

持之以恒

Every thing that has a beginning has an end.

【转】LDA必读的资料

Posted on 2012 年 3 月 27 日 by 恒

2012@4@15添加，一个大牛写的介绍，貌似需翻墙

<http://tedunderwood.wordpress.com/2012/04/07/topic-modeling-made-just-simple-enough/>

2012@4@18添加David M.Blei主

页：<http://www.cs.princeton.edu/~blei/publications.html>，上面有布雷最新的文章：[Introduction to probabilistic topic models](#)

以下内容来自网络，但是作者已经不可考啦，抱歉没法找到原始引用

关于LDA并行化：

那么若利用MapReduce实现，怎样的近似方法好呢？

斯坦福的ScalaNLP项目值得一看：

[http://nlp.stanford.edu/javanlp/scala/scaladoc/scalanlp/cluster/DistributedGibbsLDA\\$object.html](http://nlp.stanford.edu/javanlp/scala/scaladoc/scalanlp/cluster/DistributedGibbsLDA$object.html)

另外还有NIPS2007的论文：

Distributed Inference for Latent Dirichlet Allocation

http://books.nips.cc/papers/files/nips20/NIPS2007_0672

ICML2008的论文：

Fully Distributed EM for Very Large

Datasets<http://www.cs.berkeley.edu/~jawolfe/pubs/08-icml-em>

(3) D. M. Blei, et al., "Hierarchical Topic Models and the Nested Chinese Restaurant Process," NIPS, 2003.

(4) Blei的LDA视频教程: http://videlectures.net/mlss09uk_blei_tm/

(5) Teh的关于Dirichlet Processes的视频教程:
http://videlectures.net/mlss07_teh_dp/

(6) Blei的毕业论文: <http://www.cs.princeton.edu/~blei/papers/Blei2004.pdf>

(7) Jordan的报告: http://www.icms.org.uk/downloads/mixtures/jordan_talk.pdf

(8) G. Heinrich, "Parameter Estimation for Text Analysis,"
<http://www.arbylon.net/publications/text-est.pdf>

基础知识:

(1) P. Johnson and M. Beverlin, "Beta Distribution,"
<http://pj.freefaculty.org/ps707/Distributions/Beta.pdf>

(2) M. Beverlin and P. Johnson, "The Dirichlet Family,"
<http://pj.freefaculty.org/stat/Distributions/Dirichlet.pdf>

(3) P. Johnson, "Conjugate Prior and Mixture Distributions",
<http://pj.freefaculty.org/stat/TimeSeries/ConjugateDistributions.pdf>

(4) P.J. Green, "Colouring and Breaking Sticks: Random Distributions and Heterogeneous Clustering",
<http://www.maths.bris.ac.uk/~mapjg/papers/GreenCDP.pdf>

(5) Y. W. Teh, "Dirichlet Process",
<http://www.gatsby.ucl.ac.uk/~ywtteh/research/npbayes/dp.pdf>

(6) Y W Teh and M I Jordan "Hierarchical Bayesian Nonparametric Models

dirichlet.pdf

(8) 北邮论坛的LDA导读: [导读]文本处理、图像标注中的一篇重要论文Latent Dirichlet Allocation, http://bbs.byr.edu.cn/article/PR_AI/2530?p=1

(9) Zhou Li的LDA Note: [http://lsa-lda.googlecode.com/files/Latent Dirichlet Allocation note.pdf](http://lsa-lda.googlecode.com/files/Latent%20Dirichlet%20Allocation%20note.pdf)

(10) C. M. Bishop, "Pattern Recognition And Machine Learning," Springer, 2006.

代码:

(1) Blei的LDA代码 (C): <http://www.cs.princeton.edu/~blei/lda-c/index.html>

(2) BLEI的HLDA代码 (C): <http://www.cs.princeton.edu/~blei/downloads/hlda-c.tgz>

(3) Gibbs LDA (C++): <http://gibbslda.sourceforge.net/>

(4) Delta LDA (Python):
<http://pages.cs.wisc.edu/~andrzej/research/deltaLDA.tgz>

(5) Griffiths和Steinberger的Topic Modeling工具箱:
http://psiexp.ss.uci.edu/research/programs_data/toolbox.htm

(6) LDA (Java): <http://www.arbylon.net/projects/>

(7) Mochihashi的LDA (C, Matlab): <http://chasen.org/~daiti-m/dist/lda/>

(8) Chua的LDA (C#):
<http://www.mysmu.edu/phdis2009/freddy.chua.2009/programs/lda.zip>

(9) Chua的HLDA (C#):
<http://www.mysmu.edu/phdis2009/freddy.chua.2009/programs/hlda.zip>

(2) B. C. Russell, et al., "Using Multiple Segmentations to Discover Objects and their Extent in Image Collections," in Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, 2006, pp. 1605-1614.

(3) J. Sivic, et al., "Discovering objects and their location in images," in Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on, 2005, pp. 370-377 Vol. 1.

(4) F. C. T. Chua, "Summarizing Amazon Reviews using Hierarchical Clustering,"
<http://www.mysmu.edu/phdis2009/freddy.chua.2009/papers/amazon.pdf>

(5) F. C. T. Chua, "Dimensionality Reduction and Clustering of Text Documents,"
<http://www.mysmu.edu/phdis2009/freddy.chua.2009/papers/probabilisticIR.pdf>

(6) D Bacciu, "Probabilistic Generative Models for Machine Vision,"
http://www.math.unipd.it/~sperduti/AI09/bacciu_unipd_handouts.pdf

This entry was posted in [Academics](#) and tagged [LDA](#) by [恒](#). Bookmark the [permalink](#) [<http://www.xperseverance.net/blogs/2012/03/657/>].

3 THOUGHTS ON “【转】LDA必读的资料”



drew meyers

on 2012 年 11 月 30 日 at 下午 5:20 said:

It's a shame you don't have a donate button! I'd without a doubt donate to this brilliant blog! I guess for now i'll settle for bookmarking and adding your RSS feed to my Google account. I look forward to new updates and will share this site with my Facebook group.

Chat soon!

