doi:10.1631/FITEE.1400352

题目：大规模文本数据的主题建模

目的：研究大规模数据的主题模型在线推理算法，针对随机变分推理算法中随机梯度误差较大的问题，提出一种移动平均随机变分推理算法。

创新点：使用多次迭代的随机梯度移动平均值近似代替真实随机梯度，以此减小随机梯度和真实梯度间的误差。

方法：以主题模型的基础模型潜在狄利克雷分配为载体展开研究。考虑不同次迭代的文本子集具有不同的词汇（表 1），使用不同次迭代的随机项移动平均值近似代替真实随机梯度的随机项。为尽可能保证算法的精度，使用最近 R 次迭代的随机项（图 2）并验证所提算法的收敛性。

结论：在随机变分推理算法基础上，提出一种移动平均随机变分推理算法，实现更好的文本主题建模效果和更快的收敛速度。

关键词：潜在狄利克雷分配；主题模型；在线学习；移动平均值