The Poisson Gamma Belief Network

Mingyuan Zhou[#], Yulai Cong[*], and Bo Chen[*]

[#] IROM Department, The University of Texas at Austin, Austin, TX, USA

[*] School of Electronic Engineering, Xidian University, Xi'an, Shaanxi, China

THE UNIVERSITY OF TEXAS AT AUSTIN McCOMBS SCHOOL OF BUSINESS

西安电子科技大学 XIDIAN UNIVERSITY

Introduction

The Poisson gamma belief network (PGBN) is proposed to infer a multilayer representation of high-dimensional count vectors.

- The PGBN factorizes each of its layers into the product of a connection weight matrix and the nonnegative real hidden units of the next layer.
- The PGBN's hidden layers are jointly trained with an upward-downward Gibbs sampler.
- The gamma-negative binomial process combined with a layer-wise training strategy allows the PGBN to infer the width of each layer given a fixed budget on the width of the first layer.
- Example results illustrate interesting relationships between the width of the first layer and the inferred network structure, and demonstrate that the PGBN can add more layers to increase its performance gains over Poisson factor analysis.

Hierarchical Model and Properties

Poisson Gamma Belief Network (PGBN)

$$\boldsymbol{\theta}_j^{(T)} \sim \mathrm{Gam}\left(\boldsymbol{r}, 1/c_j^{(T+1)}\right),$$
$$\cdots$$
$$\boldsymbol{\theta}_j^{(t)} \sim \mathrm{Gam}\left(\boldsymbol{\Phi}^{(t+1)}\boldsymbol{\theta}_j^{(t+1)}, 1/c_j^{(t+1)}\right),$$
$$\cdots$$
$$\boldsymbol{x}_j^{(1)} \sim \mathrm{Pois}\left(\boldsymbol{\Phi}^{(1)}\boldsymbol{\theta}_j^{(1)}\right),\quad \boldsymbol{\theta}_j^{(1)} \sim \mathrm{Gam}\left(\boldsymbol{\Phi}^{(2)}\boldsymbol{\theta}_j^{(2)}, p_j^{(2)}/\left(1-p_j^{(2)}\right)\right).$$

PGBN vs Sigmoid belief network (SBN)

$$P\left(\boldsymbol{x}_j^{(1)}, \{\boldsymbol{\theta}_j^{(t)}\}_t \,\middle|\, \{\boldsymbol{\Phi}^{(t)}\}_t\right) = P\left(\boldsymbol{x}_j^{(1)} \,\middle|\, \boldsymbol{\Phi}^{(1)}, \boldsymbol{\theta}_j^{(1)}\right)\left[\prod_{t=1}^{T-1} P\left(\boldsymbol{\theta}_j^{(t)} \,\middle|\, \boldsymbol{\Phi}^{(t+1)}, \boldsymbol{\theta}_j^{(t+1)}\right)\right] P\left(\boldsymbol{\theta}_j^{(T)}\right)$$

PGBN: $$P\left(\theta_{vj}^{(t)} \,\middle|\, \boldsymbol{\phi}_{v:}^{(t+1)}, \boldsymbol{\theta}_j^{(t+1)}, c_{j+1}^{(t+1)}\right) = \frac{\left(c_{j+1}^{(t+1)}\right)^{\boldsymbol{\phi}_{v:}^{(t+1)}\boldsymbol{\theta}_j^{(t+1)}}}{\Gamma\left(\boldsymbol{\phi}_{v:}^{(t+1)}\boldsymbol{\theta}_j^{(t+1)}\right)}\left(\theta_{vj}^{(t)}\right)^{\boldsymbol{\phi}_{v:}^{(t+1)}\boldsymbol{\theta}_j^{(t+1)}-1} e^{-c_{j+1}^{(t+1)}\theta_{vj}^{(t)}}$$

SBN: $$P\left(\theta_{vj}^{(t)} = 1 \,\middle|\, \boldsymbol{\phi}_{v:}^{(t+1)}, \boldsymbol{\theta}_j^{(t+1)}, b_v^{(t+1)}\right) = \sigma\left(b_v^{(t+1)} + \boldsymbol{\phi}_{v:}^{(t+1)}\boldsymbol{\theta}_j^{(t+1)}\right)$$

Properties of PGBN:

Lemma 1 (Augment-and-conquer the PGBN). *With $p_j^{(1)} := 1 - e^{-1}$ and*

$$p_j^{(t+1)} := -\ln(1-p_j^{(t)}) \Big/ \left[c_j^{(t+1)} - \ln(1-p_j^{(t)})\right]$$

for $t = 1, \ldots, T$, one may connect the observed (if $t = 1$) or some latent (if $t \geq 2$) counts $\boldsymbol{x}_j^{(t)} \in \mathbb{Z}^{K_{t-1}}$ to the product $\boldsymbol{\Phi}^{(t)}\boldsymbol{\theta}_j^{(t)}$ at layer t under the Poisson likelihood as

$$\boldsymbol{x}_j^{(t)} \sim \mathrm{Pois}\left[-\boldsymbol{\Phi}^{(t)}\boldsymbol{\theta}_j^{(t)} \ln\left(1-p_j^{(t)}\right)\right].$$

Corollary 2. With $m_{kj}^{(t)(t+1)} := x_{\cdot jk}^{(t)} := \sum_{v=1}^{K_{t-1}} x_{vjk}^{(t)}$, we can propagate the latent counts $x_{vj}^{(t)}$ of layer t upward to layer $t+1$ as

$$\left\{\left(x_{vj1}^{(t)}, \ldots, x_{vjK_t}^{(t)}\right) \,\middle|\, x_{vj}^{(t)}, \boldsymbol{\phi}_{v:}^{(t)}, \boldsymbol{\theta}_j^{(t)}\right\} \sim \mathrm{Mult}\left(x_{vj}^{(t)}, \frac{\phi_{v1}^{(t)}\theta_{1j}^{(t)}}{\sum_{k=1}^{K_t}\phi_{vk}^{(t)}\theta_{kj}^{(t)}}, \ldots, \frac{\phi_{vK_t}^{(t)}\theta_{K_tj}^{(t)}}{\sum_{k=1}^{K_t}\phi_{vk}^{(t)}\theta_{kj}^{(t)}}\right)$$

$$\left(x_{kj}^{(t+1)} \,\middle|\, m_{kj}^{(t)(t+1)}, \boldsymbol{\phi}_{k:}^{(t+1)}, \boldsymbol{\theta}_j^{(t+1)}\right) \sim \mathrm{CRT}\left(m_{kj}^{(t)(t+1)}, \boldsymbol{\phi}_{k:}^{(t+1)}\boldsymbol{\theta}_j^{(t+1)}\right)$$

Projecting and ranking latent factors:

$$\mathbb{E}\left[\boldsymbol{x}_j^{(1)} \,\middle|\, \boldsymbol{\theta}_j^{(t)}, \{\boldsymbol{\Phi}^{(\ell)}, c_j^{(\ell)}\}_{1,t}\right] = \left[\prod_{\ell=1}^{t}\boldsymbol{\Phi}^{(\ell)}\right]\frac{\boldsymbol{\theta}_j^{(t)}}{\prod_{\ell=2}^{t} c_j^{(\ell)}}$$

$$\mathbb{E}\left[\boldsymbol{\theta}_j^{(t)} \,\middle|\, \{\boldsymbol{\Phi}^{(\ell)}, c_j^{(\ell)}\}_{t+1,T}, \boldsymbol{r}\right] = \left[\prod_{\ell=t+1}^{T}\boldsymbol{\Phi}^{(\ell)}\right]\frac{\boldsymbol{r}}{\prod_{\ell=t+1}^{T+1} c_j^{(\ell)}}$$

Projection of the factors of layer *t* :

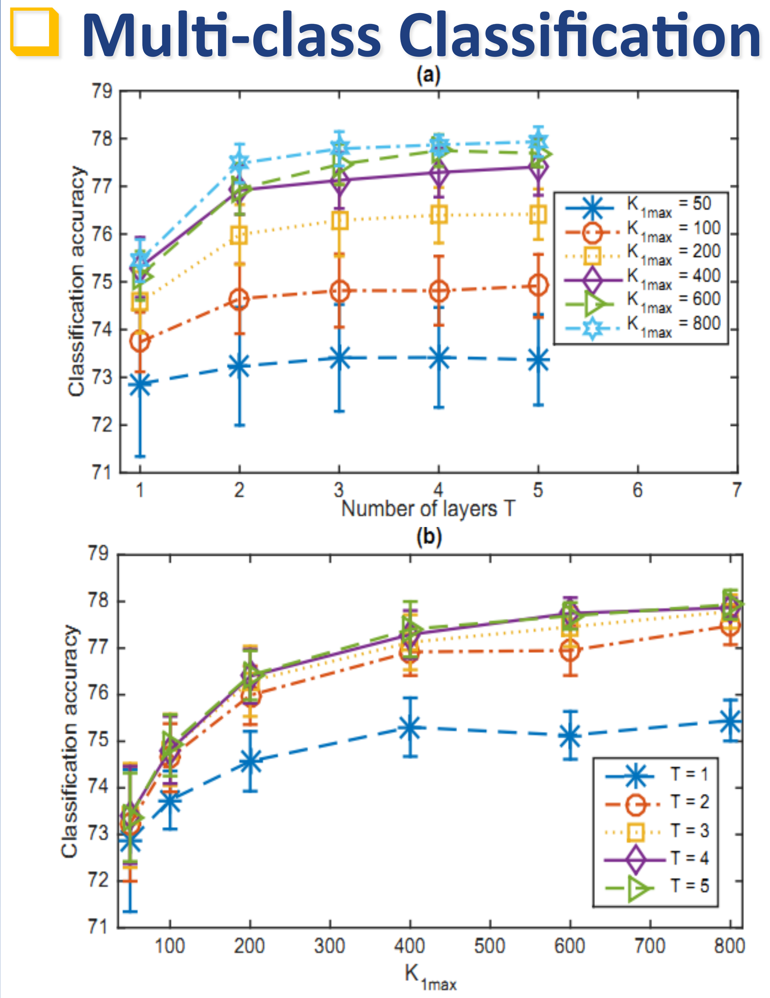
$$\prod_{\ell=1}^{t}\boldsymbol{\Phi}^{(\ell)}$$

Weights of the factors of layer *t* :

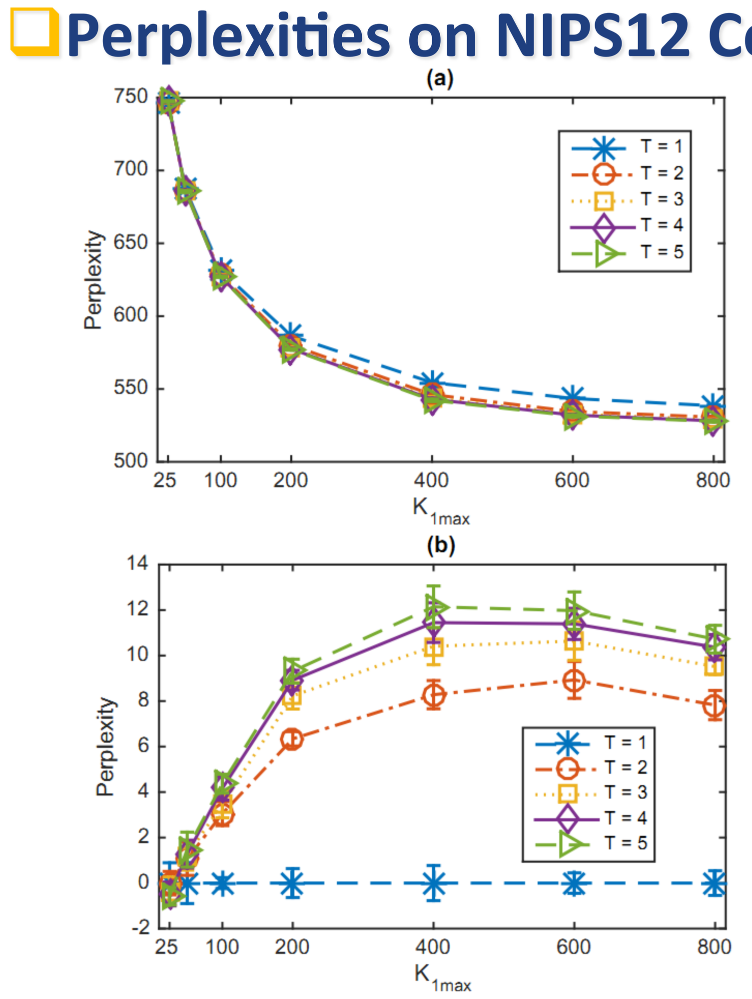
$$\boldsymbol{r}^{(t)} := \left[\prod_{\ell=t+1}^{T}\boldsymbol{\Phi}^{(\ell)}\right]\boldsymbol{r}$$

```
 1: for T = 1, 2, ..., T_max do Jointly train all the T layers of the network
 2:     Set K_{T-1}, the inferred width of layer T-1, as K_{T max}, the upper bound of layer T's width.
 3:     for iter = 1 : B_T + C_T do Upward-downward Gibbs sampling
 4:         Sample {z_ji}_{j,i} using collapsed inference; Calculate {x^(1)_{vjk}}_{v,k,j}; Sample {x^(2)_{vj}}_{v,j} ;
 5:         for t = 2, 3, ..., T do
 6:             Sample {x^(t)_{vjk}}_{v,j,k} ; Sample {φ^(t)_k}_k ; Sample {x^(t+1)_{vj}}_{v,j} ;
 7:         end for
 8:         for t = T, T-1, ..., 2 do
 9:             Sample c^(t+1)_j and calculate p^(t+1)_j ; Sample r if t = T; Sample {θ^(t)_j}_j ;
10:         end for
11:         Sample p^(2)_j and Calculate c^(2)_j ;
12:         if iter = B_T then
13:             Prune layer T's inactive factors {φ^(T)_k}_{k: x^(T)_{··k}=0}, let K_T = Σ_k δ(x^(T)_{··k} > 0), and update r;
14:         end if
15:     end for
16:     Output the posterior means (according to the last MCMC sample) of all remaining factors {φ^(t)_k}_{k,t} as
        the inferred network of T layers, and {r_k}_{k=1}^{K_T} as the gamma shape parameters of layer T's hidden units.
17: end for
```

Multi-class Classification



Perplexities on NIPS12 Corpus



Topics of Layer One on 20newsgroups

1 doesn new problem work probably let said years ll long question course	11 team game hockey season games nhl year league teams players play cup	2 michael netcom andrew new uucp internet mark steve opinions mike mail net	12 sale new shipping offer condition price asking sell cover cd interested best	3 thanks mail email help fax advance looking hi info information phone send	13 power output input high signal low voltage chip radio battery circuit data	4 god jesus christian bible christ christians life faith believe man christianity church	14 law government rights state court laws states public case civil legal federal	5 mac apple mhz modem ram bit card speed port board simms memory	15 bike dod ride bikes riding motorcycle sun left road bnr ama honda	6 available software version ftp server file sun program unix mit information code	16 clipper government chip key encryption nsa phone crypto secure keys netcom public		

Topics of Layer Three

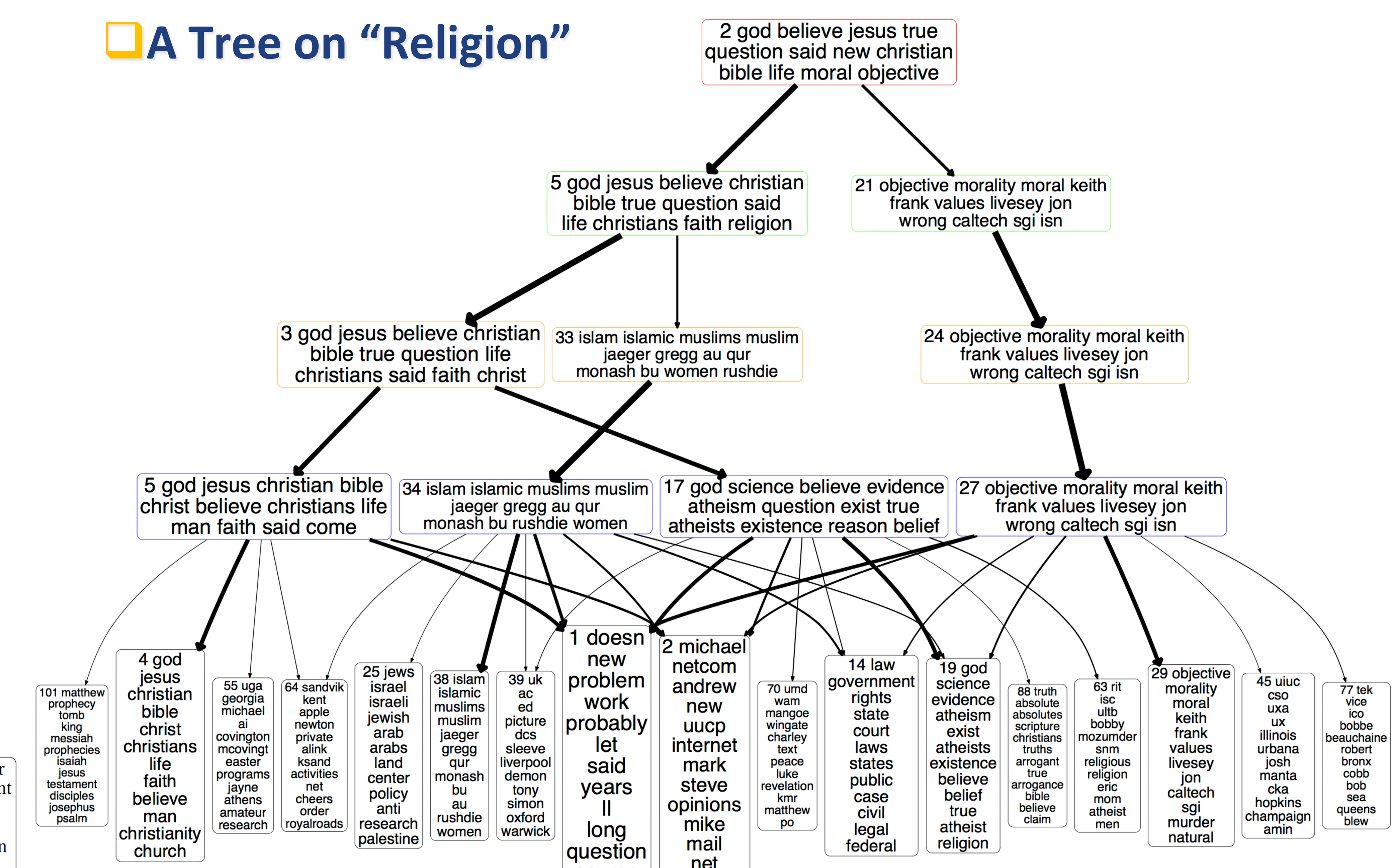
1 thanks mail email information help fax software info advance looking new hi	2 mac apple thanks bit mhz card modem ram speed port board memory	3 god jesus believe christian bible true question life christians said faith christ	4 car cars new engine miles dealer price drive year road buy speed	5 windows dos file files thanks problem program using run running win ms
6 team game hockey season games nhl year play league new players teams	7 year game team baseball runs games season players hit win league years	8 god jesus christian bible christ believe christians church life faith man said	9 bike dod ride bikes riding motorcycle got sun left new said ll	10 window thanks widget application using display available program mail server set software
11 power thanks mail output input high work line using low phone circuit	12 government key encryption clipper chip nsa phone public keys secure crypto escrow	13 gun guns firearms crime control law weapons state new self government said	14 disease doctor pain new help treatment medical years said problem thanks patients	15 card monitor video vga windows thanks drivers cards mode mail graphics ati

Topics of Layer Five

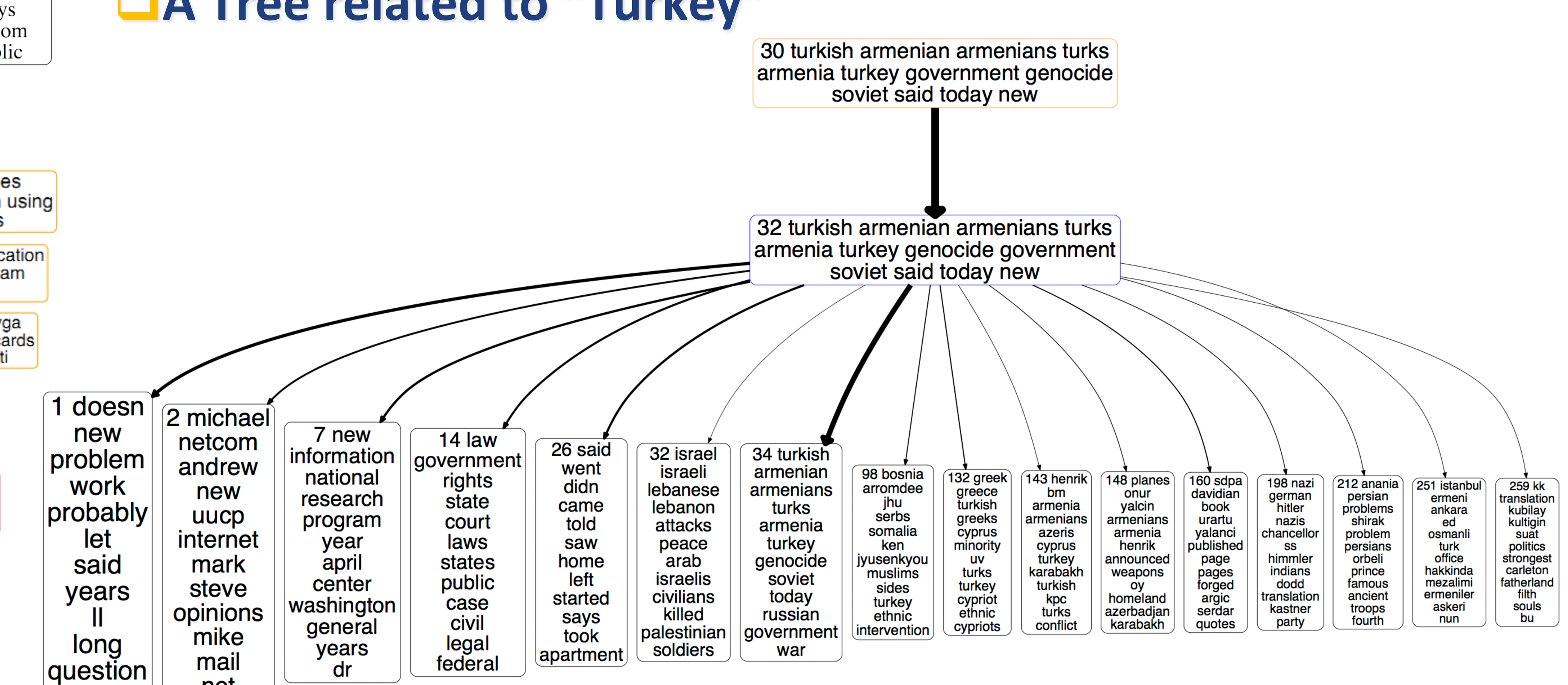
1 windows thanks file dos window using mail program problem help files work	2 god believe jesus true question said new christian bible life moral objective	3 car bike new cars dod engine got thanks price road ll miles	4 thanks mail email help information fax software info new advance looking hi	5 card thanks new mac apple mail drive bit work video mb problem
6 space nasa new gov work years long earth said orbit ll year	7 team game hockey season games nhl year play new league players teams	8 god jesus christian bible believe christ christians life new church said man	9 key chip government clipper encryption keys nsa phone public escrow secure new	10 card thanks mac windows apple mail drive bit problem work mb new
11 car bike new cars dod engine got road said ll miles ride	12 thanks new mail power email work phone high help sale price using	13 year game team baseball games runs season players hit win league years	14 disease new doctor pain help treatment said years thanks problem medical work	15 israel jews israeli armenian turkish new state government said armenians years law

Example Results

A Tree on "Religion"



A Tree related to "Turkey"



NIPS 2015