

Wyznaczanie okresu gwiazd zmiennych metodą
minimum entropii na podstawie danych
fotometrycznych z projektu Pi of the Sky

Mikołaj Grzędzielski

Streszczenie

ćwiczenie polegało na implementacji i dyskusji stosowania metody minimalnej entropii do poszukiwania okresu gwiazd zmiennych. Metoda ta jest efektywna dla danych okresowych o dowolnej postaci zmienności, lecz tylko jeżeli poszukujemy częstości w danym przedziale.

1 Poszukiwanie periodyczności

W pomiarach fizycznych często spotykamy się z danymi, co do których chcielibyśmy oczekiwać periodycznych zmian w czasie. Podstawowe przykłady to sygnały akustyczne, elektromagnetyczne, bądź też gwiazdy zmienne. Podstawową metodą wyłuskiwania okresowości z danych jest transformata Fouriera którą z powodzeniem można stosować do obliczania okresów danych o w miarę regularnym próbkowaniu, takim że okres próbkowania jest dużo mniejszy od okresu sygnału, co oznacza też, że odstępów czasowych pomiędzy poszczególnymi próbkami muszą być tego samego rzędu wielkości.

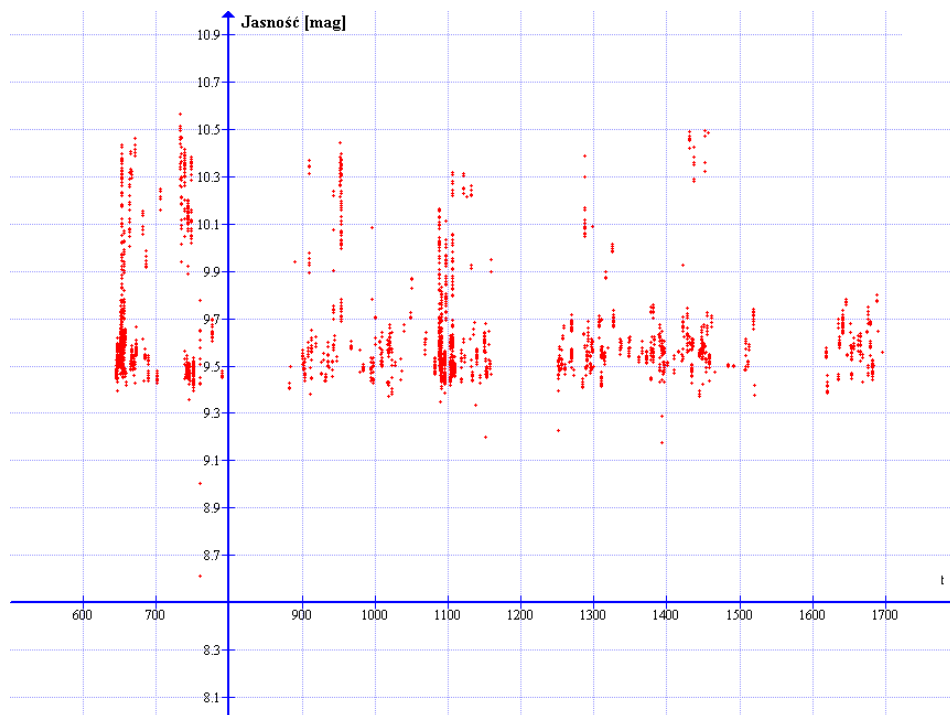
Dane astronomiczne jednak nie spełniają tego założenia ze względu na to że w danym miejscu na Ziemi gwiazdy widoczne są tylko w niektórych porach dnia, a w pewnych wypadkach, tylko w niektórych porach roku. Częstości próbkowania towarzyszą więc zmienność dzienna i roczna, charakteryzujące się tym, że w niektórych porach roku czy dnia pomiary w ogóle nie są przeprowadzane

Na rysunku nr 1 umieszczono wykres danych z przykładowej gwiazdy numer 453131 (dane pochodzą z eksperymentu Pi of the Sky) (gwiazda zaćmieniowa typu Algol)

Na rysunku widać, jak postać danych została zniekształcona przez zmienność wynikającą z pór roku. Wyraźne piki jasności widoczne w odstępach kilkudniowych są zauważalne, lecz regularność ich występowania została mocno zaburzona. Warto jeszcze dodać, że dane oznaczają ujemny logarytm z jasności, tak więc pik 'do góry' oznacza de facto spadek jasności.

2 Metoda najmniejszej entropii

Jedną z metod poszukiwania okresu, właściwszą dla danych mniej regularnie rozłożonych w czasie jest metoda poszukiwania minimum entropii informacji. Alternatywną metodą jest między innymi metoda długości sznurka będąca rozszerzeniem metody Laflera- Kinmana zaproponowana przez Dworetzky'ego w 1982 roku, opierająca się na policzeniu całkowitej odległości między punktami na fazowanej krzywej blasku (długość sznurka). Im mniejsza wartość długości sznurka, tym krzywa jest 'gładsza', a więc najmniejsza wartość jest dla prawdziwego okresu [2]



Rysunek 1: Przykładowa gwiazda - zależność jasności od czasu

2.1 Normalizacja

Rozpoczemy od normalizacji danych doświadczalnych, dzieląc różnicę jasności dla danej chwili czasu m_i jasności minimalnej m_{min} przez różnicę między jasnością maksymalną m_{max} a minimalną

$$(1) \quad \tilde{m}_i = \frac{m_i - m_{min}}{m_{max} - m_{min}}$$

Wskutek czego otrzymujemy liczby z przedziału $[0, 1]$.

2.2 Próbkowanie dla danej częstości

Jeśli dane są już znormalizowane, przystępujemy do kolejnego etapu, mającego na celu zmieszczenie wszystkich danych doświadczalnych w kwadracie jednostkowym (co będzie potrzebne dla wyznaczenia entropii informacji).

W tym celu punkt o współrzędnych (\tilde{x}_i, t_i) odwzorowujemy na punkt zawarty w kwadracie jednostkowym, według następującego wzoru

$$(2) \quad (\tilde{x}_i, t_i) \mapsto \left(\tilde{x}_i, \frac{t_i - t_0}{p}\right)$$

Gdzie t_0 oznacza czas wykonania pierwszego pomiaru, a p próbkowany okres. Wartość p powinna być większa niż najmniejsza spotykana różnica czasu pomiędzy pomiarami i mniejsza niż czas trwania całych pomiarów.

W wyniku powyższej procedury otrzymujemy dwuwymiarową mapę w postaci kwadratu jednostkowego, na którym zaznaczone są punkty odpowiadające poszczególnym pomiaram. Dla różnych wartości okresów konfiguracje są różne i to właśnie pojęcie entropii mówi nam o uporządkowaniu tych punktów na mapie. Jeśli próbkowany okres odpowiada rzeczywistemu okresowi, wtedy punkty układają się w uporządkowane struktury np. w linie czy pasma co oznacza niską wartość entropii, a jeżeli nie odpowiada on okresowi, wtedy rozłożenie punktów jest chaotyczne i entropia jest wysoka.

Aby zmierzyć entropię dla danego okresu, konieczne jest podzielenie kwadratu jednostkowego na $n \times n$ kwadratowych obszarów 'boksów'. Entropię informacji definiujemy zgodnie z wzorem [1]

$$(3) \quad S(p) = - \sum_i \sum_j \mu_{ij} \log \mu_{ij}$$

gdzie μ_{ij} oznacza ilość punktów w obszarze $[\frac{i-1}{n}, \frac{i}{n}] \times [\frac{j-1}{n}, \frac{j}{n}]$ dzieloną przez całkowitą liczbę punktów pomiarowych N . Oczywiście jest fakt, że im większe n , tym większa dokładność pomiaru okresu, jednakże zbyt duża jego wartość może wydłużyć znacząco czas działania programu. Program przetestowano więc dla wartości

$$(4) \quad n = \sqrt[4]{N}$$

Gdzie N oznacza całkowitą ilość punktów pomiarowych użytych w ćwiczeniu.

Oznacza to więc, że jeżeli wszystkie punkty skupione będą w jednym 'boksie' to wtedy entropia wynosić będzie 0, a jeśli będą one rozłożone równomiernie to $2 \log n$, zawsze jednak entropia będzie wartością dodatnią. Ilość boksów n^2 może być kwadratem dowolnej liczby naturalnej, w praktyce jednak nie ma sensu aby n^2 było większe od N

3 Wyniki

Powyższy algorytm zastosowano do napisania programu wyszukującego minimalną wartość entropii informacji `s8.exe`, który w wersji `s7.exe` drukował dodatkowo zależność entropii od próbkowanego okresu. Osobną kwestią jest sam dobór próbkowania. Założyliśmy, że poszukiwany okres jest większy od średniej odległości między punktami pomiarowymi wynoszącej około 0.3 dnia a mniejszy od połowy długości trwania pomiarów. (dla opisanej wcześniej gwiazdy nr 453131).

Jeżeli badamy okres na przestrzeni kilku rzędów wielkości, więc sensowne będzie zastosowanie próbkowania logarytmicznego, w którym każda następną częstotliwość jest większa od poprzedniej określoną ilość razy, na przykład 1%. Rozdzielczość próbkowania zwiększa dokładność pomiaru, ale jednocześnie wydłuża czas działania programu.

Rozdzielczość (%)	Zmierzony okres	Czas działania programu (s)
1	1.00001	0.9
0.5	2.9903	1.79
0.2	1.9938	4.58
0.1	0.99743	8.94
0.05	0.997231	19.14
0.02	0.994	47.07
0.01	0.997	92.48

Tabela 1: Minimalne wartości okresu oraz czas wykonywani programu dla próbkowanego przedziału $[0.3, 3]$

Rozdzielczość (%)	Zmierzony okres	Czas działania programu (s)
1	1.002	1.4
0.5	2.9903	2.75
0.2	1.9938	6.91
0.1	0.99743	13.75
0.05	0.997231	27.88
0.02	0.9999	70.61
0.01	0.9972	149.87

Tabela 2: Minimalne wartości okresu oraz czas wykonywani programu dla próbkowanego przedziału $[0.3, 10]$

W tabelach nr 1 i 2 podano czas działania programu `s8.exe` w zależności od rozdzielności i uzyskany wynik przedstawiono w tabelach 1 i 2

Z danych przedstawionych w tabeli widać że czas działania programu jest z dobrym przybliżeniem proporcjonalny do rozdzielności próbkowania

Okres zmienności tej gwiazdy, mierzony inną metodą wynosił 2.519653 dnia. Jak wynika z danych przedstawionych na wykresie, okres dla którego entropia była najniższa był bliski 1, 2 bądź 3 dni. Wniosek jest taki że dla tego przedziału dzienna zmienność obserwacji jest znacznie bardziej zauważalna niż zmienność związana z okresowymi zmianami jasności. Aby poradzić sobie z tym problemem, możemy albo zwiększyć ilość boksów (z obecnej \sqrt{N} czyli 49), co jednak znacznie wydłużyłoby czas działania programu, albo zawęzić swój obszar poszukiwań, wyłączając z obszaru próbkowania otoczenia liczb całkowitych.

Najprostszą tego realizacją będzie zawężenie próbkowanego przedziału do $[2.2, 2.8]$. Kolejną korzyścią z takiego postępowania jest skrócenie czasu

Rozdzielczość (%)	Zmierzony okres	Czas działania programu (s)
1	2.52884	0.11
0.5	2.51713	0.22
0.2	2.52016	0.5
0.1	2.52033	1.04
0.05	2.51915	1.9
0.02	2.51996	4.76
0.01	2.51998	9.64
0.005	2.51999	19.31
0.002	2.51994	47.64

Tabela 3: Minimalne wartości okresu oraz czas wykonywania programu dla próbkowanego przedziału [2.2, 2.8]

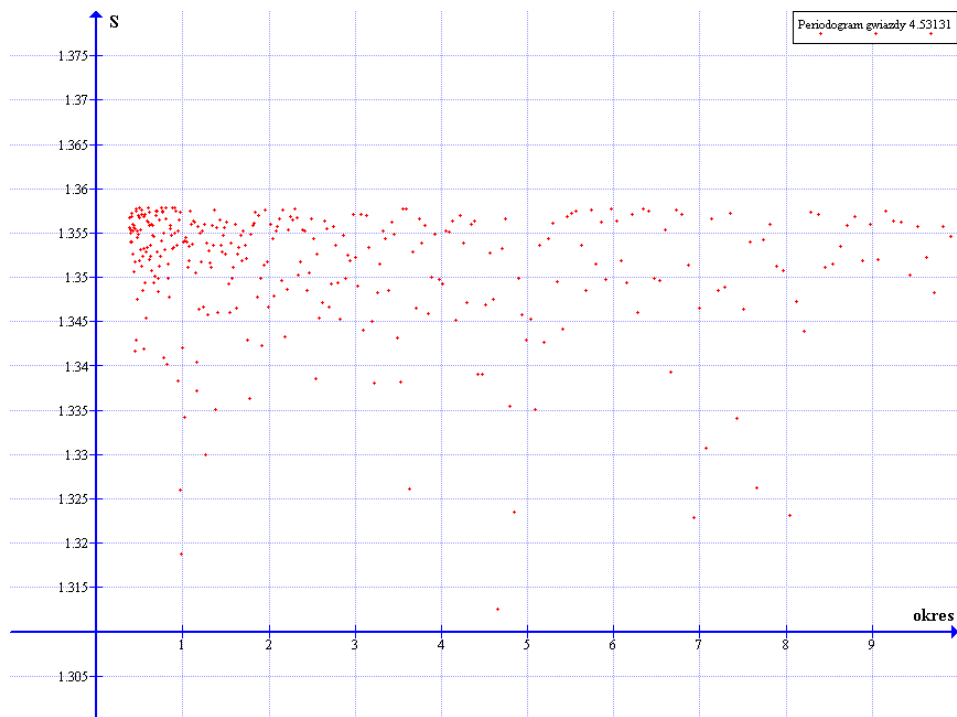
działania programu np. dla rozdzielczości 1% wynosi on 0.11s wobec 1.4s w poprzednim przypadku. Wyniki dla przedziału [2.2, 2.8], odpowiadające oczekiwanemu okresowi, umieszczono w tabeli nr 3.

Program przetestowano na przykładowej gwiazdzie obserwowanej przez detektor Pi of the Sky w Las Campanas w Chile w sezonie 2006-2009. Na rysunku nr 2 przedstawiono krzywą blasku gwiazdy. Poniżej widoczny jest wykres entropii w zależności od próbkowanego okresu.

Pokazuje on, że minima entropii dla zmienności dobowej są znacznie głębsze od tych związanych z rzeczywistą periodycznością zmian jasności gwiazdy.

4 Podsumowanie

Wyżej przedstawiona metoda wydaje się być nieefektywna, z powodu konieczności silnego zawężania próbkowanego przedziału częstości, przez co nie nadaje się ona do wykorzystania na skalę masową do danych zebranych przez eksperyment Pi of the Sky.



Rysunek 2: Periodogram gwiazdy 453131

Literatura

- [1] P.M. Cincotta, A.Helmi, M.Mendez, J.A. Nunez, H.Vucetich (1998),
Astronomical time-series analysis - II. A search for periodicity using
the Shannon entropy
- [2] M.Siudek (2010), Klasyfikacja gwiazd zmiennych na podstawie analizy
danych fotometrycznych w eksperymencie „Pi of the Sky”