

Metoda ALS w systemach rekomendacji

Maciej Chudziński
Dariusz Burczyński
Paweł Chmiel

Naszym zadaniem było sprawdzenie skuteczności algorytmu ALS w zależności od dobranych danych i parametrów. W celu otrzymania precyzyjniejszych wyników wstępnie wyselektowaliśmy produkty pod względem rodzaju produktu (kategoria), ilości ocen (dodawaliśmy produkty, które przekroczyły określony próg ocen) oraz ograniczyliśmy ilość użytkowników (dobieraliśmy tych, którzy mają powyżej 1 ocenę).

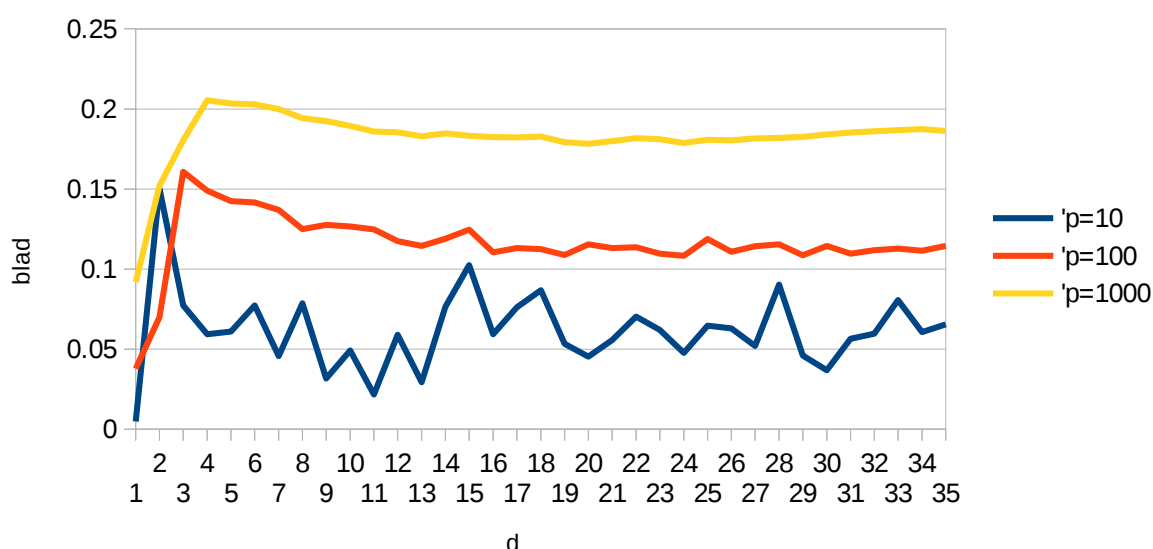
By stwierdzić poprawność algorytmu utworzyliśmy 2 zbiory - testujący i sprawdzający. Zbiór sprawdzający posiadał wszystkie uwzględnione oceny dla produktów. Zbiór testujący posiadał część ocen zakrytych - dla każdego użytkownika i produktu, posiadających przynajmniej 2 oceny, jedna została zakryta. Zastosowanie tego środka miało na celu wyeliminowanie nieprawidłowości, które mogły tworzyć „sztuczny błąd” np. produkt bez ocen.

Do określenia poprawności algorytmu pomogło nam także śledzenie wartości funkcji celu, która malała i zbiegała do pewnej liczby.

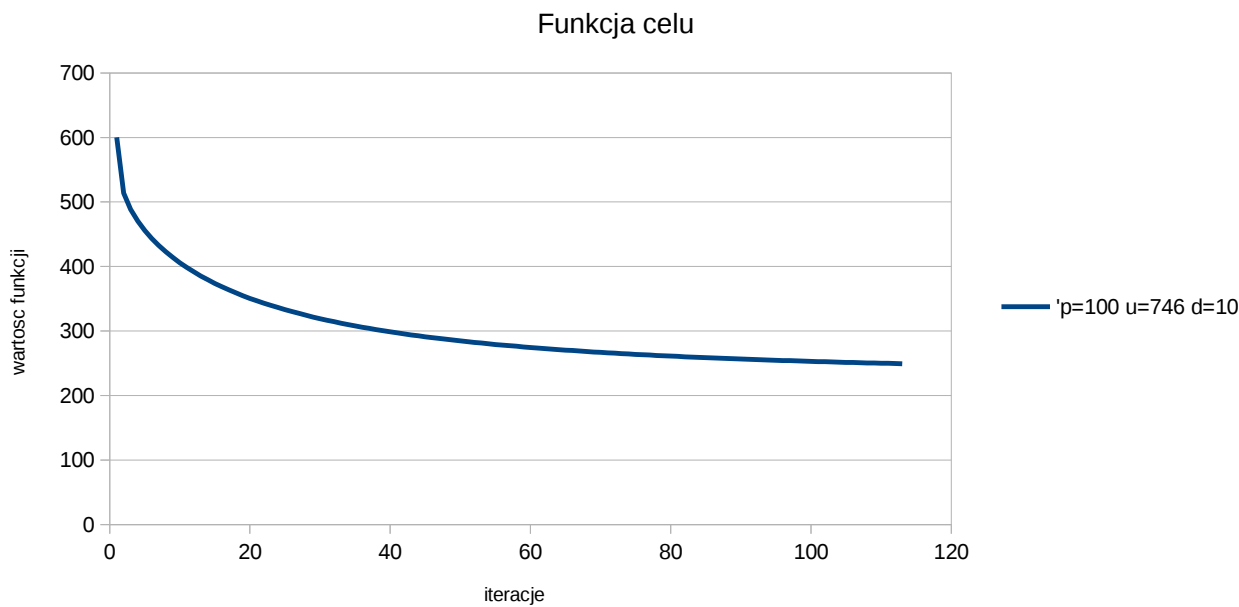
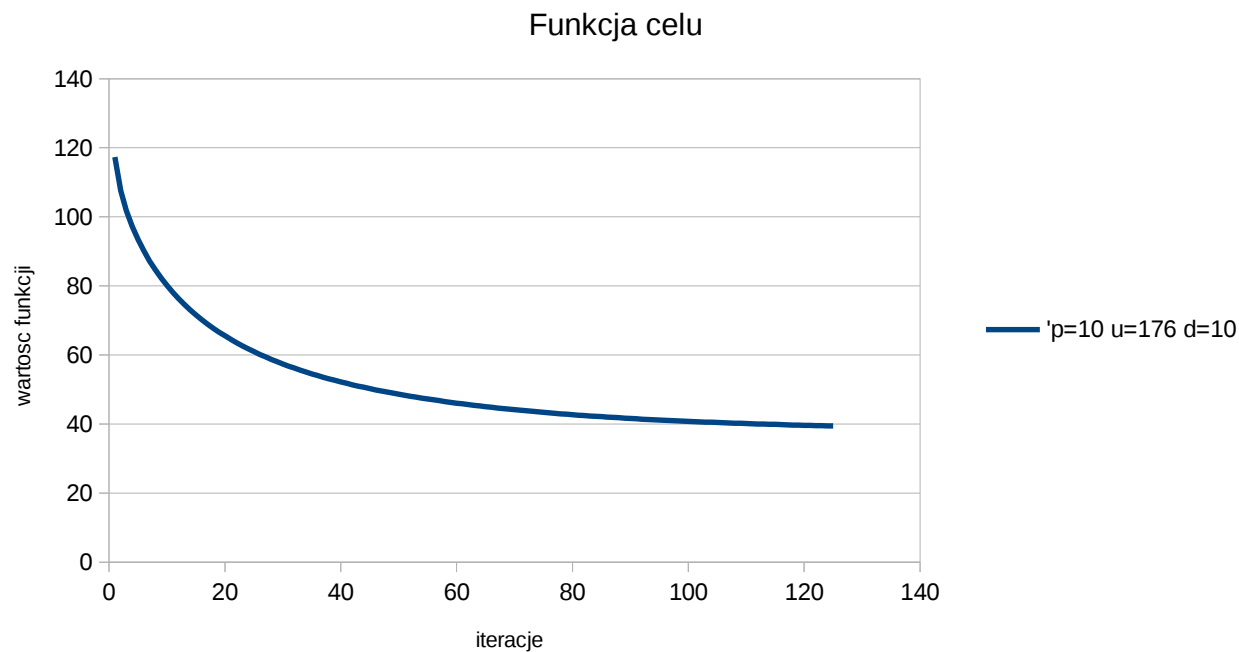
Badanie lambdy:

	Lambda									
	1,00E+02	1,00E+01	1,00E+00	1,00E-01	1,00E-02	1,00E-03	1,00E-04	1,00E-05	1,00E-06	1,00E-07
prod x user	Średni błąd									
10x2	0,454545	0,454533	0,0938	0,221191	0,0008	0,208655	0,220218	0,088798	0,162488	0,133958
100x166	0,440449	0,397968	0,310198	0,284682	0,23685	0,265567	0,215706	0,498254	0,281176	0,453569
1000x2846	0,364513	0,284088	0,166088	0,135986	0,157395	0,1842	0,255318	0,284424	0,306985	0,317453

Wykres błędu względem parametru d dla 50 iteracji



Wykresy funkcji celu do momentu różnicy 0.001



Funkcja celu

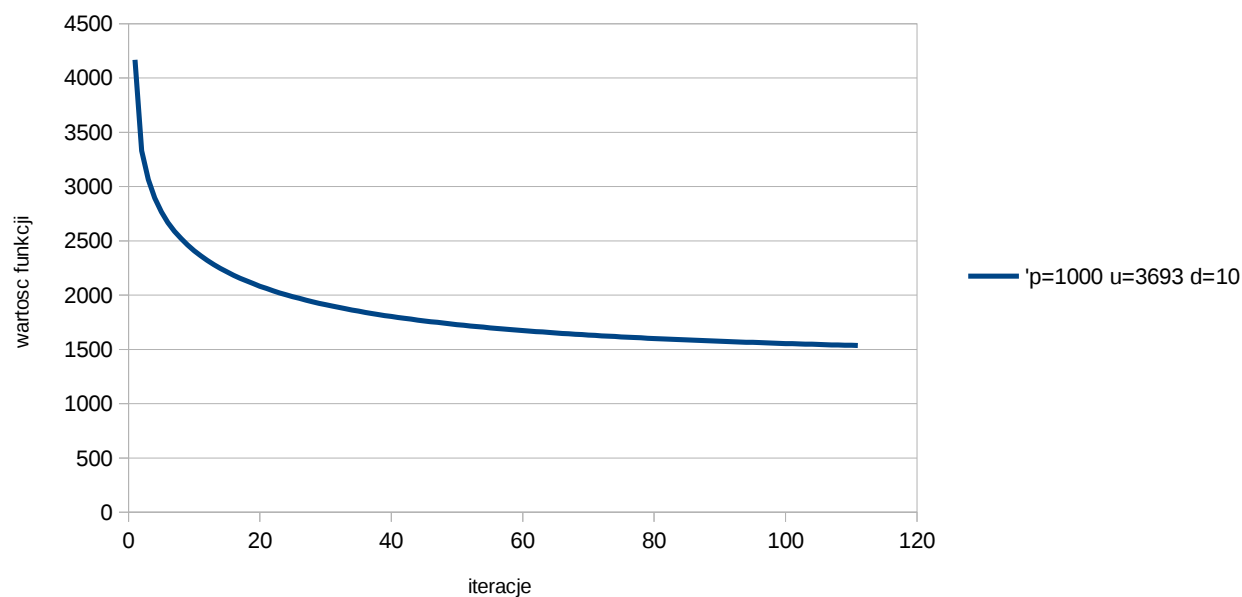


Tabela czasów i błędów dla różnych d:

Stopień różnicy 0.001	P=10 U=176		P=100 U=746		P=1000 U=3693	
d	Błąd	Czas	Błąd	Czas	Błąd	Czas
5	0,077968399004303	00:00:00.6367300	0,09042109815562	00:00:06.5699650	0,165700470184733	00:03:09.3445780
10	0,054802763562937	00:00:00.7077576	0,0871862646056	00:00:07.0829011	0,160607420606536	00:02:54.2608697
15	0,059515208246405	00:00:01.3007074	0,0823219051464	00:00:07.0624462	0,156373520322202	00:02:41.7084212
20	0,061166186358546	00:00:01.6921528	0,08130870110144	00:00:08.9969912	0,147358084070897	00:02:10.7814685
25	0,064305530515701	00:00:02.4751519	0,07594457691424	00:00:09.8493207	0,148151433785206	00:02:06.7425129
30	0,056066736872605	00:00:02.8596081	0,07420606936581	00:00:12.4135914	0,148815537584166	00:02:11.8517277
35	0,050712569470019	00:00:03.1766237	0,07835222533433	00:00:11.6034369	0,149424424332986	00:02:08.2119962

100 iteracji	P=10 U=176		P=100 U=746		P=1000 U=3693	
d	Błąd	Czas	Błąd	Czas	Błąd	Czas
5	0,07717880407254	00:00:00.5267077	0,097937665066541	00:00:04.8529855	0,168010622986122	00:02:24.8664978
10	0,059672623033573	00:00:00.5994320	0,09587424689992	00:00:06.3448377	0,159608341933425	00:02:35.9062305
15	0,064346572490539	00:00:01.1572373	0,083158284762334	00:00:08.4382805	0,154628004837126	00:02:45.6873645
20	0,058489060000623	00:00:01.8280648	0,081291739203514	00:00:11.9909854	0,152113775483914	00:02:57.5376587
25	0,061859369602958	00:00:02.8823671	0,078611223803012	00:00:16.8273457	0,153502722114439	00:03:29.2201744
30	0,052038378890814	00:00:04.2277726	0,076667158790359	00:00:23.7244013	0,157324709237373	00:04:01.0052439
35	0,050910830370935	00:00:06.1367157	0,081527136707213	00:00:32.1924735	0,159923527237686	00:04:49.4679020