









































































































































| Gibbs sampling | | | | |
|----------------|-------------|--------|------------------------------------|----|
| | | | iteration 1 | |
| i | W_i | d_i | Z_i | |
| 1 | MATHEMATICS | 1 | 2 | |
| 2 | KNOWLEDGE | 1 | $\frac{1}{2}$ | |
| 3 | RESEARCH | 1 | 1 | |
| 4 | WORK | 1 | 2 | |
| 5 | MATHEMATICS | 1 | 1 | |
| 6 | RESEARCH | 1 | 2 | |
| 7 | WORK | 1 | 2 | |
| 8 | SCIENTIFIC | 1 | 1 | |
| 9 | MATHEMATICS | 1 | 2 | |
| 10 | WORK | 1 | 1 | |
| 11 12 | SCIENTIFIC | 2 2 | 1 | |
| | KNOWLEDGE | - | 1 | |
| | • | • | | |
| | • | | · | |
| 50 | JOY | 5 | 2 | |
| | | | | |
| | | C | Eric Xing @ CMU, ACL Tutorial 2012 | 70 |
















































































Automatic annotation



True caption birds tree Corr-LDA birds nest leaves branch tree

GM–LDA water birds nest tree sky

GM–Mixture tree ocean fungus mushrooms coral



True caption fish reefs water Corr-LDA

fish water ocean tree coral

GM–LDA water sky vegetables tree people

GM–Mixture fungus mushrooms tree flowers leaves

© Eric Xing @ CMU, ACL Tutorial 2012











































































































































































































































































| Expe | | | | | | |
|-------|---|----------|------------|----------------|-----------------|--|
| • Sm | ing data Iall: Treebank 3 Ige: FBIS-Beijin | | · · | , | lish words). | |
| | Train | #Doc. | #Sent. | #To English | kens Chinese | |
| | Treebank | 316 | 4172 | 133K | 105K | |
| | FBIS.BJ | 6,111 | 105K | 4.18M | 3.54M | |
| | Sinorama | 2,373 | 103K | 3.81M | 3.60M | |
| | XinHua | 19,140 | 115K | 3.85M | 3.93M | |
| | FOUO | 15,478 | 368K | 13.14M | 11.93M | |
| | Test | 95 | 627 | 25,500 | 19,726 | |
| Word | Alignment | Accuracy | / & I rans | ation Qua | ality | |
| ● F-n | neasure | | | | - | |
| • BLI | EU | | | | | |
| | | | | | | |
| | | | | | | |
| | 246 | | | | | |



| 0 | pics | | |
|----|---|----|--|
| T1 | Teams, sports, disabled, games members, people, cause, water, national, handicapped | T1 | 人, 残疾, 体育, 事业, 水, 世界, 区, 新华社, 队员, 记者 |
| T2 | Shenzhen, singapore, hongkong, stock, national, investment, yuan, options, million, dollar | T2 | 深圳,深,新,元,有,股,香港,国有 外资,新华社 |
| Т3 | Chongqing, company, takeover, shenzhen, tianjin, city, national, government, project, companies | ТЗ | 国家, 重庆, 市, 区, 厂, 天津, 政府 项目, 国, 深圳 |
| T4 | Hongkong, trade, export, import, foreign, tech., high, 1998, year, technology | Τ4 | 香港, 贸易, 出口, 外资, 合作, 今年, 项目, 利用, 新, 技术 |
| Т5 | House, construction, government, employee, living, provinces, macau, anhui, yuan | Т5 | 住房,房,九江,建设,澳门,元,职 工,目前,国家,占,省 |
| Т6 | Gas, company, energy, usa, russia, france, chongqing, resource, china, economy, oil | Т6 | 公司, 天然气, 两, 国, 美国, 记者, 关系, 俄, 法, 重庆 |





| Systems 1-gram 2-gram 3-gram 4-gram BLE Hiero Sys. 73.92 40.57 23.21 13.84 30 Gale Sys. 75.63 42.71 25.00 14.30 32 |
|--|
| |
| |
| HM-BITAM 76.77 42.99 25.42 14.04 33 |
| Ground Truth 76.10 43.85 26.70 15.73 34 |
| found fruth 70.10 45.85 20.70 15.75 54 |





























| Conclusion | |
|--|-----|
| GM-based topic models are cool Flexible Modular Interactive | |
| There are many ways of implementing topic models unsupervised supervised | |
| Efficient Inference/learning algorithms GMF, with Laplace approx. for non-conjugate dist. MCMC | |
| Many applications Word-sense disambiguation Image understanding | |
| Network inference © Eric Xing @ CMU, ACL Tutorial 2012 | 266 |

