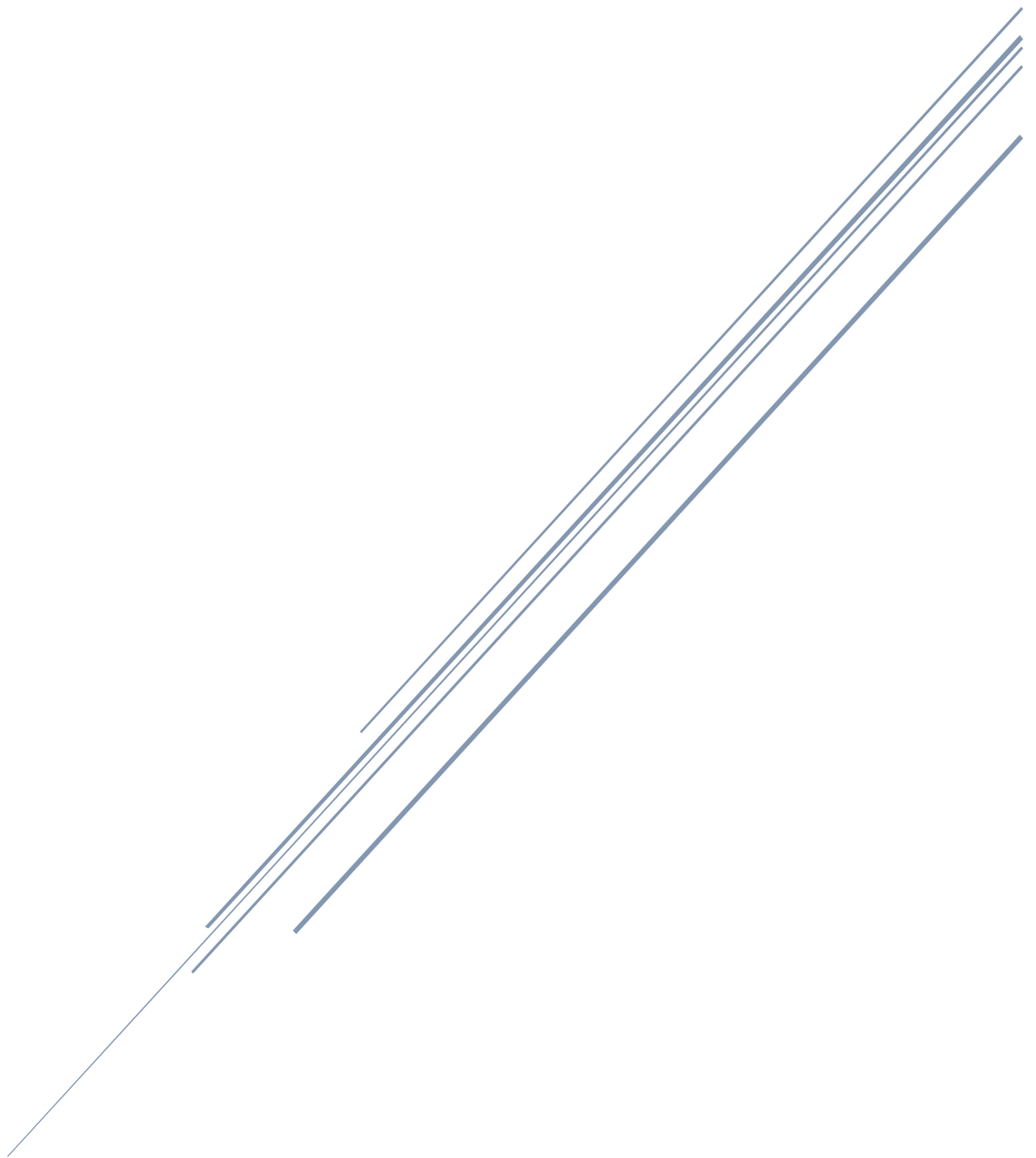


SZTUCZNA INTELIGENCJA

Algorytmy genetyczne



20.03.2018

Michał Kanak 206906

Spis treści

Wprowadzenie	2
Algorytm genetyczny – definicja.....	2
Metody selekcji.....	2
Operatory kodowania	2
Krzyżowanie.....	2
Mutacja	2
Typowe działanie algorytmu	2
Zastosowanie.....	3
Badania.....	3
Zachowanie stworzonego algorytmu, także w odniesieniu do ilości danych	3
Badanie.....	3
Wnioski.....	6
Opis zastosowanych metod selekcji oraz wpływ na wyniki	6
Metoda Ruletki	6
Badanie.....	6
Metoda Turniejowa.....	6
Badanie.....	7
Wnioski dla metod selekcji	7
Optymalna wielkość populacji oraz generacji dla wybranych danych	7
Badanie.....	7
Wnioski.....	9
Wpływ parametru prawdopodobieństwa mutacji.....	10
Badanie.....	10
Wnioski.....	11
Wpływ parametru prawdopodobieństwa krzyżowania	11
Badanie.....	11
Wnioski.....	12
Podsumowanie	12
Literatura.....	13

Wprowadzenie

Algorytm genetyczny – definicja

Swoją nazwę zawdzięcza podobieństwo do ewolucji biologicznej, ponieważ jego autor właśnie z obserwacji biologii zaczerpnął inspiracje do swojej pracy. Algorytm jest nazywany ewolucyjnym. Algorytm rozwiązuje problem, który jest definiowany jako środowisko w którym istnieje pewna populacja osobników. Każdy osobnik zawiera swój genotyp, zbiór chromosomów odpowiadających za cechy problemu i opisuje go.

Metody selekcji

Rozważymy dwie metody selekcji:

1. Ruletka – Polega na zbudowaniu wirtualnego koła, którego wycinki odpowiadają osobnikom, z tym że, im lepszy osobnik, tym proporcjonalnie więcej miejsca zajmuje na kole, metoda da słabiej rozpoznaje osobniki „dobre”.
2. Ranking – Metoda ta polega na ocenie każdego z osobników za pomocą funkcji oceny, sortujemy osobniki wg. oceny, ustawiamy w szeregu. Najlepsze osobniki mają możliwość rozmnożenia, a ostatnie są usuwane z populacji. Metoda selekcji lepiej sprawdza się pod względem przystosowania osobników, lecz daje większe prawdopodobieństwo, że zmiany w genotypach nie będą znaczące.

Operatory kodowania

Krzyżowanie

Operator krzyżowania polega na łączeniu cech w różnorodnych kombinacjach, zazwyczaj wybieranych losowo, może być realizowane przez rozcięcie genotypów u rodziców w tym samym miejscu oraz połączenie lewej części „matki” z prawą częścią „ojca” i na odwrót. Stosuje się też metody liczenia średniej z genotypu obydwu rodziców lub wykorzystuje operacje logiczne.

Mutacja

Polega na wprowadzaniu losowych zmian w wybranych (przez współczynnik prawdopodobieństwa) osobników, odpowiada za wprowadzenia różnorodności w populacjach i zapobieganie stagnacji, może ułatwić odnalezienie optymalnego rozwiązania wyłamując schemat działania. Prawdopodobieństwo mutacji nie powinno być większe niż 1%, w przeciwnym razie może przynieść odwrotny skutek do zamierzonego.

Typowe działanie algorytmu

1. Losujemy pewną populację początkową
2. Populacja jest poddawana selekcji, która definiuje osobniki wybrane do ewentualnej reprodukcji.
3. Wybrane osobniki są poddawane z pewnym prawdopodobieństwem operatorom ewolucyjnym:
 - a. Krzyżowaniu – łączenie genotypów wybranych par
 - b. Mutacja – losowe zmiany w genotypie
4. Nowo utworzone pokolenie dołączają do populacji, najgorsze są zastępowane. Po określonej ilości kroków lub po uzyskaniu zadowalającego wyniku algorytm przerywa działanie w przeciwnym wypadku wraca do punktu 2.

Zastosowanie

Algorytmy genetyczne stosują się w celu przeszukiwania dużych przestrzeni rozwiązań, gdzie zależy nam na czasie w jakim znajdziemy rozwiązanie bliskie optymalnemu, niekoniecznie optymalne. A także w problemach NP-trudnych, poprzez swoje nieschematyczne działanie (przy odpowiednio dobranych parametrach) rozwiązuje problemy szybciej niż zwykłe algorytmy, a także tam gdzie algorytmy nie zostały stworzone.

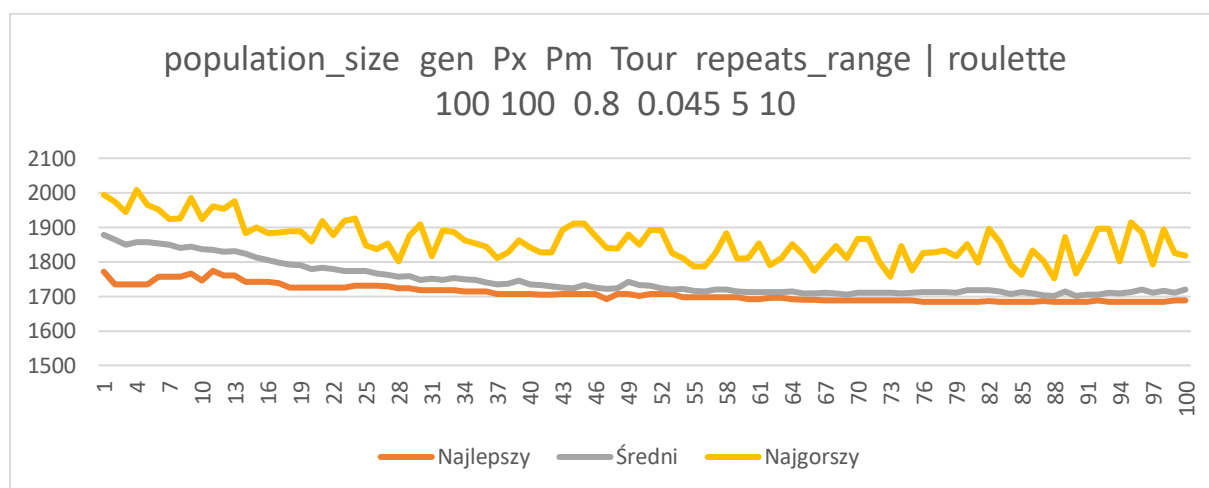
Badania

Zachowanie stworzonego algorytmu, także w odniesieniu do ilości danych

Algorytm jest bardzo stabilny, szybki, został sprawdzony przez dużą ilość danych oraz liczbę uruchomień. Wszystkie badania zostały przeprowadzone po wielokrotnym sprawdzeniu działania algorytmu. **Odchylenie standardowe** oraz **optimum** pochodzą z podwójnej pętli algorytmu, tak by zostały wygenerowane populacje z 10 osobnych unikalnie wylosowanych populacji początkowych. (np. populacje z 10 różnych planet). Wykresy zostały stworzone z populacji pierwszej planety. Do sprawozdania zostało wybrane kilka z wielu tablic danych. Są one dołączone jako załącznik nr.1.

Badanie

HAD12



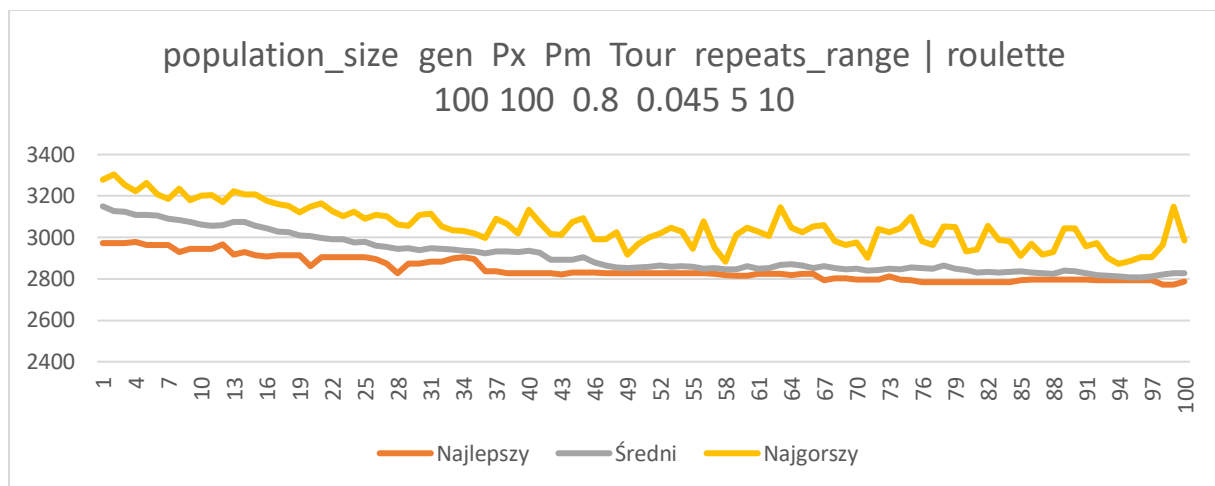
Wyniki dla wybranych parametrów to:

Odchylenie	29,269121	Optimum	1716
------------	-----------	---------	------

Średnie globalne:

Najlepszy	Średni	Najgorszy
1703.2	1710.8	1777.6

HAD14



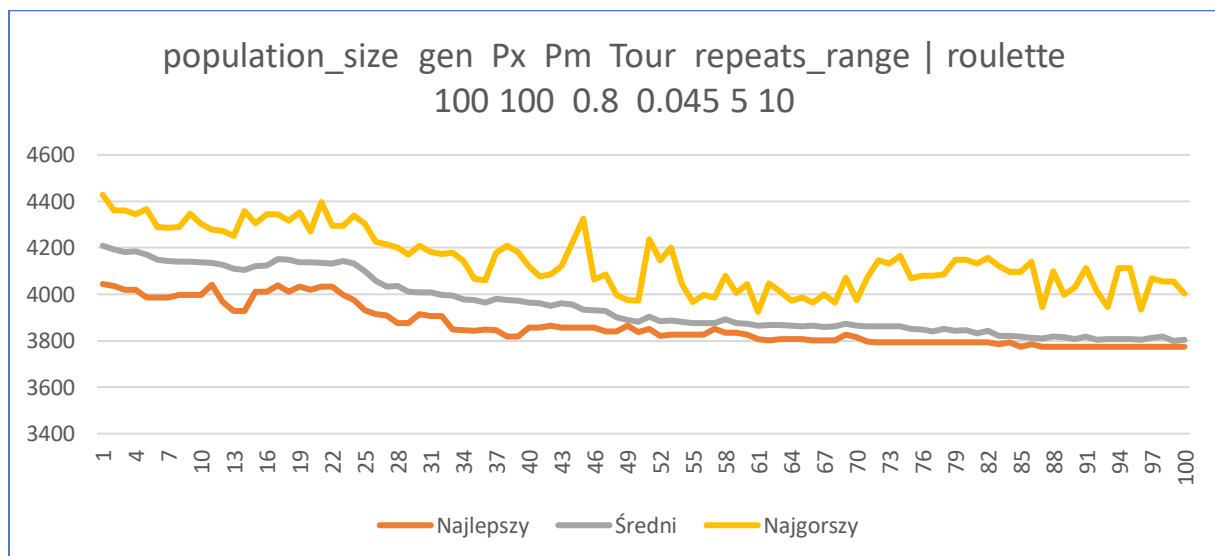
Wyniki dla wybranych parametrów to:

Odchylenie	53,3245907	Optimum	2788
------------	------------	---------	------

Średnie globalne:

Najlepszy	Średni	Najgorszy
2788.0	2819.2	2972.6

HAD16



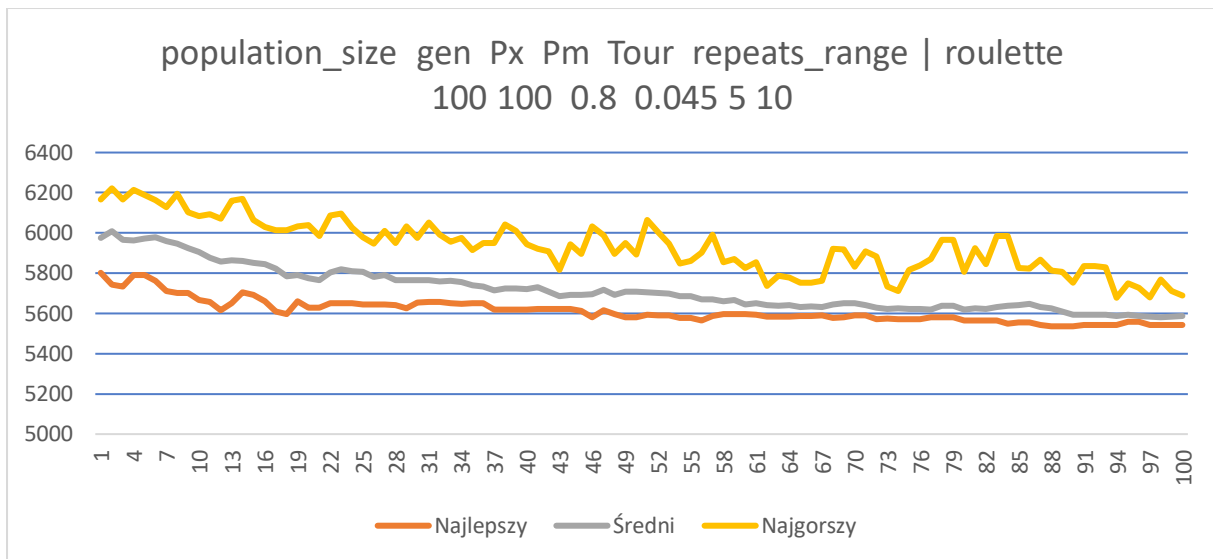
Wyniki dla wybranych parametrów to:

Odchylenie	66,(6)	Optimum	3852
------------	--------	---------	------

Średnie globalne:

Najlepszy	Średni	Najgorszy
3842.0	3880.0	4060.6

HAD 18



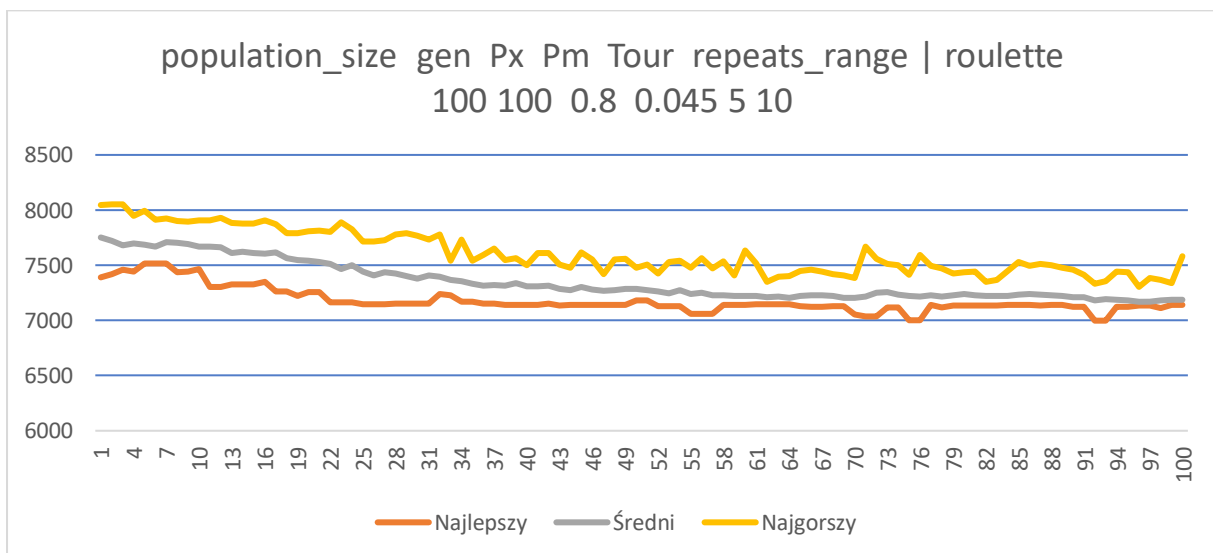
Wyniki dla wybranych parametrów to:

Odchylenie	73,03983	Optimum	5454
------------	----------	---------	------

Średnie globalne:

Najlepszy	Średni	Najgorszy
5516.8	5581.6	5799.0

HAD20



Wyniki dla wybranych parametrów to:

Odchylenie	110,9324	Optimum	7086
------------	----------	---------	------

Średnie globalne:

Najlepszy	Średni	Najgorszy
7156.0	7230.1	7571.8

Wnioski

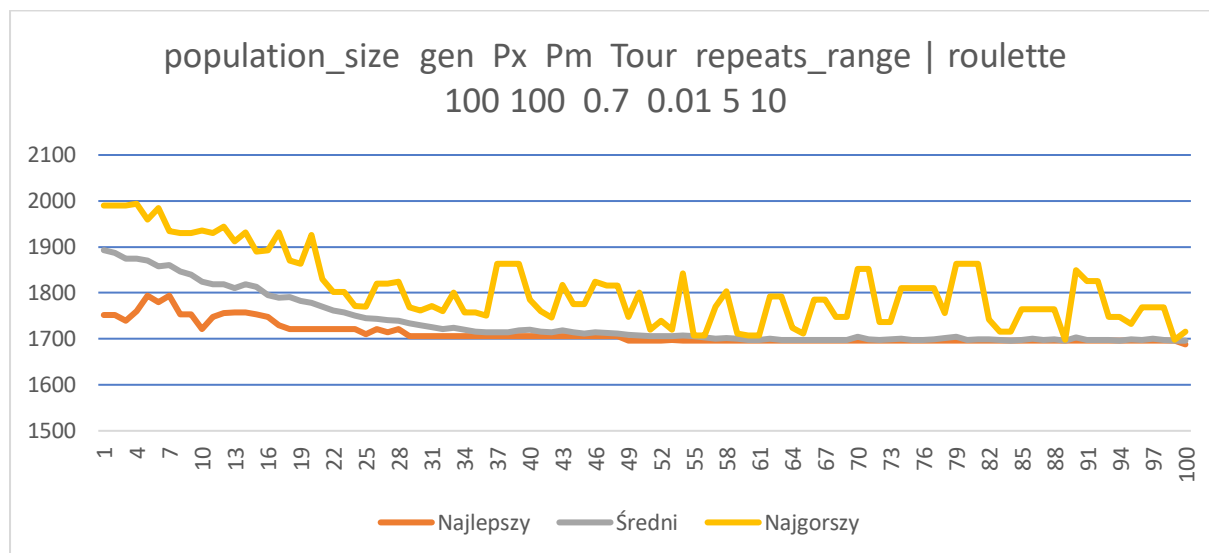
Z wyżej załączonych wykresów wnioskuję, że algorytm działa sprawnie, zachowując swoje właściwości niezależnie od wielkości macierzy danych. Czas wykonywania rośnie w rozsądnym tempie. Prawdopodobnie dzięki zastosowanemu trickowi dotyczącemu zapisywania wyniku funkcji kosztu przy danym genotypie. Przez co musi być liczony tylko raz na każdą modyfikację genotypu. Kolejne populacje dążą do optymalnego wyniku, bez znaczących niepożądanych odchyień.

Opis zastosowanych metod selekcji oraz wpływ na wyniki

Metoda Ruletki

Metoda polega na zbudowaniu wirtualnego koła (można ją też przedstawić na wykresie), poszczególne wycinki koła odpowiadają konkretnym osobnikom. Im bardziej obiecujący osobnik (lepiej przystosowany), tym większy obszar koła zajmuje. Im większy obszar tym większe prawdopodobieństwo udziału w krzyżowaniu. Niestety ewolucja tej metody selekcji spada z pokolenia na pokolenie. Może być także nieskuteczny w przypadku osobników bardzo do siebie podobnych, w tej kwestii lepiej sprawdza się metoda turniejowa.

Badanie



Wyniki dla wybranych parametrów to:

Odchylenie	19.691	Optimum	1716
------------	--------	---------	------

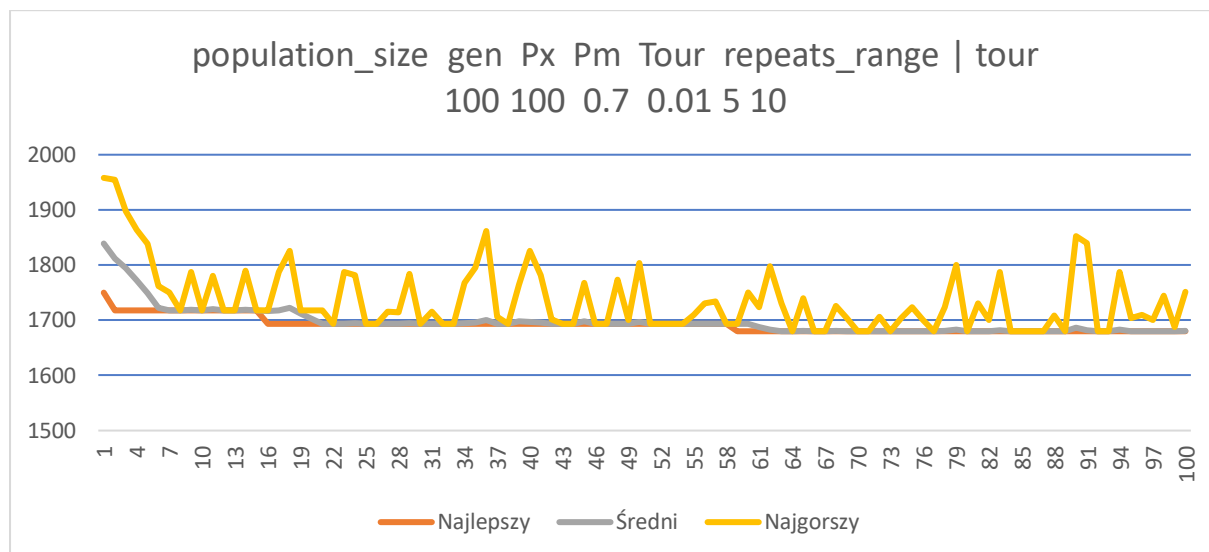
Średnie globalne:

Najlepszy	Średni	Najgorszy
1703.2	1710.8	1777.6

Metoda Turniejowa

Metoda turniejowa przez lepsze radzenie sobie z podobnymi do siebie osobnikami lepiej sprawdza się w badanym przypadku, co też widać na poniższym wykresie. Polega ona na określeniu oceny na podstawie odpowiednio zaprojektowanej funkcji, później osobniki są sortowane od najlepszego do najgorszego, pierwsi z nich mają możliwość krzyżowania się ze sobą. W tej metodzie wadą może być niewrażliwość na różnice między sąsiednimi genotypami.

Badanie



Wyniki dla wybranych parametrów to:

Odchylenie	26.271	Optimum	1680
------------	--------	---------	------

Średnie globalne:

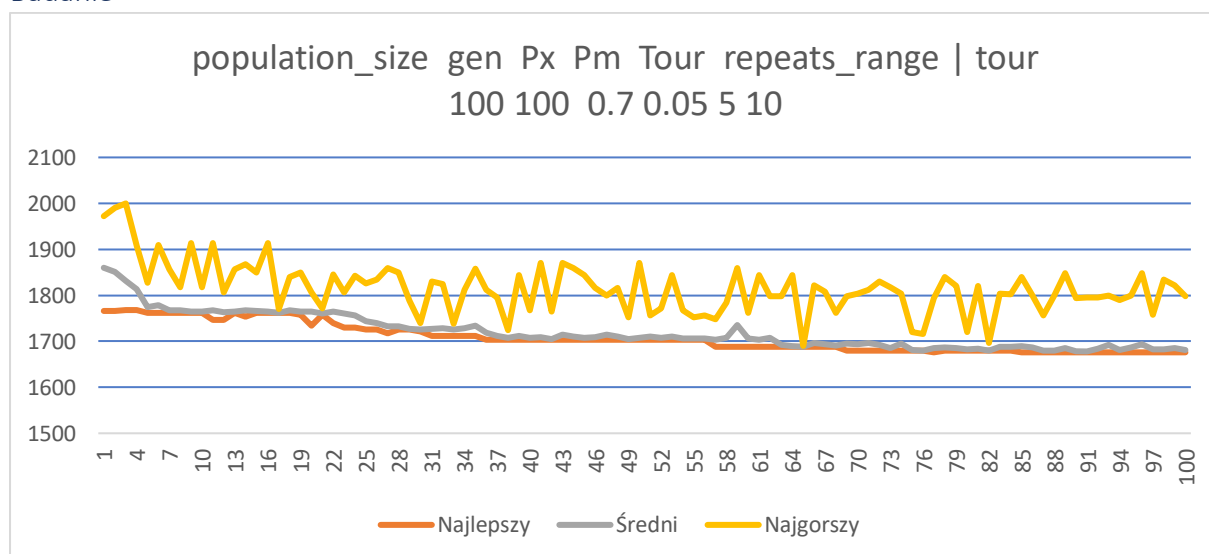
Najlepszy	Średni	Najgorszy
1678.4	1678.7	1704.0

Wnioski dla metod selekcji

Metoda turniejowa szybciej oraz gwałtowniej zbiega się bliżej określonego optimum, w tym konkretnym rozważanym przypadku o wiele lepiej się sprawdza, pomimo większego odchylenia standardowego. Turniej potrzebuje także mniejszej ilości populacji do uzyskania zadowalającego wyniku, co może być kluczowe dla rozwiązania problemu.

Optymalna wielkość populacji oraz generacji dla wybranych danych

Badanie

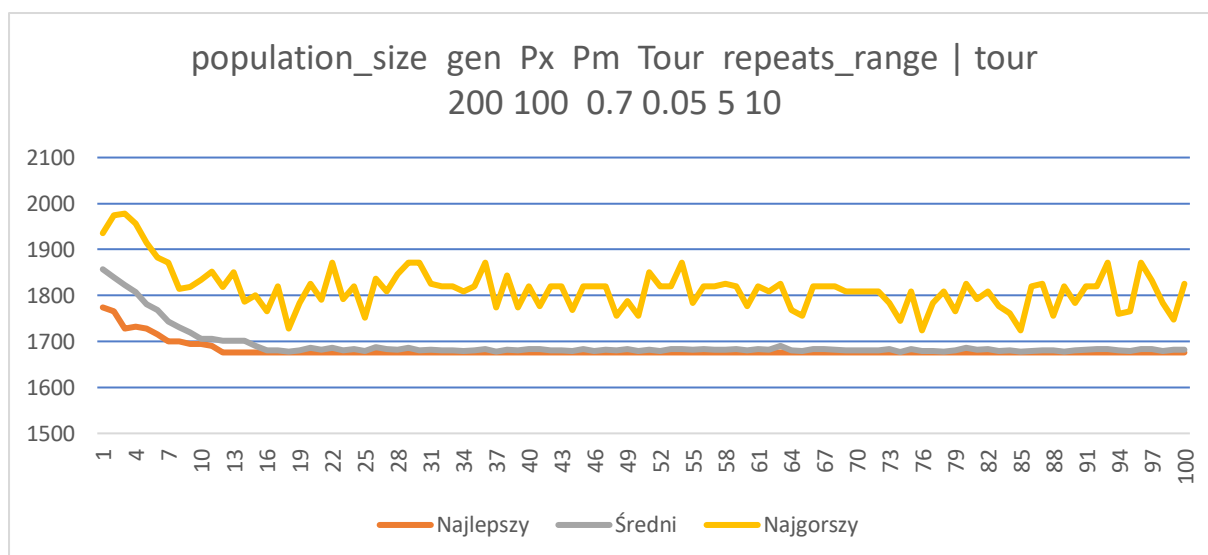


Wyniki dla wybranych parametrów to:

Odchylenie	26.37282	Optimum	1682
------------	----------	---------	------

Średnie globalne:

Najlepszy	Średni	Najgorszy
1672.0	1678.9	1809.6

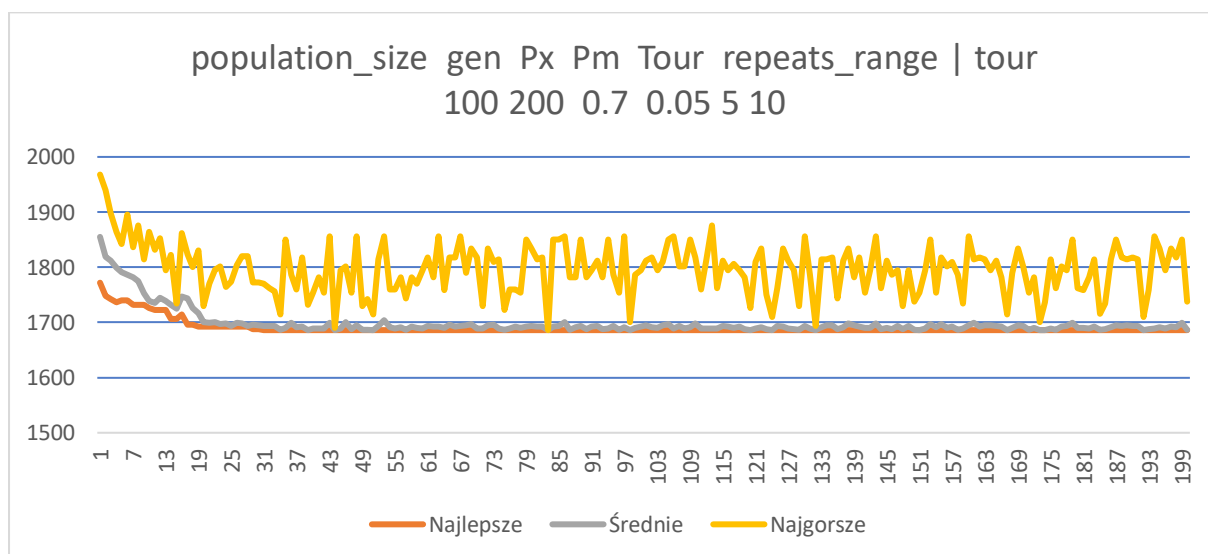


Wyniki dla wybranych parametrów to:

Odchylenie	21.5283	Optimum	1680
------------	---------	---------	------

Średnie globalne:

Najlepszy	Średni	Najgorszy
1666.6	1671.9	1800.6

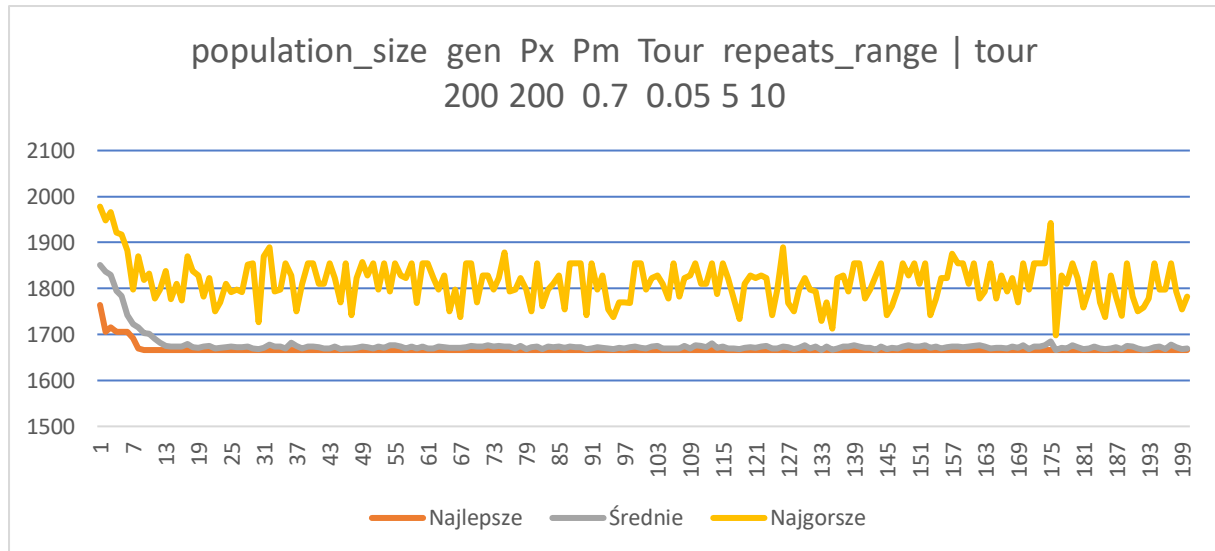


Wyniki dla wybranych parametrów to:

Odchylenie	21.63703	Optimum	1662
------------	----------	---------	------

Średnie globalne:

Najlepszy	Średni	Najgorszy
1670	1677.7	1807



Wyniki dla wybranych parametrów to:

Odchylenie	14.4374	Optimum	1678
------------	---------	---------	------

Średnie globalne:

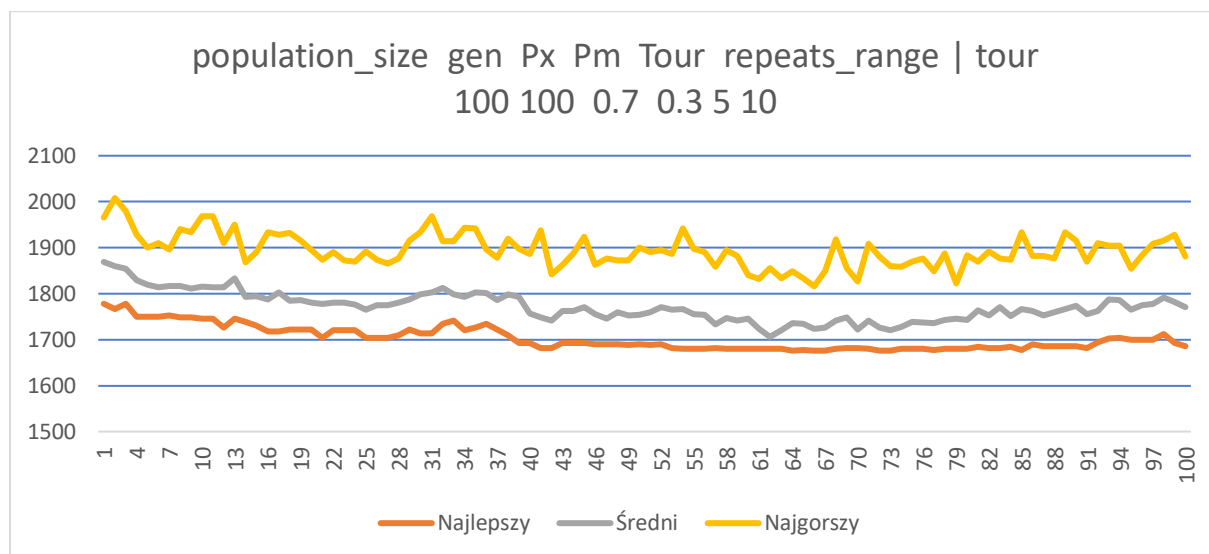
Najlepszy	Średni	Najgorszy
1670.8	1677.1	1804.8

Wnioski

Wyniki są zbliżone do siebie, ponieważ algorytm zaczyna ustawiać populację w niedalekiej granicy optimum bardzo szybko, wystarczy około 10-20 generacji by zaobserwować to zjawisko, zmiana odchylenia na mniejszą wraz ze zwiększeniem ilości generacji oraz ilości osobników w populacji tylko potwierdza to stwierdzenie. Wyniki średnie i najlepsze są bardzo podobne i bardzo szybko zbiegają się ku optimum, działanie jest stabilne i daje wynik najbliższy optymalnemu, już w około 10-20 pokoleniu osiągamy cel algorytmu.

Wpływ parametru prawdopodobieństwa mutacji

Badanie

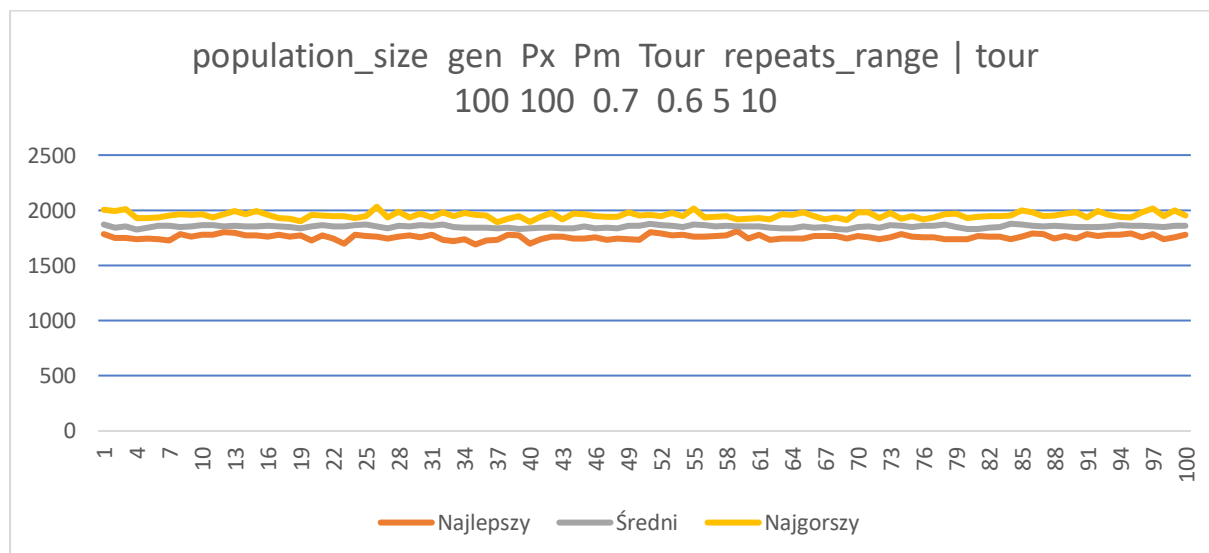


Wyniki dla wybranych parametrów to:

Odchylenie	29.18	Optimum	1712
------------	-------	---------	------

Średnie globalne:

Najlepszy	Średni	Najgorszy
1685.2	1761.6	1890.4



Wyniki dla wybranych parametrów to:

Odchylenie	22.81	Optimum	1770
------------	-------	---------	------

Średnie globalne:

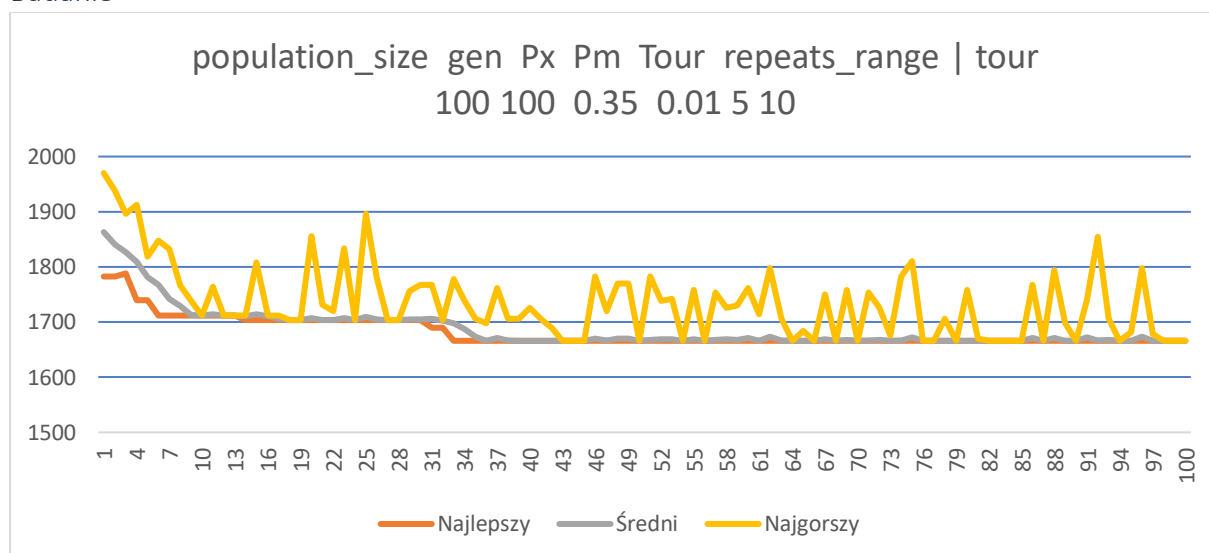
Najlepszy	Średni	Najgorszy
1752.2	1849.2	1966.2

Wnioski

Gdy prawdopodobieństwo mutacji jest zbyt duże algorytm, może być daleki od optymalnego rozwiązania, gdy będzie zbyt małe będzie potrzebne o wiele więcej generacji by algorytm dotarł do swojego optimum. Dlatego należy określić taki punkt, by mutacje występowały, lecz nie na tyle często by wprowadziły duży stopień losowości. Dobrym rozwiązaniem w tym przypadku jest ustalenie P_m na około 0.15

Wpływ parametru prawdopodobieństwa krzyżowania

Badanie

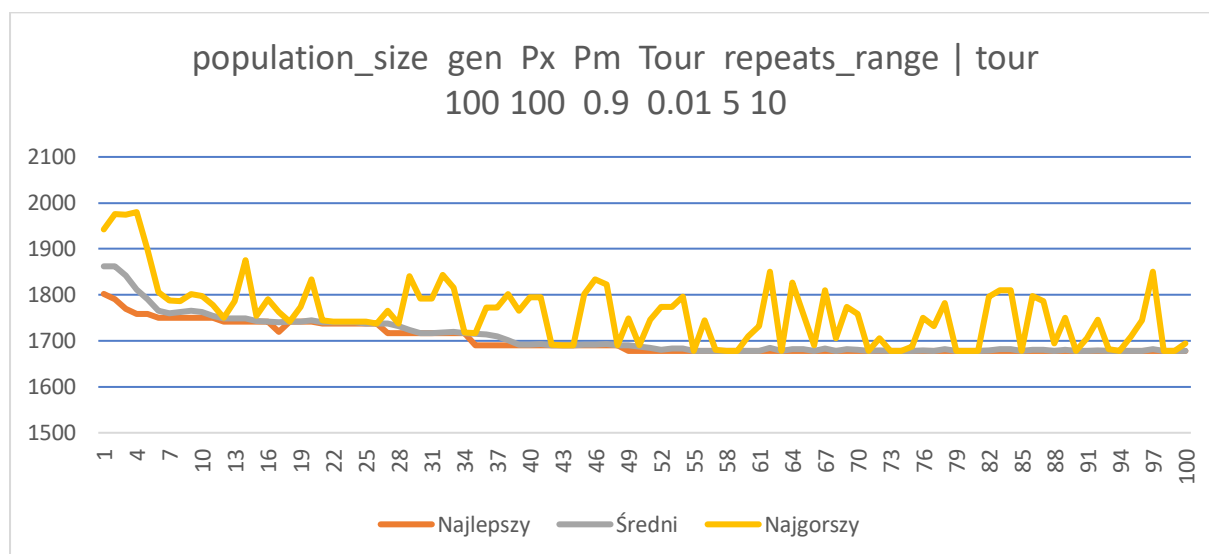


Wyniki dla wybranych parametrów to:

Odchylenie	25.277	Optimum	1690
------------	--------	---------	------

Średnie globalne:

Najlepszy	Średni	Najgorszy
1684.0	1684.4	1711.6



Wyniki dla wybranych parametrów to:

Odchylenie	23.35	Optimum	1720
------------	-------	---------	------

Średnie globalne:

Najlepszy	Średni	Najgorszy
1691.2	1691.9	1726.0

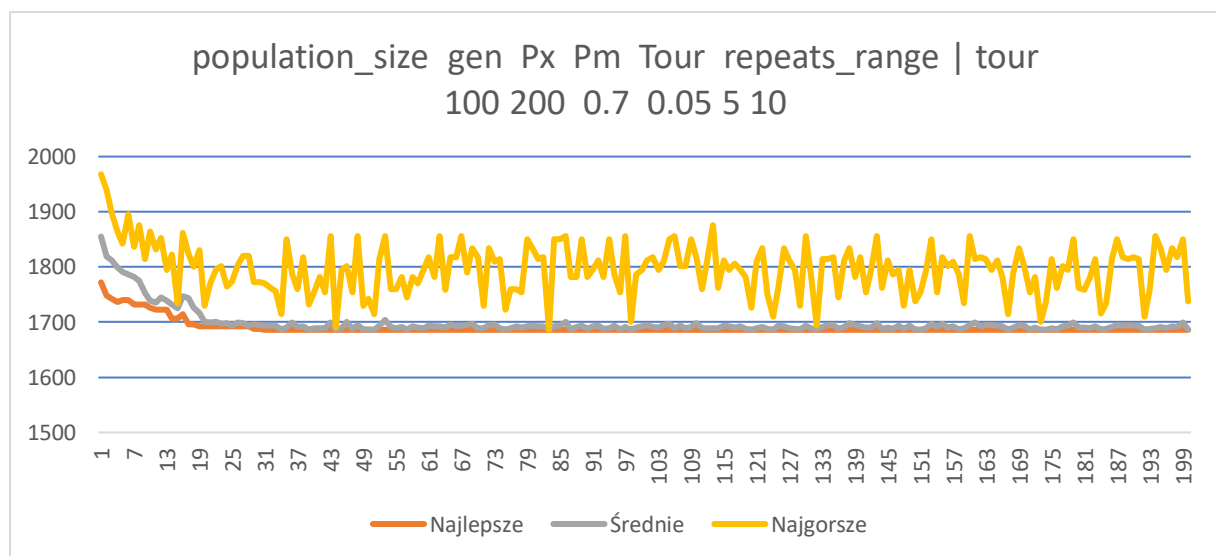
Wnioski

Jeśli zbyt szybko zmutujemy dużą ilość osobników wprowadzimy możliwość pogorszenia wyników działania algorytmu. Parametr należy dobrać tak, by działanie nie było zbyt gwałtowne, w przeciwnym wypadku doprowadzimy do znalezienia lokalnego optimum wokół którego algorytm pozostanie do końca swojego działania, wyniki średnie i najlepsze są bardzo podobne i bardzo szybko zbiegają się ku optimum, działanie jest stabilne i daje wynik najbliższy optymalnemu, już w około 30-40 pokoleniu osiągamy cel algorytmu.

Podsumowanie

Najkorzystniejsze parametry jakie udało mi się dobrać to:

population_size	gen	Px	Pm	Tour
100	200	0.7	0.05	5



Wyniki dla wybranych parametrów to:

Odchylenie	21.63703	Optimum	1662
------------	----------	---------	------

Średnie globalne:

Najlepszy	Średni	Najgorszy
1670	1677.7	1807

Rozwiązaniem jest genotyp:

8	10	5	11	12	2	6	1	3	7	4	9
---	----	---	----	----	---	---	---	---	---	---	---

Jak widać na załączonym wykresie wyniki średnie i najlepsze są bardzo podobne i bardzo szybko zbiegają się ku optimum, działanie jest stabilne i daje wynik najbliższy optymalnemu, już w około 20 pokoleniu osiągamy cel algorytmu.

Profil z PyCharm:

Name	Call Count	Time (ms)	Own Time (ms) ▼
fitness	101000	15302 74,1%	9446 45,8%
<method 'item' of 'numpy.ndarray' objects>	29088000	5856 28,4%	5856 28,4%
repair	69826	906 4,4%	808 3,9%
sample	115956	2402 11,6%	778 3,8%
_randbelow	552130	824 4,0%	565 2,7%
tournament_selection	1000	3013 14,6%	454 2,2%

Literatura

- Instrukcja do sprawozdania oraz udostępnione na laboratoriach materiały
- <http://www.genetic-programming.com/>
- https://pl.wikipedia.org/wiki/Algorytm_genetyczny