

Dominik CUBALA, Longin STOLC  
Politechnika Gdańska

## GENEROWANIE PLANU ZAJĘĆ Z ZASTOSOWANIEM ALGORYTMÓW GENETYCZNYCH

**Streszczenie.** Poniższa praca opisuje problem automatycznego generowania rozkładu zajęć. Do tego celu został stworzony program PLANIX, który przy wykorzystaniu algorytmów genetycznych w kilka minut potrafi poradzić sobie ze średniej klasy problemem. Praca zawiera opis rdzenia programu oraz wyniki przeprowadzonych badań.

### USING GENETIC ALGORITHM IN AUTOMATED TIMETABLING

**Summary.** This paper shows subject matter of automated timetabling. We provide computer program which takes advantage of genetic algorithm and composes finished schedule in several minutes. This paper includes specification such as coding, initialization, construction of fitness function, selection and reproduction procedures. We also present results of our research.

#### 1. Wstęp

Ręczne ułożenie planu lekcji w średniej wielkości szkole zajmuje człowiekowi od kilku do kilkudziesięciu godzin. Co więcej, tak skonstruowany plan może być niedoskonały pod pewnymi względami. Z tych powodów zwrócono uwagę na możliwość automatyzacji procesu układania rozkładu zajęć. Do tego celu został stworzony program PLANIX.

Ze względu na ogromną liczbę ograniczeń i duże rozmiary zagadnienia metoda, polegająca na przejrzeniu całej przestrzeni rozwiązań, nie jest w ogóle brana pod uwagę. Wtedy konieczne jest zastosowanie jednej z wyspecjalizowanych technik obliczeniowych. Jedną z takich metod są niezwykle elastyczne algorytmy genetyczne. Więcej na ich temat można znaleźć w [2], [3], [4].

#### 2. Sformułowanie zagadnienia

##### 2.1. Reprezentacja genotypu i sposoby jego kodowania

Genotyp w systemie PLANIX reprezentowany jest w sposób bezpośredni. Każda sala odpowiada jednemu chromosomowi w genotypie. Każdy gen chromosomu reprezentuje inny przedział czasowy. Wszystkie geny chromosomów są zakodowane,

przyjmując wartości całkowite większe bądź równe zero. Wartość zero oznacza, iż gen jest wolny od zajęć. Pozostałe wartości odpowiadają numerowi zajęcia z zadanej przez układającego macierzy odwzorowującej wszystkie spotkania, które muszą zostać zaplanowane.

Genotyp w systemie PLANIX wskazuje jedynie dzień, godzinę lekcyjną oraz salę, w której odbywają się poszczególne zajęcia. Jednak korzystając z macierzy zajęć, można łatwo wywnioskować rozkłady zajęć dla poszczególnych klas oraz nauczycieli. Taki właśnie wyników genotyp z rozkodowanymi rozkładami zajęć dla wszystkich klas i nauczycieli przedstawia rysunek 1. Na jego podstawie można stwierdzić np., iż klasa  $c_0$  ma w poniedziałek na pierwszych trzech lekcjach w sali  $s_2$  geografę z nauczycielem  $t_1$ .

		Termin																		
		Poniedz.			Wtorek			Środa			Czwartek			Piątek						
		0	1	2	3	0	1	2	3	0	1	2	3	0	1	2	3			
sale	$s_0$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	11	0
	$s_1$	6	0	0	12	0	0	0	3	0	0	2	0	0	0	1	0	0	0	0
	$s_2$	4	0	8	0	0	0	7	0	0	5	0	0	0	9	0	0	0	0	0
kl.	$c_0$	4		12			7			2				10	1					
	$c_1$	6		8			3			5				9	11					
	$f_0$			12			3			2				10	1					
	$t_1$	4		8			7			5				9	11					
nauc.	$t_1$	4		8			7			5				9	11					
	$t_2$	6																		

Rys. 1. Przykładowy genotyp z rozkodowanymi rozkładami zajęć

## 2.2. Inicjalizacja genotypu

W systemie PLANIX zaimplementowano deterministyczno-losową strategię inicjalizacji genotypu. Każde zajęcie zostało umieszczone w jednej z przyporządkowanych temu zajęciu sal, tak aby termin i sala, w którym ma ono zostać umieszczone, nie były zabronione dla danej klasy, nauczyciela bądź sali. Na początku rozmieszczane są zajęcia, które mogą odbywać się tylko w jednej sali. Spośród tych zajęć pierwszeństwo mają te, których czas trwania wynosi 3 godziny lekcyjne. Proces ten powtarza się dla zajęć, które mogą odbywać się w dwóch, trzech salach itd., aż w genotypie zostaną umieszczone wszystkie zajęcia.

## 2.3. Funkcja oceny

Podczas konstrukcji planu musi być spełniony szereg różnych warunków. Dzieli się je na dwie grupy [1]:

- twarde – takie, które muszą być spełnione, aby rozkład był dopuszczalny,
- miękkie – takie, które powinny być spełnione, a liczba ewentualnych niespełnionych ograniczeń jest minimalizowana.

Funkcja oceny w systemie PLANIX (1) jest funkcją opartą na sumie dwóch funkcji ważonych: funkcji kary twardej oraz funkcji kary miękkiej.

$$f = f_{twarda} + f_{miękka} \quad (1)$$

gdzie:  $f$  – funkcja oceny,  $f_{twarda}$  – funkcja kary twardej,  $f_{miękka}$  – funkcja kary miękkiej.

Ze względu na rodzaj reprezentacji genotypu oraz sposób jego inicjalizacji większość ograniczeń twardej jest spełniona. Dwa różne zajęcia nigdy nie będą się

odbywały w tej samej sali, gdyż każdy chromosom w genotypie reprezentuje inną salę. Pewne jest również, iż każdy nauczyciel ma odpowiednią liczbę zajęć z każdą klasą, ponieważ właśnie takie pary klasa-nauczyciel umieszczane są w genotypie. Ponadto, okres trwania poszczególnych zajęć jest zawsze odpowiedni, gdyż podczas inicjalizacji zajęcia w genotypie umieszczane są na tylu genach, ile godzin trwa dane zajęcia. Sposób inicjalizacji genotypu zapewnia nam również, iż zajęcia odbywają się w jednej z dozwolonych sal oraz iż nie odbywają się w terminach, w których ze względu na ogólnie ustaloną niedostępność sal, nauczycieli bądź terminów nie mogą się odbywać.

Sposób reprezentacji oraz inicjalizacji genotypu nie zapewnia jednak stworzenia osiągalnego planu. Dlatego w systemie PLANIX zaimplementowano odpowiednie algorytmy, oceniające następujące twarde konflikty:

- konflikt nauczycieli ( $k_{nau}$ ) – jednoczesne prowadzenie wielu zajęć w tym samym terminie przez tego samego nauczyciela,
- konflikt klas ( $k_{kla}$ ) – jednoczesne odbywanie się wielu zajęć w tym samym terminie przez tę samą klasę.

Po wyznaczeniu wartości konfliktów nauczycieli oraz klas dla wszystkich zajęć w genotypie wyznacza się łączną karę twardą dla danego genotypu (2). Aby dany genotyp reprezentował dopuszczalny plan zajęć, funkcja ta musi mieć wartość równą zero.

$$f_{twarda} = \sum_{z=0}^{n_z-1} (w_{nau} k_{nau_z} + w_{kla} k_{kla_z}) \quad (2)$$

gdzie:  $n_z$  – liczba zajęć,  $w_{nau}$ ,  $w_{kla}$  – wagi konfliktu nauczycieli oraz klas,  $k_{nau_z}$ ,  $k_{kla_z}$  – liczba poszczególnych konfliktów dla danego zajęcia

Podobnie jak wśród ograniczeń twardych, również dla ograniczeń miękkich można wyróżnić kilka rodzajów konfliktów:

- konflikt okienek klas ( $k_{okk}$ ) – klasy nie powinny mieć okienek,
- konflikt godzin początkowych ( $k_{gpo}$ ) – klasy powinny zaczynać zajęcia na pierwszej godzinie lekcyjnej,
- konflikt równomierności ( $k_{row}$ ) – zajęcia dla poszczególnych klas powinny być równomiernie rozłożone na wszystkie dni tygodnia,
- konflikt przedmiotów ( $k_{prz}$ ) – dla tej samej klasy zajęcia z tego samego przedmiotu powinny być równomiernie rozłożone,
- konflikt okienek nauczycieli ( $k_{okn}$ ) – nauczyciele nie powinni mieć okienek,
- konflikt ilości ( $k_{ilo}$ ) – nauczyciel nie powinien przekraczać pewnej dziennej liczby zajęć w ciągu dnia.

#### 2.4. Selekcja i procedury reprodukcji

W systemie PLANIX zastosowano selekcję metodą rankingową. Liczba kopii każdego osobnika wprowadzana do puli rodzicielskiej ustalana jest w następujący sposób: każdy kolejno lepszy osobnik ma tyle razy większe szanse wejścia od swojego słabszego poprzednika, ile wynosi współczynnik selekcji.

Poza tym w systemie PLANIX zastosowano elitarną strategię reprodukcji. Najlepszy osobnik zawsze trafia do puli rodzicielskiej, a dodatkowo nie podlega on procesowi krzyżowania i mutacji, aby nie zostać zatraconym.

## 2.5. Operatory krzyżowania i mutacji

W systemie PLANIX zastosowano krzyżowanie równomierne, tzn. krzyżowaniu podlegają po kolei wszystkie zajęcia, dla danej pary rodziców, z prawdopodobieństwem 50% znalezienia się w każdym z dwóch nowych potomków.

Algorytm mutacji z określonym prawdopodobieństwem dla pojedynczego zajęcia wybiera nowy, wolny termin z zachowaniem reguł opisanych w procesie inicjalizacji.

## 3. Strojenie algorytmu genetycznego

Do przeprowadzenia symulacji wykorzystano rzeczywiste dane jednego z gimnazjów. W szkole tej uczniowie podzieleni są na 17 klas, w których zajęcia prowadzi 28 nauczycieli uczących jednego lub dwóch przedmiotów. Szkoła posiada 17 pomieszczeń lekcyjnych (w tym dwie sale gimnastyczne oraz pracownię komputerową). Łączna liczba wszystkich zajęć odbywających się w ciągu jednego tygodnia wynosi 420.

### 3.1. Badanie strategii inicjalizacji genotypu

W badaniu efektywności strategii inicjalizacji porównywano wartości funkcji kary dla najlepszego i najgorszego osobnika w populacji oraz średnią karę dla całej populacji przy różnych wartościach współczynnika zachłanności  $k$ .

Z przeprowadzonych eksperymentów wynika generalnie, że im większa jest wartość współczynnika zachłanności  $k$ , tym lepsze osobniki są generowane. Istnieje jednak pewna graniczna wartość  $k=100$ , przy której dalsze zwiększanie współczynnika zachłanności nie daje już efektu w postaci coraz to lepszych osobników. Należy w tym miejscu wspomnieć, iż zapotrzebowanie na moc obliczeniową rośnie liniowo wraz ze wzrostem  $k$ .

Do dalszych eksperymentów wartość współczynnika zachłanności zostaje ustalona na poziomie  $k=100$ .

### 3.2. Dobór parametrów algorytmu genetycznego

Aby odpowiednio dobrać parametry ewolucji, takie jak prawdopodobieństwa selekcji, mutacji oraz krzyżowania należało przeprowadzić wiele eksperymentów.

Na początku zbadano wpływ współczynnika selekcji dla trzech wartości: 1, 0,75, 0,5. Najlepiej ewolucja przebiega dla współczynnika selekcji równego 0,75. Średnia wartość kary dla tej populacji, dla tysięcznego pokolenia, wynosi 145 i jest niższa od średniej wartości kar dla innych populacji. Najlepszy osobnik z populacji, dla współczynnika selekcji równego 1, ma zbliżoną wartość kary co średnia dla populacji dla współczynnika równego 0,5. Wynika z tego, że średnio połowa genotypów ze zbioru osobników dla współczynnika 0,5 jest lepiej przystosowana niż najlepszy osobnik populacji ze współczynnikiem równym 1. Podobna zależność

występuje pomiędzy populacjami dla współczynników wynoszących odpowiednio 0,75 i 0,5. Średnio połowa populacji dla pierwszego współczynnika jest lepiej przystosowana niż najlepszy osobnik z populacji dla drugiego współczynnika. Zatem, do dalszych badań wartość współczynnika selekcji została ustalona na 0,75.

Kolejnym parametrem, który znacząco wpływa na proces ewolucji, jest prawdopodobieństwo krzyżowania. Najwyższą średnią karę, w okolicach 1000 pokolenia, odnotowujemy dla populacji, której prawdopodobieństwo krzyżowania wynosi 0,75. Zatem, wartość tego współczynnika krzyżowania można od razu odrzucić. Z kolei, dla wartości prawdopodobieństw równych 1 oraz 0,5, średnie wartości kar dla obydwu populacji są do siebie zbliżone. Przed 300 pokoleniem, osobniki z populacji z prawdopodobieństwem krzyżowania równym 0,5 są lepiej przystosowane od osobników z populacji z prawdopodobieństwem równym 1. Sytuacja ta powyżej 400 pokolenia wygląda dokładnie odwrotnie. Tak więc do dalszych eksperymentów prawdopodobieństwo krzyżowania ustalone zostaje na 1.

Kolejną ważną rzeczą, na drodze do ułożenia dobrego planu, jest odpowiedni dobór prawdopodobieństwa mutacji. Dla prawdopodobieństwa mutacji równego 0,005 średnia kara dla populacji znacznie odbiega od najlepszego osobnika. Wiadomo przecież, że z dwóch słabszych osobników ciężko jest utworzyć osobnika dobrego. Wniosek z tego jest jeden: mutacji podlega zbyt duża liczba genów.

Spróbujmy zatem zmniejszyć prawdopodobieństwo mutacji. Dla populacji z prawdopodobieństwem równym 0,002 najlepszy osobnik ma coraz mniejszą wartość funkcji kary. Jednak różnica między tym osobnikiem a średnią dla populacji jest ciągle zbyt duża. Z tego względu wartość kary dla tej populacji maleje bardzo powoli. O wiele gorzej sytuacja przedstawia się dla populacji z prawdopodobieństwem mutacji wynoszącym 0,0005. W tym wypadku różnica między średnią karą populacji a karą dla najlepszego osobnika jest zbyt mała. Efekt tego jest taki, że wszystkie osobniki stają się do siebie bardzo podobne. Ciężko jest zatem wygenerować nowy lepszy genotyp. Z powyższych spostrzeżeń nasuwa się jeden wniosek. Należy wybrać wartość prawdopodobieństwa mutacji z przedziału pomiędzy 0,0005 a 0,002. Taka właśnie wartość 0,0012 została wybrana dla trzeciej populacji. Przy takiej wartości prawdopodobieństwa mutacji algorytm osiąga bardzo dobre wyniki. Różnica pomiędzy średnią karą dla populacji a karą dla najlepszego osobnika jest w miarę stała i na tyle odpowiednia, żeby zachować różnorodność populacji.

### 3.3. Działanie algorytmu a wagi poszczególnych typów konfliktów

Do tej pory wszystkie eksperymenty przeprowadzane były dla pewnych ustalonych wartości wag poszczególnych typów konfliktów. Kolejne eksperymenty określają, w jaki sposób zmiana wartości każdego konfliktu z osobna wpływa na pozostałe konflikty. Można domyślać się, że zmniejszenie wartości kary jednego z konfliktów może niekorzystnie wpływać na zmiany innych konfliktów.

Po przeanalizowaniu wpływu poszczególnych wag konfliktów na wartość kary osobnika można dobrać takie wagi, które zapewnią, iż ułożony plan będzie planem dobrym, tzn. będzie planem osiągalnym, a dodatkowo pozostałe wartości konfliktów będą jak najbardziej zminimalizowane. Ostateczne dobrane wagi konfliktów przedstawia tabela 1.

Tabela 1  
Ostateczne wagi poszczególnych typów konfliktów

Waga	$w_{kla}$	$w_{nau}$	$w_{prz}$	$w_{okk}$	$w_{gpo}$	$w_{row}$	$w_{okn}$	$w_{ilo}$
Wartość	6	4	6	8	6	2	2	1

#### 4. Wnioski końcowe

Podsumowując, w pełni automatyczny, oparty na algorytmie genetycznym system PLANIX układa plan w kilka minut. Stworzony rozkład zajęć nie łamie żadnych twardych ograniczeń, a więc jest planem osiągalnym. Natomiast ograniczenia miękkie program stara się jak najbardziej zminimalizować, poprawiając w ten sposób jakość ułożonego planu.

Opisana metoda z powodzeniem może znaleźć zastosowanie w problemach o mniejszej skali złożoności, takich jak: planowanie dyżurów w szpitalu, harmonogramowanie obłożenia obiektów w klubie sportowym itp. Jej mocną stroną jest duża elastyczność i spora, jak na algorytm genetyczny, szybkość działania.

#### LITERATURA

1. Burke E. K., Elliman D. G., Weare R.: A Genetic Algorithm Based University Timetabling System. Practice and Theory of Automated Timetabling, 1996.
2. Goldberg D. E.: Algorytmy genetyczne i ich zastosowania. WNT, Warszawa 1998.
3. Michalewicz Z.: Algorytmy genetyczne + struktury danych = programy ewolucyjne. WNT, Warszawa 1999.
4. Rutkowska D., Piliński M., Rutkowski L.: Sieci neuronowe, algorytmy genetyczne i systemy rozmyte. PWN, Warszawa 1997.

Recenzent: Prof. dr hab. inż. Antoni Niederliński

#### Abstract

Manually compose timetabling in average school takes us from several to a few dozen hours. Therefore we use genetic algorithm to solve this problem automatically furthermore much quicker and better.

Genotype is represented in direct way. Each classroom replies one chromosome in genotype. Each chromosome's gene represents different time range. Value 0 means that gene is free from class. Remaining values reply number of class in class matrix.

We use random strategy of inicialization. Each class is putted in genotype in place which is not prohibited for the sake of teacher, class or term. Fitness fuction is composed as sum broken restrictions. We use elitary strategy of reproduction and multi-point crossover, normal mutation which theirs values of probability is selected at random.