

Projekt 2

Aleksander Milach

04/07/2020

Opis problemu, cel działania programu

Otrzymaliśmy dane, zawierające informacje o poborze prądu przez pięć urządzeń elektrycznych w pewnym gospodarstwie domowym. Pomiary wykonywane były w bardzo krótkich odstępach czasu, jest ich tyle samo dla każdego urządzenia. Naszym zadaniem będzie zbadanie, do którego urządzenia najbardziej ‘pasuje’ podany nam inny ciąg danych o poborze prądu, ściślej, sklasyfikowanie tego ciągu, jako jedno z pięciu urządzeń, dla których pomiary otrzymaliśmy na początku.

Do rozwiązania tego problemu skorzystamy z ukrytych modeli Markowa. Dla każdego z urządzeń wyuczymy model Markowa na ciągu obserwacji przyporządkowanych do tego urządzenia z najbardziej wiarogodną ilością stanów. W tym momencie, program ‘wie’ jak powinny wyglądać ciągi obserwacji pod warunkiem, że są pomiarami dla określonego urządzenia. Następnie dla każdego z plików zawierających pomiary, które mamy za zadanie sklasyfikować, obliczymy wiarogodność ciągu dla danych dla każdego modelu. Jako wynik klasyfikacji uznamy urządzenie, z którym związany jest model o największej wiarogodności.

Wyniki i komentarze

Poniżej umieściłem wiarygodności modeli obliczone, dla ilości stanów równych n . Najlepsze wyniki oznaczone są kursywą. Każdy model uczył się na zbiorze danych, a następnie liczył dla niego wiarogodność. Dla lodówki obliczyłem dla nieco większych wartości parametru, ponieważ likelihood był sporo mniejszy od pozostałych. Co prawda przez to program działa zauważalnie dłużej, ale otrzymany wynik optymalnego modelu ($n = 18$) jest bardziej już bliższy pozostałym. Dla trzech z pięciu urządzeń najlepszy wynik osiągnął model dla największej liczby stanów. Prawdopodobnie więc dla większych n uzyskałbym jeszcze lepszą wiarogodność. Jednak ze względu na wydłużający się przez to znacznie czas działania, zdecydowałem się na to tylko dla lodówki, ponieważ tam likelihood dla $n \in [2, \dots, 8]$ był sporo mniejszy od pozostałych.

	n = 2	n = 3	n = 4	n = 5	n = 6	n = 7	n = 8
lighting2	-60152	-54789	35426	37117	<i>42926</i>	38633	38882
lighting5	-59818	42799	49598	50763	55278	56150	<i>57121</i>
lighting4	-68129	21357	26273	30462	33099	34433	<i>36107</i>
refrigerator	-38975	-27066	-26114	-23904	-23814	-23811	-22445
microwave	-12937	-11131	62293	62307	75962	<i>75970</i>	67809

	n = 9	n = 10	n = 11	n = 12	n = 13	n = 14	n = 15	n = 16	n=17	n=18
refrigerator	-21691	-21247	-20913	-20908	-20068	-21037	-20655	31526	32634	<i>33069</i>

W następujących tabelach umieszczone są wyniki klasyfikacji dla udostępnionych plików testowych. W każdym z przypadków poza plikiem szóstym wynik dla prawdziwego urządzenia był znacznie wyższy niż dla innych urządzeń. Dla ostatniego pliku ta różnica była nieco mniejsza. Każdy z plików został prawidłowo zaklasyfikowany.

	lighting2	lighting5	lighting4	refrigerator	microwave
dev1	<i>12434</i>	-462	-9441	3488	-15040
dev2	<i>10081</i>	-211505	-807856	-13955	-31560
dev3	-29938	-218717	-144546	<i>9202</i>	-32167
dev4	-63643	-1490303	-1301741	-65565	<i>45203</i>
dev5	-18200	<i>18009</i>	-19518	5809	-13758
dev6	7288	-12623	<i>9724</i>	1312	-14473