy rozwiązać metodą consensusu. Przejdźmy zatem do wyznaczenia consensusu na podstawie kryterium C_1.

Twierdzenie 2

Dany jest zbiór $X = \{x_1, \ldots, x_N\}$. Niech $t_i(j)$ będzie liczbą wystąpień elementu x_j w klasach Z_i^L, ($i = 1,..., K$; $j = 1, ..., N$; $L = 1, ..., M$) profilu $Z = \left\{Z^{(1)}, Z^{(2)}, \ldots, Z^{(M)}\right\}$ oraz $CON = \left\langle CON_1, CON_2, \ldots, CON_K \right\rangle$ będzie consensusem spełniającym kryterium C_1 względem danego profilu, wówczas każdy element x_j należy w consensusie CON do klasy z indeksem, pod którym występuje on najwięcej razy w profilu P.

Dowód 1

Niech $CON = \left\langle CON_1, CON_2, \ldots, CON_K \right\rangle$ będzie consensusem spełniającym kryterium C_1 względem profilu $Z = \left\{Z^{(1)}, Z^{(2)}, \ldots, Z^{(M)}\right\}$. Z twierdzenia 2 wynika, że:

$$\sum_{i=1}^{M} \left[\mu\left(CON, Z^{(i)}\right) \right] = \min_{Q \in \mathrm{VU_K} X} \sum_{i=1}^{M} \sum_{i=1}^{K} \sum_{j=1}^{N} (q_{ij} - z_{ij}^{(i)}) = \min_{Q \in \mathrm{VU_K} X} \sum_{i=1}^{K} \sum_{j=1}^{N} \sum_{i=1}^{M} (q_j - z_j^{(i)}). \tag{9}$$

Ponieważ $\sum_{i=1}^{M} (q_j - z_j^{(i)})$ dla $j=1,...,N$ zależy wyłącznie od zmiennej q, więc

$$\min_{Q \in \mathrm{VU_K} X} \sum_{i=1}^{K} \sum_{j=1}^{N} \sum_{i=1}^{M} (q_j - z_j^{(i)}) = \min_{q_1} \sum_{i=1}^{M} (q_1 - z_1^{(i)}) + \ldots + \min_{q_N} \sum_{i=1}^{M} (q_N - z_N^{(i)}). \tag{10}$$

Składniki sumy $\sum_{i=1}^{M}(q_j - z_j^{(i)})$ są równe 0 w przypadku, gdy $q_j = z_j^{(i)}$, w przeciwnym wypadku są równe 1. Suma ta osiąga minimum wtedy, gdy wybierzemy q równe wartości najczęściej występującej wśród $z_j^{(i)}$.

Na podstawie twierdzenia 2 możemy opracować algorytm consensusu według kryterium C_1 dla niepełnych pokryć uporządkowanych. Algorytm działa w ten sposób, że dla każdego elementu zbioru X sprawdzamy, ile razy wystąpił ten element w danej klasie wszystkich pokryć i zapamiętujemy maksymalną liczbę wystąpień. Następnie umieszczamy ten element w consensusie w klasach, w których wystąpił on najwięcej razy i przechodzimy do kolejnego elementu zbioru X. Algorytm kończy się w momencie sprawdzenia wszystkich elementów, a otrzymany consensus jest consensusem według kryterium C_1.

Dane: Profil $Z = \{Z^{(1)}, Z^{(2)}, ..., Z^{(M)}\}$ zbioru X.

Wynik: Consensus $CON = \langle CON_1, CON_2, \ldots, CON_K \rangle$ według kryterium C_1 względem Z.

BEGIN

Krok 1: Przyjmujemy $CON_1 = CON_2 = \ldots = CON_K = \varnothing$.

Krok 2: Przyjmujemy j:=1.

Krok 3: Przyjmujemy i:=1, max:=0.

Krok 4: Jeżeli $t_i(j) > max$, to $max = t_i(j)$.

Krok 5: Jeżeli $i<K$, to $i:=i+1$. Przechodzimy do kroku 4.

Krok 6: Przyjmujemy i:=1.

Krok 7: Jeżeli $t_i(j) = max$, to $CON_i := CON_i \cup \{x_j\}$.

Krok 8: Jeżeli $i<K$, to $i:=i+1$. Przechodzimy do kroku 7.

Krok 9: Jeżeli $j<N$, to $j:=j+1$. Przechodzimy do kroku 3. Jeżeli $j \geq N$, to END.

END.

Złożoność algorytmu wynosi O(NKM).

Następnie określimy consensus na podstawie kryterium C_2. Wyznaczanie consensusu według kryterium C_2 jest problemem NP-zupełnym [Kamel 1994]. Bierzemy tutaj pod uwagę sumę kwadratów odległości. Z dokonanych obserwacji wynika, że consensus według kryterium C_2 jest wynikiem przekształcenia consensusu według kryterium C_1. Poniżej opracowano algorytm heurystyczny, pozwalający na wyznaczenie consensusu według kryterium C_2. Wynik algorytmu jest oczywiście wynikiem przybliżonym. Algorytm ten działa w następujący sposób:

Wyznaczany jest consensus według kryterium C_1 oraz obliczany jest kwadrat odległości między consensusem a profilem i przyjmowany jako minimum. Dla każdego elementu zbioru X sprawdzamy, czy występuje on w danej klasie w consensusie. Jeśli występuje, to usuwamy go z tej klasy i liczymy kwadrat odległości. Jeśli jest większy, to przechodzimy do następnej klasy, jeśli jest mniejszy, to przyjmujemy go jako consensus, a odległość do profilu jako minimalną. Jeśli element nie występuje w danej klasie w consensusie, to obliczamy, ile razy występuje w danej klasie we wszystkich pokryciach profilu. Jeśli nie wystąpi ani razu, to przechodzimy do następnej klasy, jeśli natomiast wystąpi chociaż raz, to umieszczamy go w tej klasie w consensusie oraz sprawdzamy, czy odległość otrzymanego consensusu jest mniejsza od poprzedniego. Jeśli nie, to zostawiamy poprzedni consensus jako najlepszy, jeśli tak, to przyjmujemy nowy consensus jako najlepszy, a odległość do profilu jako minimalną. Po sprawdzeniu wszystkich klas przechodzimy do kolejnego elementu zbioru X. Algorytm kończy działanie wtedy, gdy sprawdzone zostaną wszystkie elementy zbioru X, a otrzymany consensus jest consensusem według kryterium C_2.

Dane: Profil $Z = \{Z^{(1)}, Z^{(2)}, ..., Z^{(M)}\}$ zbioru X.

Wynik: Consensus $CON = \langle CON_1, CON_2, \ldots, CON_K \rangle$ według kryterium C_2 względem Z.

BEGIN

Krok 1: Przyjmujemy CON jako consensus według kryterium C_1, niech $i{:}=1$, $j{:}=1$ oraz

$$d := \sum_{i=1}^{M} \left[\mathrm{m}\left(CON, Z^{(i)}\right) \right]^2 .$$

Krok 2: Jeżeli $x_j \in CON_i$ to:

$$CON' := \left\langle CON_1, CON_i \setminus \{x_j\} \ldots, CON_K \right\rangle .$$

Przechodzimy do kroku 5.

Jeżeli $x_j \notin CON_i$, to przechodzimy do kroku 3.

Krok 3: Jeżeli $t_i(j) = 0$, to przechodzimy do kroku 6,

Jeżeli $t_i(j) > 0$, to przechodzimy do kroku 4.

Krok 4: Jeżeli $x_j \cap CON \neq \varnothing$ oraz $x_j \in CON_z$, to

$$CON' := \left\langle CON_1, CON_i \cup \{x_j\} \ldots, CON_K \right\rangle .$$

Przechodzimy do kroku 5.

Krok 5: Jeżeli $\sum_{i=1}^{M} \left[\mu\left(CON', Z^{(i)}\right) \right]^2 < d$, to:

$$d := \sum_{i=1}^{M} \left[\mu\left(CON', Z^{(i)}\right) \right]^2$$

oraz

$$CON := CON' .$$

Przechodzimy do kroku 6.

Krok 6: Jeżeli $i<K$, to $i:=i+1$. Przechodzimy do kroku 2.
Jeżeli $i\geq K$, to $i=1$. Przechodzimy do kroku 7.
Krok 7: Jeżeli $j<N$, to $j:=j+1$. Przechodzimy do kroku 2.
Jeżeli $j\geq N$, to END.
END.

Złożoność algorytmu wynosi O(NK^{2}M).

Obliczanie consensusu według kryterium C_2 dla niepełnych pokryć uporządkowanych jest problemem NP-zupełnym. Consensus optymalny to taki consensus, którego suma kwadratów odległości od poszczególnych elementów profilu jest minimalna. Aby znaleźć taki consensus, musimy sprawdzić wszystkie kombinacje elementów zbioru X we wszystkich klasach niepełnego pokrycia uporządkowanego. W takim przypadku złożoność algorytmu wynosi O((K+1)^N). Jest to bardzo duża złożoność obliczeniowa. Zauważmy, że metody consensusu mają zastosowanie w systemach rozproszonych, gdzie użytkownik nie może długo czekać na wyniki. Algorytm heurystyczny opracowany w niniejszym artykule pozwala na szybkie wyznaczenie consensusu według kryterium C_2.

4. Badanie efektywności algorytmu heurystycznego

Efektywność algorytmu heurystycznego wyznaczania consensusu według kryterium C_2 określają następujące wielkości:

– współczynnik odchylenia od consensusu optymalnego mówi nam, o ile procent consensus obliczony za pomocą algorytmu heurystycznego jest gorszy od consensusu optymalnego,
– szybkość działania algorytmu.

W celu zbadania efektywności napisano program, który umożliwia obliczanie consensusu za pomocą algorytmu optymalnego i heurystycznego.

Symulacje wykonano na komputerze o następujących parametrach: procesor Celeron 2000MHz, pamięć 1 GB, twardy dysk 40 GB. Przeprowadzono 100 symulacji algorytmu optymalnego i heurystycznego. Każda symulacja składa się z następujących kroków: losowanie profilu (30 niepełnych pokryć uporządkowanych, każde pokrycie jest 8-klasowe), obliczanie consensusu optymalnego, obliczanie consensusu według algorytmu heurystycznego.

Wnioski z przeprowadzonych badań efektywności:

– consensus według kryterium C_2 obliczony za pomocą algorytmu heurystycznego (przedstawionego w niniejszym artykule) jest gorszy od optymalnego consensusu według kryterium C_2 średnio o 1,93%, co jest wynikiem bardzo dobrym,
– w ok. 60% przypadków consensus obliczony według algorytmu heurystycznego jest identyczny z consensusem optymalnym,
– najgorszy consensus według algorytmu heurystycznego jest oddalony od consensusu optymalnego o 11,20%,

– algorytm heurystyczny działa dużo szybciej, ponieważ jego złożoność wynosi $O(NK^2M)$, natomiast złożoność algorytmu optymalnego wynosi $O((K+1)^N)$, w programie symulacyjnym liczenie consensusu optymalnego zajmuje ok. 17 sekund, natomiast algorytm heurystyczny liczy consensus w czasie 0,2 sekundy.

5. Podsumowanie

W artykule opracowano metody consensusu dla niepełnych pokryć uporządkowanych, które są jedną ze struktur informacji eksperckiej (lub agenckiej) znajdującą wykorzystanie w m.in. systemach ekspertowych lub wieloagentowych. Wykorzystanie metod consensusu pozwala na uzgodnienie jednej wersji rozwiązania problemu i przedstawienie tej wersji użytkownikowi, dzięki czemu nie musi on się zastanawiać nad wyborem spośród wielu rozwiązań.

Wybór consensusu według kryterium C_2 jest problemem NP-zupełnym, więc opracowany algorytm jest algorytmem heurystycznym, dlatego w końcowej części dokonano badania efektywności algorytmu. Okazało się, że algorytm działa bardzo szybko, a przy tym otrzymany consensus w niewielkim stopniu odbiega od consensusu optymalnego.

Przedstawione w artykule zagadnienia można w przyszłości rozszerzyć o metody consensusu dla hierarchicznych niepełnych pokryć uporządkowanych oraz ważonych niepełnych pokryć uporządkowanych.

Literatura

Chantemargue F., Courant M., Dagaeff T., Robert A., *A Pragmatic Approach to Conflict*, ECAI 98, 13th European Conference on Artificial Intelligence, John Wiley & Sons, Ltd., 1998.

Daniłowicz C., Nguyen N.T., *Metody wyboru reprezentacji podziałów i pokryć uporządkowanych*, Oficyna Wydawnicza PWr, Wrocław 1992.

Daniłowicz C., Nguyen N.T., *The computation complexity of some nonlinear integer programming problems*, Raport BGiOINT, PWr, Seria PRE nr 86, Wrocław 1987.

Hernes M., Nguyen N.T., *Deriving Consensus for Incomplete Ordered Partitions*, [w:] *Intelligent Technologies for Inconsistent Knowledge Processing,* N.T. Nguyen (red.), Advanced Knowledge Intelligence, Australia 2004.

Kamel M., *Identifying, classifying, and resolving semantic conflicts in distributed heterogeneous databases*, „Journal of Database Management" 1994 no 6.

Nguyen N.T., *Metody wyboru consensusu i ich zastosowanie w rozwiązywaniu konfliktów w systemach rozproszonych*, Oficyna Wydawnicza Politechniki Wrocławskiej, 2002.

Sobieska-Karpińska J., Hernes M., *Susceptibility to Consensus of Conflict Situation in Intelligent Multi-Agent Decision Support System*, [w:] *Information Management*, B.F. Kubiak, A. Korowicki (red.), Gdansk University Press, Gdańsk 2009.

Sobieska-Karpińska J., Hernes M., *Rozwiązywanie konfliktów w systemach rozproszonych za pomocą metod consensusu*, [w:] Informatyka ekonomiczna nr 12, A. Nowicki (red.), UE, Wrocław 2008.

CONSENSUS METHODS FOR INCOMPLETE ORDERED COVERINGS

Summary: Consensus methods for solving conflicts in incomplete ordered coverings are presented in this paper. The incomplete ordered coverings are expert information structures in many cases. This structure should be useful when an expert classifies all elements of a set in such a way that every element belongs to one class or to many classes. The consensus methods are one of the methods for solving knowledge conflicts in distributed systems. A definition of the incomplete ordered covering, transformation of the incomplete ordered covering in the other incomplete ordered covering, the definition of distance among incomplete ordered covering and the algorithms of consensus methods for criterion C_1 and C_2 are depicted in this work. Because the consensus algorithm according to the criterion C_2 is a heuristic algorithm, the results of algorithm verification are presented.