

Filmy - faza 2

Justyna Jankowiak, Emilia Momotko, Adrianna Sudol

Podstawowe informacje:

Liczba zebranych filmów:

[1] 124706

Wielkość zebranych danych: 1172 MB

Szczegółowe opisy funkcji można znaleźć w pakiecie FilmyJMS dostępnym na https://github.com/tyskaj/Filmy/tree/master/filmy_pakiet

Wskaźniki:

- *frakcja* - wskaźnik podobieństwa między dwoma pojedynczymi wektorami tekstowymi w którym słowa są oddzielone separatorem '@'.
- *frakcja_powtarzalnych_slow* - wskaźnik podobieństwa między dwoma tekstami
- *czas* - funkcja sprawdzająca czy dwa czasy filmów zawierają się w tych samych przedziałach.
- *rok* - funkcja sprawdzająca czy lata wydania dwóch filmów zawierają się w tych samych przedziałach.
- *porownaj_mediany* - sprawdzenie czy dwie wartości są do siebie odpowiednio bliskie (na podstawie mediany).
- *porownaj_rozrzut* - funkcja wyznacza wskaźnik podobieństwa między dwiema wartościami rozrzutu.
- *czy_zawiera* - wskaźnik sprawdzający czy jeden element zawiera się w drugim.
- *roznica* - funkcja służąca do oceny czy dany film ma podobną ocenę, liczbę użytkowników oceniających oraz liczbę recenzji. W zależności od tego do czego ją stosujemy należy zmienić wartość *param*, która określa dopuszczalną różnicę między dwoma wartościami. Dla oceny jest to 0.3, liczby użytkowników: 50000, liczby recenzji: 20.
- *czy_to_samo* - wskaźnik określający czy porównywane filmy otrzymały Oscary bądź inne nagrody.

Macierz odległości

Macierz została zbudowana za pomocą funkcji *tworz_macierz* oraz *odleglosci* z pakietu FilmyJMS w której użyto w różnych kombinacjach wyżej wymienione wskaźniki. W sumie badamy podobieństwa na podstawie 21 różnych cech filmu, które można odpowiednio ważyć pod względem ich istotności, w zależności od preferencji użytkownika.

Fragment macierzy odległości

podobienstwa.csv - OpenOffice Calc

Albany AMT

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|----|-----------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| 1 | tt0002423 | tt0003471 | tt0003489 | tt0006745 | tt0008775 | tt0008854 | tt0009062 | tt0009475 | tt0010227 | tt0010962 | tt0011742 | tt0013597 | tt0014898 | |
| 2 | tt0002423 | 1 | 0.2933563417 | 0.3962848297 | 0.3489805595 | 0.3134328358 | 0.5610485329 | 0.4770448016 | 0.3187134503 | 0.4064026044 | 0.469144601 | 0.4927044118 | 0.3067964028 | 0.3607188703 |
| 3 | tt0003471 | 0.2933563417 | 1 | 0.4597839136 | 0.3872527734 | 0.4353236031 | 0.2912457912 | 0.3076563959 | 0.4653514739 | 0.537535113 | 0.2313244316 | 0.3052631579 | 0.5577580314 | 0.3527101335 |
| 4 | tt0003489 | 0.3962848297 | 0.4597839136 | 1 | 0.2956005932 | 0.284057971 | 0.3333333333 | 0.2762148338 | 0.3125 | 0.4696239151 | 0.2283737024 | 0.3519441675 | 0.4085224641 | 0.3457943925 |
| 5 | tt0006745 | 0.3489805595 | 0.3872527734 | 0.2956005932 | 1 | 0.42417154 | 0.2373737374 | 0.3106995885 | 0.447167756 | 0.4740655988 | 0.2941176471 | 0.2560879811 | 0.2761361446 | 0.2507488233 |
| 6 | tt0008775 | 0.3134328358 | 0.4353236031 | 0.284057971 | 0.42417154 | 1 | 0.2525252525 | 0.2777777778 | 0.5703933747 | 0.4788167227 | 0.211328976 | 0.2479532164 | 0.3334799905 | 0.3584277149 |
| 7 | tt0008854 | 0.5610485329 | 0.2912457912 | 0.3333333333 | 0.2373737374 | 0.2525252525 | 1 | 0.5320855615 | 0.2549019608 | 0.2402402402 | 0.3409090909 | 0.4912698413 | 0.2496818699 | 0.1892655367 |
| 8 | tt0009062 | 0.4770448016 | 0.3076563959 | 0.2762148338 | 0.3106995885 | 0.2777777778 | 0.5320855615 | 1 | 0.3350383632 | 0.2044917258 | 0.3583877996 | 0.5765432099 | 0.2074122871 | 0.211673699 |
| 9 | tt0009475 | 0.3187134503 | 0.4653514739 | 0.3125 | 0.447167756 | 0.5703933747 | 0.2549019608 | 0.3350383632 | 1 | 0.515658454 | 0.2426470588 | 0.2768361582 | 0.3626786235 | 0.3821391485 |
| 10 | tt0010227 | 0.4064026044 | 0.537535113 | 0.4696239151 | 0.4740655988 | 0.4788167227 | 0.2402402402 | 0.2044917258 | 0.515658454 | 1 | 0.2293265132 | 0.2614035088 | 0.4365565075 | 0.3611781748 |
| 11 | tt0010962 | 0.469144601 | 0.2313244316 | 0.2283737024 | 0.2941176471 | 0.211328976 | 0.3409090909 | 0.3583877996 | 0.2426470588 | 0.2293265132 | 1 | 0.5183901919 | 0.2498543124 | 0.3602642276 |
| 12 | tt0011742 | 0.4927044118 | 0.3052631579 | 0.3519441675 | 0.2560879811 | 0.2479532164 | 0.4912698413 | 0.5765432099 | 0.2768361582 | 0.2614035088 | 0.5183901919 | 1 | 0.2264526749 | 0.3275132275 |
| 13 | tt0013597 | 0.3067964028 | 0.5577580314 | 0.4085224641 | 0.2761361446 | 0.3334799905 | 0.2496818699 | 0.2074122871 | 0.3626786235 | 0.4365565075 | 0.2498543124 | 0.2264526749 | 1 | 0.3065558633 |
| 14 | tt0014898 | 0.3607188703 | 0.3527101335 | 0.3457943925 | 0.2507488233 | 0.3584277149 | 0.1892655367 | 0.211673699 | 0.3821391485 | 0.3611781748 | 0.3602642276 | 0.3275132275 | 0.3065558633 | 1 |
| 15 | tt0015145 | 0.1853888452 | 0.293017116 | 0.2262132984 | 0.2365037594 | 0.2713271729 | 0.1786788787 | 0.1938083121 | 0.2919287212 | 0.2640788507 | 0.3667857143 | 0.3188251946 | 0.2383921652 | 0.3010925628 |
| 16 | tt0015634 | 0.2171717172 | 0.386177116 | 0.3425605536 | 0.2654739485 | 0.4022147147 | 0.1437908497 | 0.2702702703 | 0.3746461151 | 0.2723850224 | 0.2944915254 | 0.2847058824 | 0.4492462042 | 0.2746566792 |
| 17 | tt0015684 | 0.4170692432 | 0.5364739008 | 0.3585122699 | 0.346838967 | 0.4031490428 | 0.1887468031 | 0.1455337691 | 0.4174430054 | 0.4544492544 | 0.3899441341 | 0.2776762879 | 0.4754147813 | 0.4237037037 |
| 18 | tt0016308 | 0.253689969 | 0.2336633909 | 0.2299465241 | 0.2792879635 | 0.2866923406 | 0.1326164875 | 0.2100271003 | 0.297979798 | 0.3220802005 | 0.3621031746 | 0.2181069959 | 0.2870895187 | 0.2615301139 |
| 19 | tt0016461 | 0.1840238895 | 0.33041571 | 0.2189936215 | 0.3516483516 | 0.3577421815 | 0.1224105461 | 0.1368760064 | 0.3802279826 | 0.3577944862 | 0.1771978022 | 0.1504065041 | 0.3705912179 | 0.2954936361 |
| 20 | tt0016914 | 0.3765371372 | 0.2581926514 | 0.3475935829 | 0.2091038407 | 0.2955465587 | 0.2551892552 | 0.2786324786 | 0.2727272727 | 0.259230165 | 0.3883108108 | 0.3507422402 | 0.3226681812 | 0.3122521263 |
| 21 | tt0016953 | 0.4027777778 | 0.3096539162 | 0.3585526316 | 0.2017017017 | 0.2752902156 | 0.3372270894 | 0.3610024743 | 0.2786377709 | 0.3178694158 | 0.4547404547 | 0.4050600886 | 0.321459695 | 0.325203252 |
| 22 | tt0017029 | 0.2002419845 | 0.332142181 | 0.3037567968 | 0.2689346979 | 0.2882205514 | 0.1875 | 0.208994709 | 0.3027950311 | 0.3826976348 | 0.23828125 | 0.2747474747 | 0.4483742563 | 0.2178932179 |
| 23 | tt0018067 | 0.3409728662 | 0.2701525054 | 0.3720238095 | 0.218308592 | 0.314058959 | 0.305982906 | 0.3004474517 | 0.2913165266 | 0.2760416667 | 0.3036384977 | 0.3806203182 | 0.2845285256 | 0.329283139 |
| 24 | tt0018179 | 0.2437174016 | 0.3385534388 | 0.2283737024 | 0.3737604882 | 0.4785575049 | 0.1262626263 | 0.1995884774 | 0.5285517716 | 0.3790305804 | 0.375 | 0.2703150912 | 0.3436660805 | 0.3622039316 |
| 25 | tt0019151 | 0.3134328358 | 0.323852618 | 0.31543052 | 0.3194083652 | 0.4019138756 | 0.1414141414 | 0.2222222222 | 0.4111963626 | 0.4410974244 | 0.224537037 | 0.2617283951 | 0.379287611 | 0.4110592938 |
| 26 | tt0019735 | 0.2951171635 | 0.4419613198 | 0.3390783065 | 0.3120997246 | 0.3652930777 | 0.1847407358 | 0.1439645626 | 0.4060992086 | 0.4213968884 | 0.3714455529 | 0.2637992832 | 0.5447582371 | 0.2975146199 |
| 27 | tt0019824 | 0.2386506063 | 0.3459495734 | 0.221719457 | 0.2859853121 | 0.3729139923 | 0.1790123457 | 0.1944444444 | 0.339031339 | 0.2815217206 | 0.3677325581 | 0.3751803752 | 0.282501282 | 0.2977548767 |
| 28 | tt0020092 | 0.3607188703 | 0.3729530485 | 0.4046179219 | 0.2657864173 | 0.4349827388 | 0.1892655367 | 0.211673699 | 0.4082829393 | 0.2977642276 | 0.2164021164 | 0.3844506002 | 0.4210526316 | |
| 29 | tt0020198 | 0.1847867845 | 0.324770259 | 0.2788671024 | 0.3155091224 | 0.2995735608 | 0.1228070175 | 0.1932006633 | 0.3201377706 | 0.3604423818 | 0.2486832865 | 0.3223290598 | 0.3719522609 | 0.2700459202 |
| 30 | tt0020223 | 0.2933563417 | 0.4343175011 | 0.3421368547 | 0.3180856026 | 0.3023695965 | 0.1801346801 | 0.1409897292 | 0.3254440665 | 0.4099661906 | 0.3707822086 | 0.2666666667 | 0.4256208704 | 0.3070960985 |
| 31 | tt0020230 | 0.2667810831 | 0.1874728143 | 0.2190237797 | 0.190430622 | 0.1562998405 | 0.2021881216 | 0.2819422293 | 0.1513002364 | 0.1425438596 | 0.3554467085 | 0.268115942 | 0.1933298593 | 0.2020792723 |

Arkusz1 / 1

Domyślnie

STD

Suma=0

11:48

2015-05-20

Mapa filmów





