

Методы разделения смеси распределений в условиях неполных данных

Научный руководитель – Илларионов Егор Александрович

Ситдиков Даниил Уралович

Студент (специалист)

Московский государственный университет имени М.В.Ломоносова,
Механико-математический факультет, Кафедра теории вероятностей, Москва, Россия
E-mail: mrfarbo@gmail.com

Одним из параметров для анализа солнечной активности является число Вольфа, которое определяется как $R_Z = k(10g + n)$, где g - число групп солнечных пятен, n - число наблюдаемых пятен, k - нормировочный коэффициент. Однако для вычисления числа Вольфа необходимо сначала выделить группы солнечных пятен.

В работе [1] собраны данные положений солнечных пятен, полученных из Цюрихской обсерватории, но солнечные пятна не были объединены в группы.

Мы будем рассматривать задачу разделения смеси распределений в условиях неполных данных. На основе данных о положении центров из обсерватории Гринвича мы будем каждому центру сопоставлять пятна, полученные обработкой данных из Цюрихской обсерватории.

Считая, что пятна в группе имеют нормальное распределение, для разделения смеси гауссиан, при условии известных центров мы будем пользоваться алгоритмом EM (Expectation-maximization) [2]. Благодаря данному подходу мы можем сопоставить каждому пятну некоторый центр, выделить новые центры, которые не встретились в данных обсерватории Гринвича, а также рассчитать вероятности принадлежности пятен к полученным кластерам.

Источники и литература

- 1) Illarionov, E., Arlt, R. Reconstruction of the Solar Activity from the Catalogs of the Zurich Observatory Sol Phys 297, 79 (2022).
- 2) Dempster, A.P.; Laird, N.M.; Rubin, D.B. (1977). "Maximum Likelihood from Incomplete Data via the EM Algorithm"