

Autoreferat rozprawy doktorskiej

SYSTEM KOMPUTEROWY UŁATWIAJĄCY WYKORZYSTANIE INFORMACJI O ZJAWISKACH SOCJALNO-EKONOMICZNYCH PRZY WYBORZE FIRM INWESTUJĄCYCH NA DANYM TERENIE

mgr inż. Janusz Rybarski

PROMOTOR: prof. dr hab. inż. Ryszard Tadeusiewicz – Akademia Górniczo-Hutnicza, WEAIiE

RECENZENCI: prof. dr hab. inż. Ewa Dudek-Dyduch -Akademia Górniczo-Hutnicza, WEAIiE
dr hab. inż., prof. UEK Jacek Wołoszyn – Uniwersytet Ekonomiczny w Krakowie

W rozprawie doktorskiej zaprezentowano projekt systemu komputerowego wykorzystującego informacje socjalno-ekonomiczne do wspomagania procesu wyboru firm inwestujących na danym terenie. Z uwagi na to, iż dotychczas najczęściej stosowane systemy tego rodzaju (systemy wspomagania decyzji) wykorzystują głównie dane liczbowe zgromadzone w bazach danych lub (przy większych systemach) w hurtowniach danych i nie wymagają skomplikowanych procesów kodowania czy też zaawansowanych algorytmów obliczeniowych, autor w niniejszej pracy podjął próbę wykorzystania danych różnego rodzaju (nie tylko numerycznych) służących do opisanie firm, branż oraz środowiska inwestycyjnego. Praca zawiera opis systemu komputerowego stworzonego w oparciu o środowisko Eclipse, opis implementacji poszczególnych rozwiązań i algorytmów, opis zdefiniowanych funkcji obliczających wartości poszczególnych koszyków firm (zestawu wybranych firm), jak również zawiera informacje na temat sposobu kodowania niezbędnych danych. W pracy znajduje się także opis i proponowany sposób rozwiązania problemów pojawiających się w procesie implementacji i testowania poszczególnych rozwiązań i algorytmów. Część empiryczna pracy zawiera opis eksperymentalnej metody weryfikacji działania stworzonego systemu komputerowego poprzez porównanie sugerowanych przez system rozwiązań z decyzjami podejmowanymi (w warunkach symulacji) przez wybraną grupę ludzi.

Cel i zakres pracy

Od momentu, gdy Polska weszła do Unii Europejskiej można zaobserwować rosnące zainteresowanie inwestorów zagranicznych lokowaniem swoich przedsięwzięć na terenie naszego kraju. To, o czym należy jednak pamiętać to ryzyko, które niesie za sobą każda z planowanych inwestycji, bowiem to nie tylko inwestor ponosi ryzyko ulokowania swojego przedsięwzięcia w danym regionie naszego kraju, ale także lokalna społeczność, a co za tym idzie, także i urzędnicy odpowiedzialni za proces podejmowania decyzji. Co więcej, można zauważyć, iż dla urzędników nie tylko realizacja inwestycji ma znaczenie, ale w głównej mierze jej wpływ na możliwości rozwoju lokalnej społeczności, co w dłuższym horyzoncie czasowym owocuje w postaci kolejnych, czasami jeszcze większych inwestycji. Ze względu na dużą liczbę argumentów, parametrów i okoliczności, jakie trzeba wziąć pod uwagę w procesie podejmowania decyzji coraz

trudniej jest decydować o tak rozbudowanych procesach, jakimi są wielomilionowe inwestycje bez odpowiedniego wspomaganie informatycznego.

Analizując powyższe problemy autor niniejszej rozprawy doszedł do wniosku, iż możliwe jest stworzenie komputerowego systemu wykorzystującego informacje socjalno-ekonomiczne, wspomagającego proces podejmowania decyzji. Stwierdzenie to zostało zawarte jako teza omawianej rozprawy.

Omawiając problem wyboru firm inwestujących na danym terenie, autor doszedł do wniosku, iż problem ten można zdefiniować w sposób analogiczny do binarnego problemu plecakowego jako wielowymiarowy binarny problem plecakowy. Binarny problem plecakowy polega na wybraniu z określonej liczby elementów, do plecaka, tych elementów, których sumaryczna wartość jest dla wybierającego jak największa przy ograniczonej pojemności plecaka. W przypadku problemu wyboru firm inwestujących na

danym terenie, wielość terenu przeznaczanego pod inwestycje, zapotrzebowanie na media (prąd, wodę, gaz) czy inne ograniczenia są równoważne wielkości plecaka. Zapotrzebowanie danej firmy na teren inwestycyjny i media odpowiada wielkości (ciężarowi) wybieranego przedmiotu. Wartość, jaką wybrana firma przedstawia dla lokalnej społeczności (wyrażona na przykład w wartości inwestycji bądź ilości zatrudnionych osób) równoważna jest z kolei wartości przedmiotu.

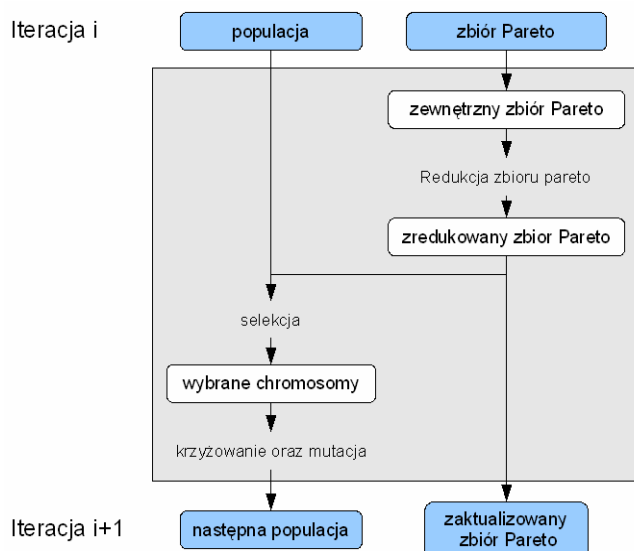
Przedstawiona powyżej krótka charakterystyka obrazuje skalę trudności problemu. W przypadku przeglądania wszystkich możliwych rozwiązań (ang. *Brute Force*) złożoność obliczeniowa przedstawionego problemu wynosi 2^N co w przypadku 20 firm daje aż 1048576 potencjalnych rozwiązań (nie koniecznie spełniających wszystkie ograniczenia).

Istnieje wiele sposobów rozwiązania omawianego problemu. W niniejszej pracy autor wykorzystał algorytm genetyczny oraz metodę optymalizacji w sensie Pareto polegającą na poszukiwaniu rozwiązań Pareto-optymalnych tzw. niezdominowanych (zadanie zostało sprowadzone do znalezienia koszyków firm, dla których nie istnieją rozwiązania lepsze ze względu na wszystkie funkcje, spełniające zadane ograniczenia).

W dostępnej literaturze przedmiotu znajdziemy opis wielu algorytmów genetycznych wykorzystywanych do poszukiwania rozwiązań Pareto-optymalnych (ang. *Multi Object Evoluntary Algorhythm - MOEA*) takie jak: *Schaffer's Vector Evaluated Genetic Algorithm (VEGA)* - Hajela and Lin's *Weighting-based Genetic Algorithm (HLGA)* – Fonseca and Fleming's *Multiobjective Genetic Algorithm (FFGA)* – , Horn, Nafpliotis, and Goldberg's *Niched Pareto Genetic Algorithm (NPGA)* - Srinivas and Deb's *Nondominated Sorting Genetic Algorithm (NSGA)*. W omawianym systemie autor, po wcześniejszym przeanalizowaniu dostępnych wyszczególnionych w pracy pozycji literaturowych porównujących działanie wyżej wymienionych algorytmów przy rozwiązywaniu problemu plecakowego, wykorzystał algorytm o nazwie: *Strange Pareto Evlountary Algorhythm (SPEA)*.

Wybrany algorytm charakteryzuje się następującymi cechami:

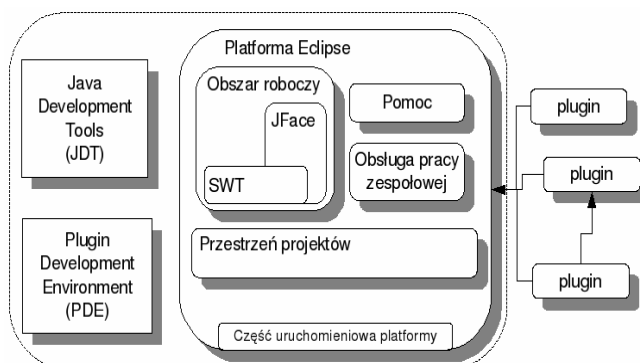
- przechowuje w zewnętrznym zbiorze (ang. *External Set*) chromosomy reprezentujące niezdominowany front Pareto,
- wykorzystuje klasteryzację w celu redukcji ilości rozwiązań,
- wykorzystuje zasadę dominacji Pareto w celu obliczenia skalarnej wartości funkcji przystosowania odpowiadającej każdemu chromosomowi.



Rys. 1: Schemat działania algorytmu SPEA

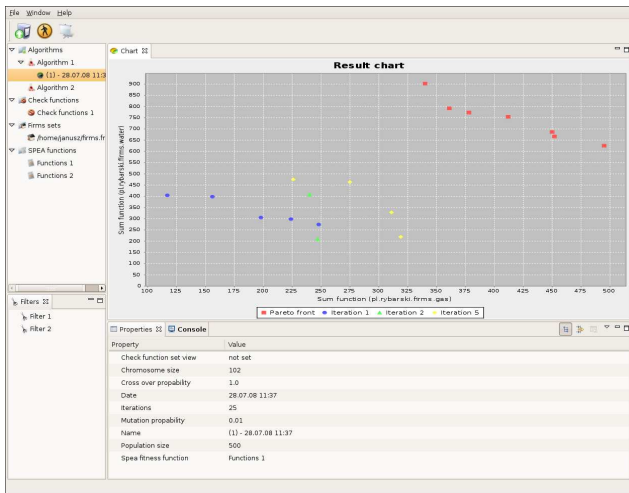
System komputerowy

Ze względu na skomplikowaną budowę tworzonego systemu oraz konieczność zapewnienia modułowości platformy umożliwiającej dalszą jej rozbudowę oraz modyfikację, autor zdecydował się na wykorzystanie otwartej platformy napisanej w języku Java o nazwie Eclipse:



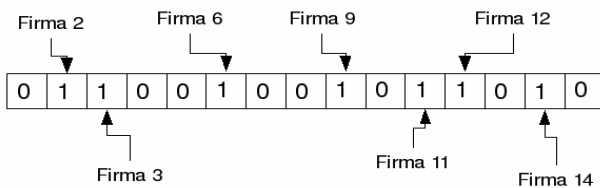
Rys. 2: Schemat platformy Eclipse

Niezależne moduły odpowiedzialne za wykonywanie operacji związanych z działaniem algorytmów genetycznych, optymalizacji Pareto czy obliczania wartości funkcji przystosowania (funkcji celu rozważanego problemu optymalizacji) zostały zaimplementowane jako wtyczki (ang. *plugins*) co umożliwia ich wykorzystanie także w innych projektach opartych o tę samą platformę.



Rys. 3: Interfejs systemu komputerowego

Poszczególne koszyki firm (zbiory firm), które podlegają procesowi decyzyjnemu, zostały zakodowane w postaci binarnej, w którym to kodowaniu, w danym chromosomie poszczególne elementy (geny) odpowiadające firmom przyjmują wartość, 1 która oznacza, iż dana firma została wybrana i może inwestować na danym terenie, bądź 0 co z kolei oznacza, iż dana firma nie będzie miała możliwości inwestycji. Długość całkowita chromosomu odpowiada całkowitej liczbie firm, które są zainteresowane inwestycją na danym terenie.



Rys. 4: Sposób kodowania firm.

Wszystkie firmy biorące udział w procesie podejmowania decyzji zostały przez autora opisane za pomocą następujących parametrów:

- nazwa firmy,
- zgłoszone zapotrzebowanie na grunty,
- zgłoszone zapotrzebowanie na media (prąd, woda, gaz),
- informacje dotyczące wpływu na środowisko (znikomy, mały, średni, duży),
- informacje na temat uciążliwości dla lokalnej społeczności (znikoma, mała, średnia, duża)
- opis produkcji,
- branża w jakiej firma działa,
- rodzaj działalności,
- struktura własności zawierająca kraj pochodzenia firmy wraz z udziałem procentowym,
- planowana wartość inwestycji w PLN,
- planowane zatrudnienie,
- preferowany rodzaj transportu.

Parametry te zostały następnie zakodowane w pliku postaci XML, którego postać gwarantuje możliwość dodawania bądź też usuwania poszczególnych danych opisujących wybrane firmy pozwalając użytkownikowi definiować dane wykorzystywane do oceny wartości poszczególnych firm.

```
<firm id="0">
  <name>Cebal Tuba Sp. z o.o.</name>
  <data>
    <parameter id="pl.rybarski.firms.size">11582.0</parameter>
    <parameter id="pl.rybarski.firms.power">21.0</parameter>
    <parameter id="pl.rybarski.firms.gas">32.0</parameter>
    <parameter id="pl.rybarski.firms.water">21.0</parameter>
    <parameter id="pl.rybarski.firms.production">tworzywa sztuczne
  - opakowania polietylenowe</parameter>
    <parameter id="pl.rybarski.firms.trade">chemiczna</parameter>
    <parameter
  id="pl.rybarski.firms.investment">1.12E7</parameter>
    <parameter id="pl.rybarski.firms.workers">45</parameter>
    <parameter id="pl.rybarski.firms.enviroment">
      <en id="enviroment" value="średnia" />
      <en id="persons" value="średnia" />
    </parameter>
    <parameter id="pl.rybarski.firms.contribution">
      <con country="Francja" value="58.0" />
      <con country="Polska" value="42.0" />
    </parameter>
    <parameter id="pl.rybarski.firms.productionType">
      <prod>produkcyjna</prod>
    </parameter>
    <parameter id="pl.rybarski.firms.transport">
      <trans>samochodowy</trans>
    </parameter>
  </data>
</firm>
```

Listing 1. Struktura pliku zawierającego informacje o firmach

Teren przeznaczony pod inwestycję został opisany za pomocą następujących danych:

- wielkości terenu inwestycyjnego,
- dostępności mediów wraz z informacją o kosztach związanych ze zwiększeniem ich wydajności (o ile taka operacja jest możliwa) zgłoszone zapotrzebowanie na media (prąd, woda, gaz),
- informacji o wydajności oczyszczalni ścieków,
- rodzaju dostępnego transportu,
- informacji na temat uciążliwości dla lokalnej społeczności (znikoma, mała, średnia, duża),
- informacji na temat infrastruktury telekomunikacyjnej znajdującej się na terenie przeznaczonym pod inwestycję.

Aby możliwa była ocena wartości poszczególnych koszyków firm oraz ich wpływu na rozwój lokalnej społeczności autor zdefiniował funkcje oceny wymienione poniżej:

funkcja sumująca - umożliwiająca obliczenie sumy dowolnie wybranego parametru firmy (wielkość zajętej

powierzchni inwestycyjnej, wartości zapotrzebowania na media, sumaryczna wartość inwestycji itp.)

$$f(ch) = \sum_{i=1}^n c_i x_i \quad (1)$$

Gdzie:

ch – chromosom

c_i – wartość parametru i -tej firmy

x_i – wartość genu oznaczająca i -tą firmę (0 lub 1)

Funkcja sumująca z ograniczeniami - odmianą zdefiniowanej powyżej funkcji sumującej jest funkcja sumująca z ograniczeniami. Umożliwia ona narzucenie nieliniowości w procesie sumowania wynikającej na przykład ze znacznego wzrostu kosztu inwestycji w przypadku przekroczenia granicznej wydajności określonych mediów (np.: konieczność zbudowania dodawanej oczyszczalni ścieków w przypadku przekroczenia ilości odpadów, którą jest w stanie przyjąć działająca oczyszczalnia ścieków)

$$f(ch) = \begin{cases} \sum_{i=0}^n c_i x_i & \text{dla } f(ch) \leq \delta \\ g(ch) & \text{dla } f(ch) > \delta \end{cases} \quad (2)$$

gdzie:

$g(ch)$ – funkcja obliczająca wartość chromosomu

Funkcja obliczająca rozkład branż w danym koszyku. Jednym z najważniejszych problemów, który należy wziąć pod uwagę podczas wyboru firm inwestujących na danym terenie jest taki dobór firm oraz branż tak, aby w przypadku problemów ekonomicznych jednej z branż, do minimum ograniczyć skutki powstałego kryzysu (na przykład grupowych zwolnień pracowników). Zaproponowana przez autora funkcja oblicza wartość odchylenia standardowego próby zawierającej ilość firm danej branży inwestującej na określonym terenie. Znak minus został dodany w celu minimalizacji wartości odchylenia (maksymalizacji współczynnika)

$$f(ch) = -\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (3)$$

Przy czym:

ch - chromosom

n – ilość branż

x_i – liczba firm z danej branży

\bar{x} - średnia ilość firm w branżach

Funkcja obliczająca potencjalną ilość osób zatrudnionych. Przy tworzeniu koszyka firm, które zainwestują na danym terenie, należy pamiętać, iż każda branża charakteryzuje się pewną specyfiką produkcji, która to z kolei wymusza takie bądź inne wykształcenie zatrudnianych osób. W celu obliczenia tak zdefiniowanego przyszłego zatrudnienia w danym regionie, autor zaproponował funkcję pozwalającą wykorzystać informacje na temat bezrobocia (jego wielkości i profilu) oraz zapotrzebowania firm na osoby o danym wykształceniu. Do obliczenia wartości funkcji wykorzystywana jest tabela określająca procentowe zapotrzebowanie branż na pracowników o danym wykształceniu. Funkcja została zdefiniowana w sposób pokazany poniżej

$$f(ch) = \sum_{i \in I} p_i \sum_{g \in G} z_g w_{ig} \quad (4)$$

Przy czym dla pokazanej funkcji zdefiniowano zbiór I będący zbiorem rodzajów wykształcenia (*podstawowe, zasadnicze, średnie, wyższe*), a także określono p_i , który jest procentową wartością określającą ilość bezrobotnych z danym wykształceniem na terenie gminy. Dla każdej firmy g zdefiniowano zmienną z_g jako wartość określającą zapotrzebowanie na określoną liczbę pracowników oraz zmienną w_{ig} oznaczającą współczynnik procentowego zapotrzebowania na pracowników o danym wykształceniu i .

Funkcja obliczająca wartość powiązań pomiędzy branżami. Ostatnią ze stworzonych funkcji jest funkcja pozwalająca ocenić sumaryczną wartość powiązań pomiędzy firmami na podstawie tabeli powiązań między branżami. Tworząc koszyk firm można, bowiem tak dobrać poszczególne firmy, aby współpracując ze sobą tworzyły coś na kształt klastra¹. Funkcja ta oparta o funkcję sumującą wykorzystuje tabelę powiązań pomiędzy branżami.

Weryfikacja poprawności implementacji funkcji:

Celem weryfikacji poprawności implementacji algorytmu genetycznego, algorytmu SPEA oraz algorytmu znajdowania rozwiązania Pareto-optimalnych zostały przeprowadzone testy mające na celu porównanie otrzymanego za pomocą metody Brute Force rozwiązania z rozwiązaniami otrzymanymi z wykorzystaniem algorytmów genetycznych. Ze względu na złożoność obliczeniową problemu plecakowego przy rozwiązywaniu metodą *brute force* autor ograniczył wielkość chromosomu (ilość wybieranych firm) oraz wielkości populacji początkowej. Została także ograniczona ilość funkcji

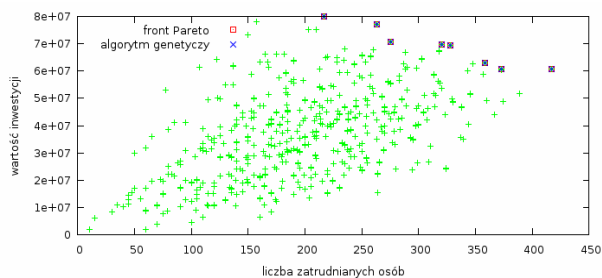
¹ Klaster- znajdująca się w geograficznym sąsiedztwie grupa przedsiębiorstw i powiązanych z nimi instytucji zajmujących się określoną dziedziną, połączoną podobieństwami i wzajemnie się uzupełniającą – źródło: www.klaster.pl

oceniających koszyki firm do funkcji obliczającej sumaryczną wartość inwestycji oraz funkcji obliczającej sumaryczną ilość zatrudnionych osób.

W tabeli poniżej zaprezentowano ustawienia algorytmu genetycznego.

Tabela 1. Parametry działania algorytmu genetycznego

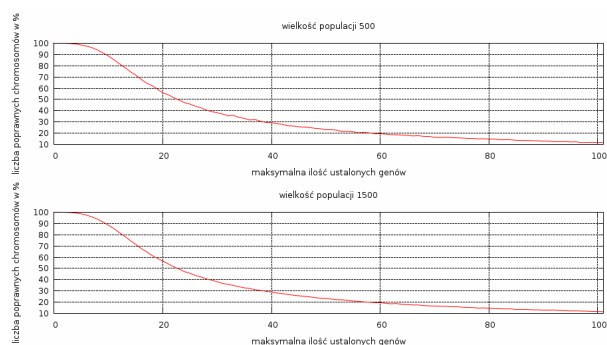
Wielkość chromosomu	20
Wielkość populacji	200
Prawdopodobieństwo krzyżowania	0,8
Prawdopodobieństwo mutacji	0,1
Ilość iteracji	300



Rys. 5: Porównanie wyników otrzymanych za pomocą algorytmu genetycznego z wynikami otrzymanymi metodą Brute Force.

Problematyka tworzenia populacji początkowej oraz problematyka rozwiązań dopuszczalnych

W rozdziale 3.6 autor przedstawił problematykę tworzenia populacji początkowej w zależności od ilości ustalonych genów w chromosomie (ilości wybranych firm w koszyku) oraz zaproponował algorytm umożliwiający określenie maksymalnej ilości ustalonych genów gwarantujących otrzymanie populacji o założonej procentowo liczbie chromosomów poprawnych.



Rys. 6: Procentowa liczba rozwiązań dopuszczalnych w zależności od ilości ustalonych genów dla populacji wielkości 500 i 1500 chromosomów.

W rozdziale 3.7 autor opisał problem otrzymywania rozwiązań znajdujących się poza przestrzenią rozwiązań dopuszczalnych. W rozdziale tym zostały omówione dwie metody rozwiązania zaistniałego problemu oraz opisany został algorytm umożliwiający naprawienie chromosomu tak, aby otrzymany chromosom spełniał wszystkie zadane

ograniczenia. Działanie zaproponowanego algorytmu polega na losowym kolejnym usuwaniu firm, które zostały wybrane i mogą inwestować na danym terenie do momentu, w którym koszyk (chromosom) spełnia wszystkie nałożone ograniczenia.

Weryfikacja działania komputerowego systemu wspomagania decyzji.

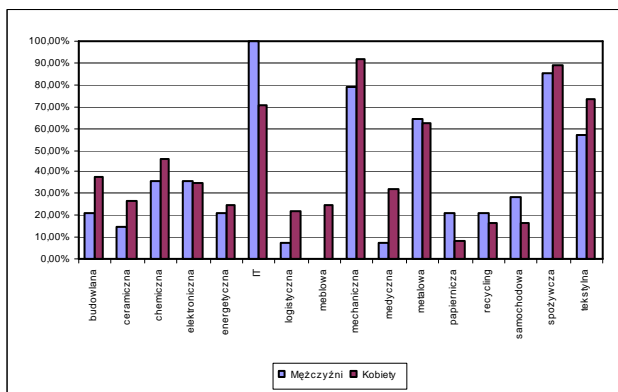
W celu przeprowadzenia weryfikacji działania systemu komputerowego wspomagania decyzji autor poprosił grupę składającą się z 37 studentów kierunku Gospodarka i Administracja Publiczna Uniwersytetu Ekonomicznego w Krakowie o podjęcie i uzasadnienie decyzji wyboru firm inwestujących na danym terenie. W pracy autor umieścił pełną analizę podjętych przez ludzi decyzji. Poniżej zaprezentowano ogólną charakterystykę rozwiązań otrzymanych przez mężczyzn oraz kobiety.

Tabela 2. Krótka charakterystyka koszyków firm otrzymanych przez mężczyzn

	Min	Max	Średnia
Ilość firm	7,00	17,00	11,43
Ilość branż	2,00	10,00	6,00
Zajęta powierzchnia	272 722,0	619 853,0	578 785,0
Prąd – zużycie	183,00	461,00	291,29
Gaz – zużycie	0,00	147,00	78,43
Woda – zużycie	29,00	373,00	173,07
Wartość inwestycji	28750000,0	1375735100,0	696524387,79
Wielkość zatrudnienia	615,00	2019,00	1222,00

Tabela 3. Krótka charakterystyka koszyków firm otrzymanych przez kobiety

	Min	Max	Średnia
Ilość firm	4,00	30,00	11,27
Ilość branż	3,00	12,00	6,76
Zajęta powierzchnia	272 722,00	619926,00	589852,99
Prąd – zużycie	113,00	614,00	235,65
Gaz – zużycie	0,00	189,00	77,24
Woda – zużycie	61,00	380,00	179,65
Wartość inwestycji	170580100,0	1282267022,0	669037353,76
Wielkość zatrudnienia	763,00	1846,00	1201,32



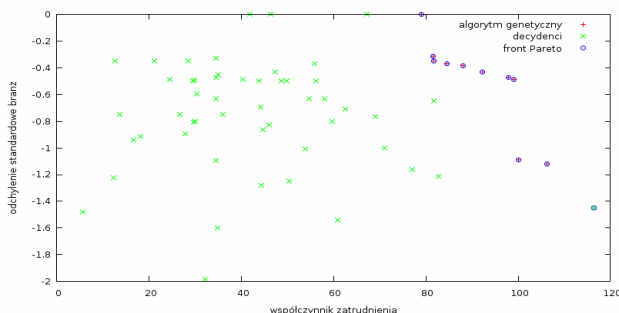
Rys. 7: Procentowy udział branż w poszczególnych koszykach – decyzje podejmowane przez ludzi.

Otrzymane wyniki zostały następnie porównane z wynikami uzyskanymi w wyniku działania stworzonego systemu komputerowego. Symulacje zostały przeprowadzone dla różnych zestawów funkcji celu jak również przy różnych zdanych ograniczeniach. W tabeli 4 znajdują się podstawowe parametry działania algorytmu genetycznego.

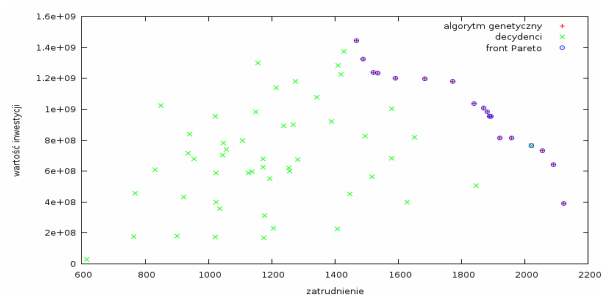
Tabela 4. Parametry działania algorytmu genetycznego

Wielkość chromosomu	102
Wielkość populacji początkowej	1200
Prawdopodobieństwo krzyżowania	0,8
Prawdopodobieństwo mutacji	0,1
Ilość iteracji	700

Otrzymane wyniki zostały przedstawione w postaci dwuwymiarowych wykresów porównujących rezultaty symulacji komputerowych z decyzjami podejmowanymi przez ludzi.



Rys. 8: Porównanie wyników symulacji z decyzjami decydentów dla 2 funkcji obliczających: odchylenie standardowe branż oraz współczynnik zatrudnienia



Rys. 9: Porównanie wyników symulacji z decyzjami decydentów dla 2 funkcji odliczających: sumaryczną wartość inwestycji oraz sumę osób zatrudnionych

Wyniki symulacji komputerowych potwierdziły postawioną na początku rozprawy tezę, iż możliwe jest stworzenie systemu komputerowego wspomagającego proces decyzyjny wykorzystującego informacje socjalno-ekonomiczne przy wyborze firm inwestujących na danym terenie. Co więcej, udało się wykazać iż system ten jest w stanie podejmować decyzje równie dobre jak decyzje podejmowane przez człowieka, gwarantując bezstronność w podejmowaniu decyzji oraz odporność na czynniki typowo ludzkie związane z procesem podejmowania decyzji (stres, zmęczenie itp.).

Publikacje doktoranta:

1. J. Rybarski, Eclipse Graphical Editing Framework - graficzny edytor danych na przykładzie edytora diagramu kształtów (Shape Diagram Editor), Software Developer's Journal nr 6(149), Czerwiec 2007, s. 40-45
2. J. Rybarski, J. Plesnar, Pojedynek gigantów - NetBeans vs Eclipse, Software Developer's Journal nr 2(146), Luty 2007, s. 24-28
3. J. Rybarski, JasperReports oraz iReports - raportowanie z baz danych, Software Developer's Journal nr 11(143), Listopad 2006, s. 52-57
4. J. Rybarski, R. Tadeusiewicz, Problem optymalnego doboru firm inwestujących w danym regionie jako informatyczny binarny problem plecakowy, Informatyka Teoretyczna i Stosowana, vol. 5, nr 9, 2006, s. 187-193
5. S. Habdank-Wojewódzki, J. Rybarski, The Kohonen Neural Network, Overload 74, Aug. 2006, ISSN 1354-3172.