

Лабораторная работа №3

1. Магнитоэнцефалография (МЭГ) является неинвазивным функциональным методом исследования головного мозга, основанным на регистрации и анализе магнитных полей, которые возникают в результате протекания электрических процессов в головном мозге. При анализе данных МЭГ встает проблема решения так называемой обратной задачи, которая состоит в восстановлении распределения активности нейронных источников на поверхности коры головного мозга. Некоторые подходы к решению такой задачи основаны на дипольном моделировании нейронных источников.

Рассматривается упрощенная модель определения характеристик статичного диполя [1], находящегося в пределах плоскости $P = \{\mathbf{r} = (r_1, r_2, 0)\}$. Система сенсоров расположена на некотором удалении h от плоскости. Источник магнитного поля характеризуется четырьмя степенями свободы: декартовы координаты положения на плоскости $p = (p_1, p_2)$; две координаты $q = (q_1, q_2)$, характеризующие дипольный момент. Компонента магнитного поля, перпендикулярная плоскости, регистрируется датчиками. Введем следующие обозначения: $x = (p, q) \in \mathbb{R}^4$; $\mathbf{q} = (q, 0) \in \mathbb{R}^3$; $\mathbf{p} = (p, 0) \in \mathbb{R}^3$. Пусть всего имеется L сенсоров. Согласно закону Био-Савара-Лапласа регистрируемая сенсором, находящимся в точке \mathbf{r}_j , компонента вектора магнитной индукции определяется следующим выражением:

$$b_j(x) = \frac{\mu_0}{4\pi} \frac{\mathbf{e} \cdot \mathbf{q} \times (\mathbf{r}_j - \mathbf{p})}{|\mathbf{r}_j - \mathbf{p}|^3}, \quad j = 1, \dots, L,$$

где μ_0 – магнитная постоянная; \mathbf{e} – единичный вектор, перпендикулярный плоскости P . Обозначим

$$b(x) = (b_1(x), \dots, b_L(x)).$$

Предполагается, что модель наблюдения имеет следующий вид

$$y = b(x) + V,$$

где шум V распределен нормально с нулевым средним и некоторой матрицей ковариаций Γ .

Требуется найти оценку характеристик диполя на основе вектора наблюдений y , предполагая, что априорное распределение характеристик диполя имеет многомерное нормальное распределение с вектором средних $(p_1^0, p_2^0, q_1^0, q_2^0)$ и диагональной матрицей ковариаций.

2. Рассматривается латентное размещение Дирихле (LDA – Latent Dirichlet Allocation) – вероятностная модель порождения текста, предназначенная для описания текстов с точки зрения их тематик. При этом тема рассматривается как некоторое распределение вероятностей в пространстве слов из общего словаря [2]. Используемые в дальнейшем обозначения приведены в табл. 1.

Таблица 1: Основные обозначения

$w \in \{1, \dots, W\}$	– номер слова в словаре
$t \in \{1, \dots, T\}$	– номер темы
N_d	– число слов в документе d
$\mathbf{w}_d = [w_{d,1}, \dots, w_{d,N_d}]$	– слова в документе d , $w_{d,n} \in \{1, \dots, W\}$
$\mathbf{z}_d = [z_{d,1}, \dots, z_{d,N_d}]$	– темы документа d , $z_{d,n} \in \{1, \dots, T\}$
$\boldsymbol{\theta}_d = [\theta_{d,1}, \dots, \theta_{d,T}]$	– вероятности тем в документе d
$\boldsymbol{\phi}_t = [\phi_{t,1}, \dots, \phi_{t,W}]$	– вероятности слов в теме t
$\boldsymbol{\Theta} = [\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_D]^T \in \mathbb{R}^{D \times T}$	– вероятности тем во всех документах
$\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_T]^T \in \mathbb{R}^{T \times W}$	– вероятности слов во всех темах
$\mathbf{W} = \{\mathbf{w}_1, \dots, \mathbf{w}_D\}$	– набор всех слов в корпусе
$\mathbf{Z} = \{\mathbf{z}_1, \dots, \mathbf{z}_D\}$	– разбиение всех слов по темам

Модель LDA задается следующим образом:

$$\begin{aligned}
 p(\mathbf{W}, \mathbf{Z}, \boldsymbol{\Theta}, \boldsymbol{\Phi} | \alpha, \beta) &= \prod_{t=1}^T p(\boldsymbol{\phi}_t | \beta) \prod_{d=1}^D p(\boldsymbol{\theta}_d | \alpha) \prod_{n=1}^{N_d} p(w_{d,n} | z_{d,n}, \boldsymbol{\Phi}) p(z_{d,n} | \boldsymbol{\theta}_d), \\
 p(\boldsymbol{\phi}_t | \beta) &= \text{Dir}(\boldsymbol{\phi}_t | \beta), \quad p(\boldsymbol{\theta}_d | \alpha) = \text{Dir}(\boldsymbol{\theta}_d | \alpha), \\
 p(w_{d,n} | z_{d,n}, \boldsymbol{\Phi}) &= \boldsymbol{\Phi}_{z_{d,n}, w_{d,n}}, \quad p(z_{d,n} | \boldsymbol{\theta}_d) = \boldsymbol{\theta}_{d, z_{d,n}},
 \end{aligned}$$

где $\text{Dir}(\cdot | \gamma)$ означает распределение Дирихле. Требуется реализовать схему Гиббса для маргинального распределения $p(\mathbf{Z} | \mathbf{W}, \alpha, \beta)$ (так называемый collapsed Gibbs sampling, см. [2, 3, 4]).

В файлах 'test1.dat' и 'test2.dat' записаны данные в виде таблицы: первый столбец – номер документа, второй столбец – номер слова из словаря, третий столбец – сколько раз текущее слово встречается в данном документе. Для первого тестового примера задать следующие значения параметров: $T = 3$; $\alpha = 1$; $\beta = 1$; для второго примера: $T = 20$; $\alpha = 0.1$; $\beta = 0.1$.

3. Рассматривается анизотропный вариант модели Изинга на прямоугольной решетке с системой соседства первого рода, для которой распределение конфигурации $\mathbf{x} = (x_1, \dots, x_n)$ (n – число узлов; $x_i \in \{-1, +1\}$) задается следующим соотношением

$$\pi(\mathbf{x}) = \frac{1}{Z} \exp(-U(\mathbf{x})),$$

где Z – нормировочная константа (статистическая сумма), а энергия определяется следующим выражением

$$U(\mathbf{x}) = -\beta_1 \sum_{\boxed{\begin{smallmatrix} \mathbf{i} & \mathbf{j} \end{smallmatrix}}} x_i x_j - \beta_2 \sum_{\boxed{\begin{smallmatrix} \mathbf{i} \\ \mathbf{j} \end{smallmatrix}}} x_i x_j.$$

Требуется реализовать процедуру генерации конфигураций модели с помощью методов МСМС.

Литература:

- [1] *E. Somersalo, A. Voutilainen, J.P. Kaipio.* Non-stationary magnetoencephalography by Bayesian filtering of dipole models. *Inverse Probl.*, 19 (2003) 1047-1063. doi: 10.1088/0266-5611/19/5/304
- [2] URL: Лекция «Латентное размещение Дирихле (LDA)»
http://www.machinelearning.ru/wiki/images/8/82/BMMO11_14.pdf
- [3] G. Heinrich. Parameter estimation for text analysis. Tech. report, 2005.
<http://www.arbylon.net/publications/text-est2.pdf>
- [4] T. L. Griffiths, M. Steyvers. Finding scientific topics. *Proceedings of the National Academy of Sciences*, 2004, 101 (suppl 1) 5228-5235.
https://www.pnas.org/content/pnas/101/suppl_1/5228.full.pdf