

СТАНОВИЩЕ

от доц. д-р Нейко Матеев Нейков, НИМХ-БАН

относно дисертационен труд на тема "Алгоритми от тип EM за статистическо оценяване в разклоняващи се стохастични процеси"

от Нина Руменова Даскалова

представен за присъждане на образователната и научна степен "доктор" по научната специалност 01.01.10 "Теория на вероятностите и математическа статистика", професионално направление 4.5 Математика.

Биографична справка: Н. Даскалова е завършила Факултет по математика и информатика, СУ "Св. Климент Охридски" през 2002г., със специализация „Приложна статистика“ към катедра Вероятности, операционни изследвания и статистика.

Общо описание на дисертацията: Дисертационният труд е изложен на 89 страници и включва увод, четири глави, литература, която съдържа 159 заглавия, и две приложения.

Основните цели на представения труд са разработване на алгоритми от тип EM за максимално правдоподобно оценяване (МПО) на параметрите на разпределението на многотипов разклоняващ се случаен процес (МТРП) при непълни данни, тяхната програмна реализация на езика R и изследване поведението на МПО върху крайни извадки с помощта на симулации за конкретни модели на МТРП.

В глава 1 са дадени основни дефиниции и резултати от теорията разклоняващите се случайни процеси, стохастичните контекстно-свободни граматика, EM алгоритъма за максимизиране на функцията на правдоподобие при непълни данни, направен е кратък литературен преглед на използването разклоняващи се случайни процеси за моделирането на реални процеси, възникващи в естествените науки и медицината.

В глава 2 е разгледана връзката на МТРП със стохастичните контекстно-свободни граматика (СКСГ), изложен е inside-outside алгоритъма (алгоритъм на динамично оптимизиране), широко използван в компютърна лингвистика, за оценяване параметрите на СКСГ. Изложението на inside-outside алгоритъма следва Durbin et al. (1998). Принос в научно отношение представлява анализа и интерпретацията на разгледаните примерите в параграф 2.3 и програмната реализация на inside-outside алгоритъма на R. Този извод се отнася също така за статията на Daskalova (2010), тъй като в началото на тази глава е упоменато, че получените резултати са публикувани.

В глава 3 са изложени основните резултати в дисертационен труд. Тези резултати са сходни с основни резултати на СКСГ, например: (1) твърдение 3.1.1; (2) inside-outside алгоритъма е тип EM алгоритъм и (3) рекурсивните оценки за пресмятане на външните и вътрешни вероятностите в inside-outside алгоритъма; източници за това са Manning and Schutze (1999), Chi and Geman (1998), Eisner, Lafferty (2000), Huang et al. (2001), Prescher (2001, 2002).

Глава 4 е посветена на анализа на данни за модел на клетъчна пролиферация, на основата на алгоритмите и програмната им реализация, разгледани в предходните глави.

В приложение 1 и 2 са описани процедури на езика R и дадени резултати от симулационните експерименти.

Оформление и изложение: Дисертацията е добре оформена. Авторката е усвоила основни методи и алгоритми на статистиката и компютърна лингвистика за анализ на данни при непълна информация, изследванията в тази област могат да продължат с моделиране на съответните параметри на МТРП по предикторни променливи.

Авторефератът на дисертацията отразява изложението в дисертацията.

Публикациите свързани с дисертацията са 3, които са самостоятелни. Не съм забелязал цитирания. Считаю, че основен принос в научно отношение в тази дисертация представлява анализа и моделирането на клетъчна пролиферация по експериментални данни, изследване на поведението на МПО върху крайната извадка от симулационните експерименти и създаването на библиотека от процедури на езика R за моделиране на МТРП.

Използвана научна литература при изготвяне на становището:

Chi, Z. and Geman, S., (1998). Estimation of Probabilistic Context-Free Grammars. *Computational Linguistics*, 24, 299-305.

Prescher, D. (2001). Inside-outside estimation meets dynamic EM. In: *Proceedings of the 7th International Workshop on Parsing Technologies (IWPT-01)*, Beijing, China. <http://www.snlp.de/prescher/papers/>

Prescher, D. (2002). A Tutorial on the Expectation-Maximization Algorithm Including Maximum-Likelihood Estimation and EM Training of Probabilistic Context-Free Grammars. *15th European Summer School in Logic, Language and Information (ESSLLI-03)*, 49 pages. Vienna, Austria. <http://www.snlp.de/prescher/papers/>

Durbin, R., Eddy, S. R., Krogh, A., and Mitchison, G. (1998). Biological sequence analysis.

Huang, X., Acero, A. and Hon, H-W. (2001). Spoken language processing: Guide to algorithms and system development. Prentice Hall.

Manning, C. and Schütze, H. (1999) Foundations of Statistical Natural Language Processing, MIT Press. Cambridge, MA

Lafferty, J.D. (2000) A Derivation of the Inside-Outside Algorithm from the EM Algorithm.

Eisner, J. Notes on the inside-outside algorithm. <http://www.cs.jhu.edu/~jason/465/iobasics.pdf>

Заклучение. Считаю, че представеният дисертационен труд отговаря на изискванията на ЗНСНЗ за образователната и научната степен "Доктор".

Дата: 17.08.2012г.
гр. София

Подпис: