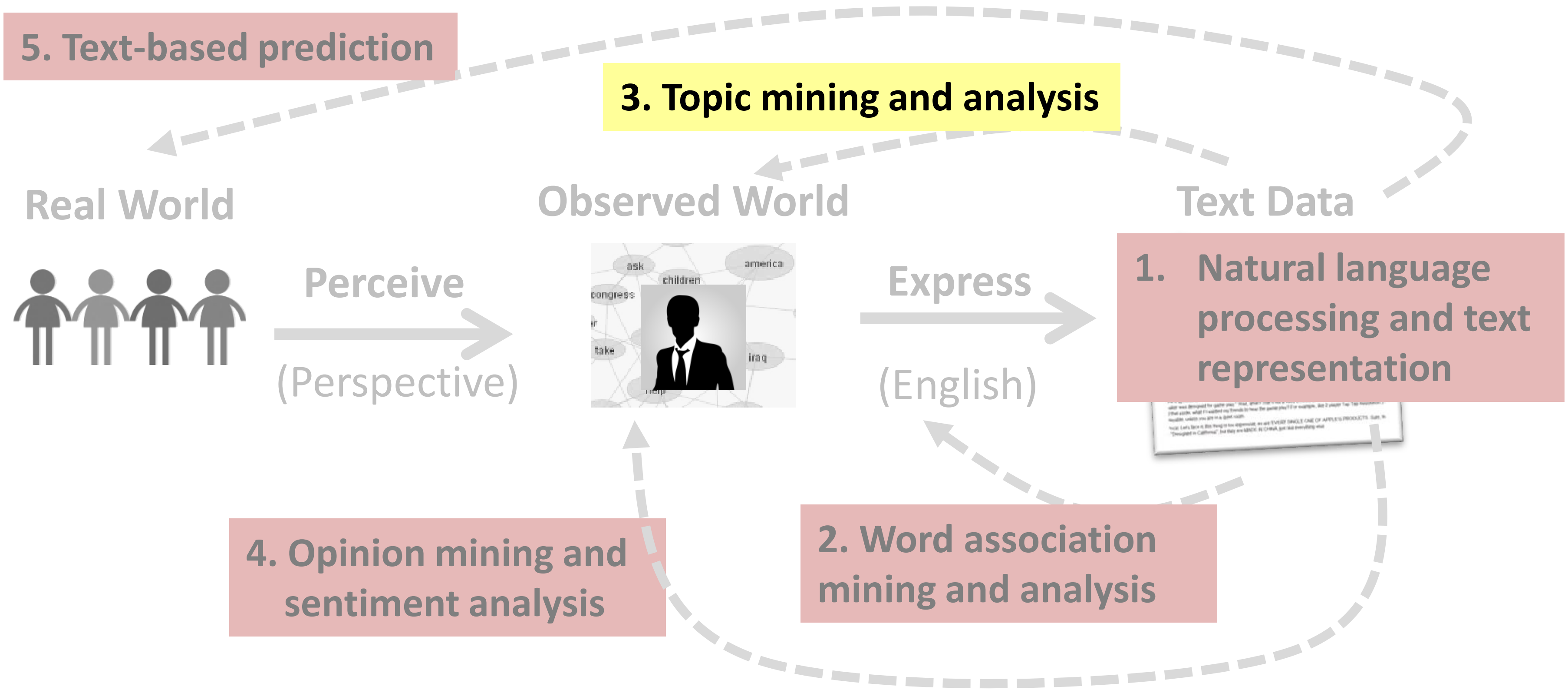
Probabilistic Latent Semantic Analysis (PLSA)

ChengXiang “Cheng” Zhai
Department of Computer Science
University of Illinois at Urbana-Champaign

Probabilistic Latent Semantic Analysis (PLSA)



Document as a Sample of Mixed Topics

Topic θ_1

government 0.3
response 0.2
...

Topic θ_2

city 0.2
new 0.1
orleans 0.05
...

. . .

Topic θ_k

donate 0.1
relief 0.05
help 0.02
...

Background θ_B

the 0.04
a 0.03
...

Blog article about "Hurricane Katrina"

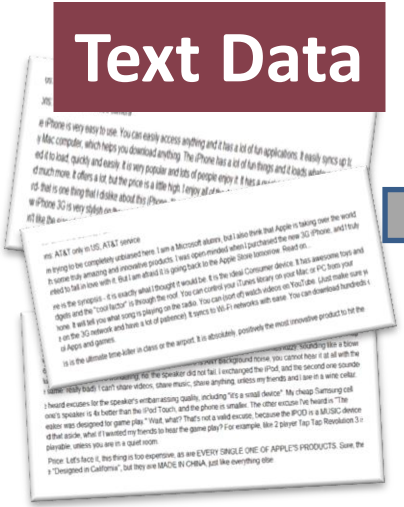
[Criticism of government response to the hurricane primarily consisted of criticism of its response to the approach of the storm and its aftermath, specifically in the delayed response] to the [flooding of New Orleans. … 80% of the 1.3 million residents of the greater New Orleans metropolitan area evacuated] …[Over seventy countries pledged monetary donations or other assistance]. …

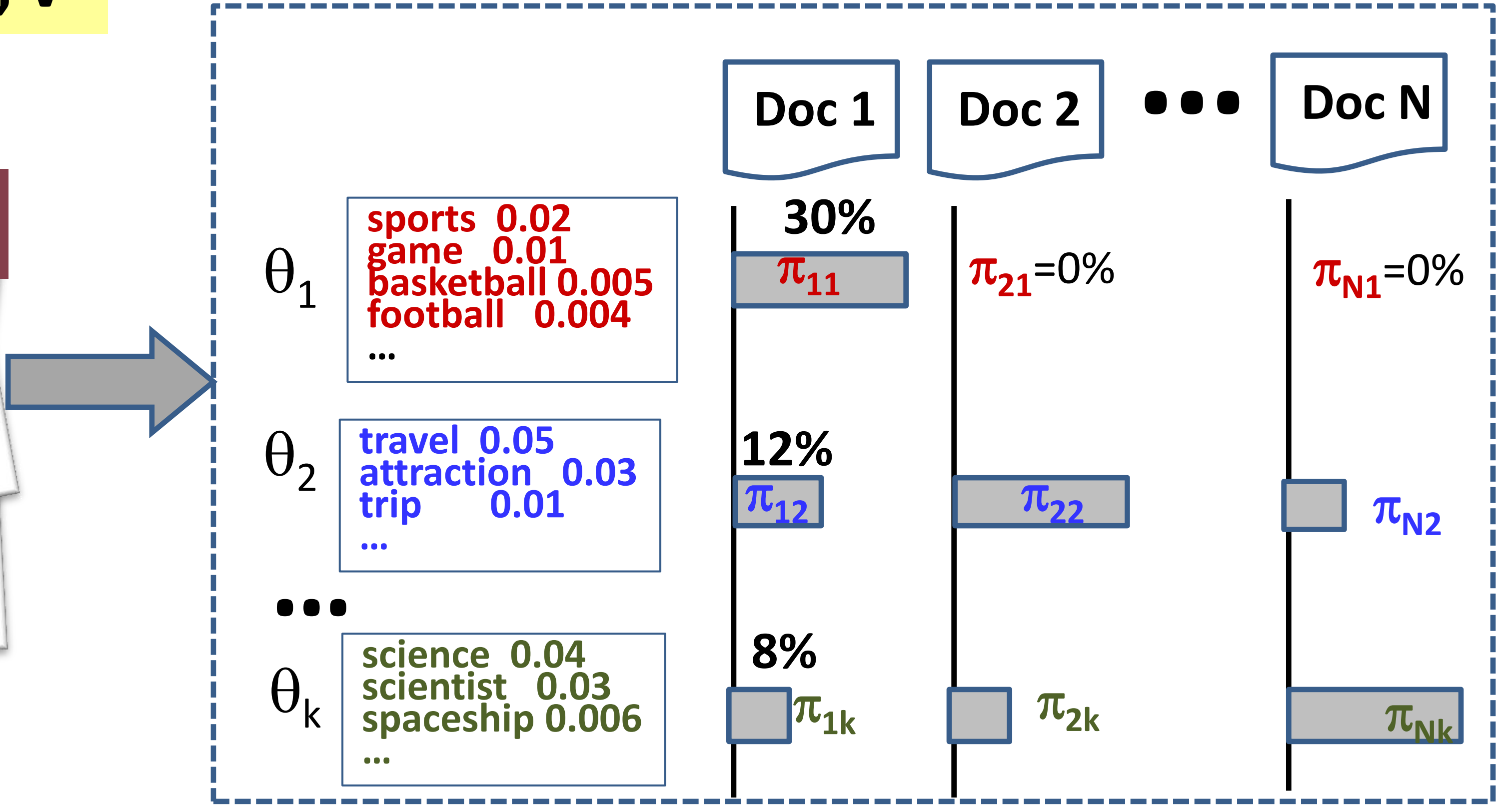
Many applications are possible if we can "decode" the topics in text…

Mining Multiple Topics from Text

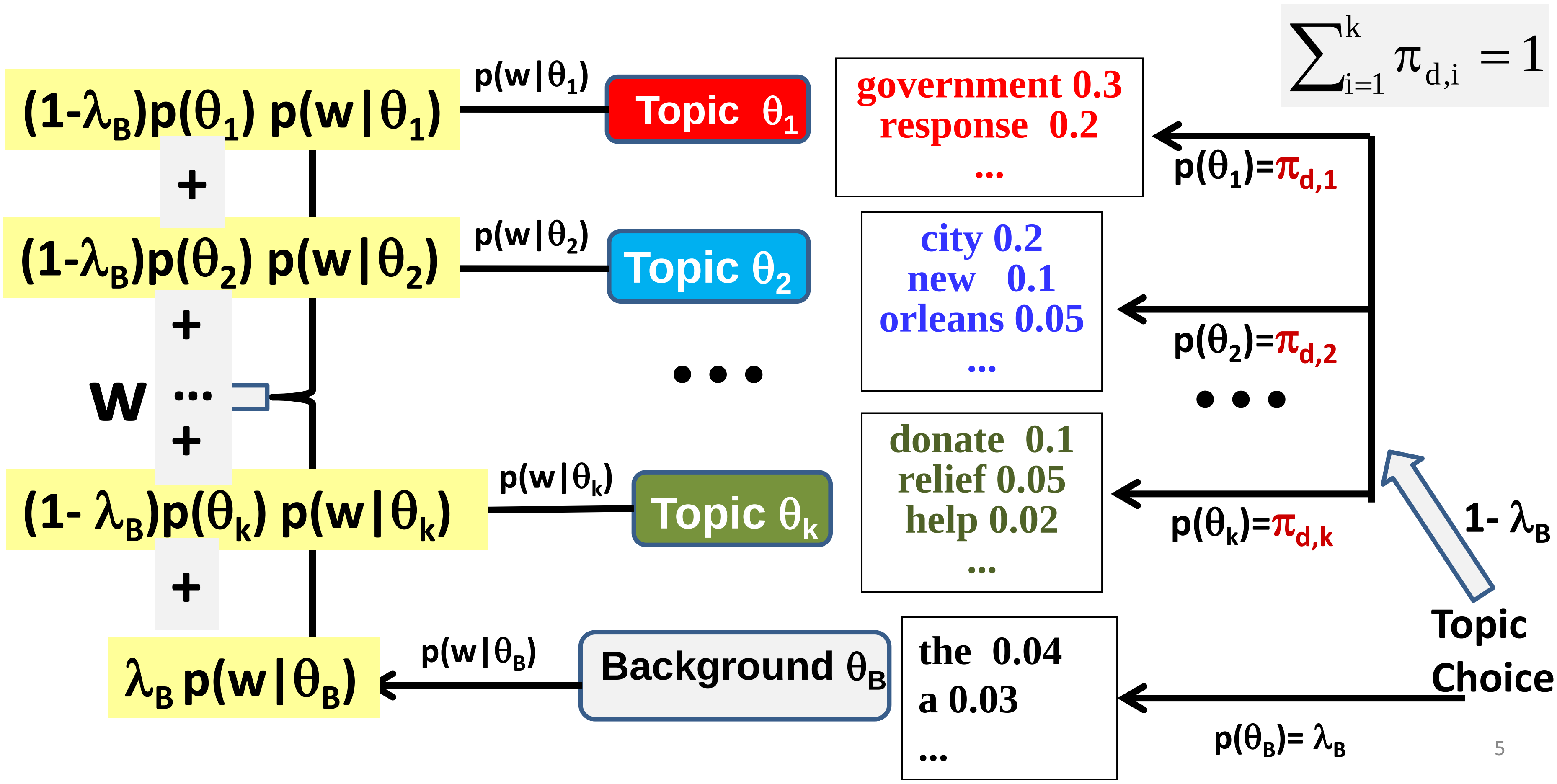
OUTPUT: $\{ \theta_1, \ldots, \theta_k \}, \{ \pi_{i1}, \ldots, \pi_{ik} \}$

INPUT: C, k, V





Generating Text with Multiple Topics: p(w)=?



Probabilistic Latent Semantic Analysis (PLSA)

Percentage of background words (known)

Background LM (known)

Coverage of topic θ $_j$ in doc d

Prob. of word w in topic θ $_j$

$$p_d(w) = \lambda_B p(w|\theta_B) + (1-\lambda_B)\sum_{j=1}^{k} \pi_{d,j}\, p(w|\theta_j)$$

$$\log p(d) = \sum_{w \in V} c(w,d) \log[\lambda_B p(w|\theta_B) + (1-\lambda_B)\sum_{j=1}^{k} \pi_{d,j}\, p(w|\theta_j)]$$

$$\log p(C|\Lambda) = \sum_{d \in C}\sum_{w \in V} c(w,d) \log[\lambda_B p(w|\theta_B) + (1-\lambda_B)\sum_{j=1}^{k} \pi_{d,j}\, p(w|\theta_j)]$$

Unknown Parameters: $\Lambda=(\{\pi_{d,j}\}, \{\theta_j\})$, j=1, …, k

How many unknown parameters are there in total?

ML Parameter Estimation

$$p_d(w) = \lambda_B p(w \mid \theta_B) + (1-\lambda_B)\sum_{j=1}^{k} \pi_{d,j}\, p(w \mid \theta_j)$$

$$\log p(d) = \sum_{w\in V} c(w,d) \log[\lambda_B p(w \mid \theta_B) + (1-\lambda_B)\sum_{j=1}^{k} \pi_{d,j}\, p(w \mid \theta_j)]$$

$$\log p(C \mid \Lambda) = \sum_{d\in C}\sum_{w\in V} c(w,d) \log[\lambda_B p(w \mid \theta_B) + (1-\lambda_B)\sum_{j=1}^{k} \pi_{d,j}\, p(w \mid \theta_j)]$$

Constrained Optimization: $\Lambda^* = \arg\max_{\Lambda} p(C \mid \Lambda)$

$$\forall j \in [1,k], \sum_{i=1}^{M} p(w_i \mid \theta_j) = 1 \qquad \forall d \in C, \sum_{j=1}^{k} \pi_{d,j} = 1$$

EM Algorithm for PLSA: E-Step

Hidden Variable (=topic indicator): $z_{d,w} \in \{B, 1, 2, \ldots, k\}$

Probability that **w in doc d** is generated from **topic θ_j**

Use of Bayes Rule

$$p(z_{d,w} = j) = \frac{\pi_{d,j}^{(n)} p^{(n)}(w \mid \theta_j)}{\sum_{j'=1}^{k} \pi_{d,j'}^{(n)} p^{(n)}(w \mid \theta_{j'})}$$

$$p(z_{d,w} = B) = \frac{\lambda_B p(w \mid \theta_B)}{\lambda_B p(w \mid \theta_B) + (1-\lambda_B) \sum_{j=1}^{k} \pi_{d,j}^{(n)} p^{(n)}(w \mid \theta_j)}$$

Probability that **w in doc d** is generated from **background θ_B**

EM Algorithm for PLSA: M-Step

Hidden Variable (=topic indicator): $z_{d,w} \in \{B, 1, 2, \ldots, k\}$

ML Estimate based on “allocated” word counts to topic θ_j

Re-estimated **probability** of **doc d** covering **topic** θ_j

$$\pi_{d,j}^{(n+1)} = \frac{\sum_{w\in V} \boxed{c(w,d)(1-p(z_{d,w}=B))p(z_{d,w}=j)}}{\sum_{j'}\sum_{w\in V} c(w,d)(1-p(z_{d,w}=B))p(z_{d,w}=j')}$$

$$p^{(n+1)}(w\,|\,\theta_j) = \frac{\sum_{d\in C} \boxed{c(w,d)(1-p(z_{d,w}=B))p(z_{d,w}=j)}}{\sum_{w'\in V}\sum_{d\in C} c(w',d)(1-p(z_{d,w'}=B))p(z_{d,w'}=j)}$$

Re-estimated **probability** of **word w** for **topic** θ_j

Computation of the EM Algorithm

- Initialize all unknown parameters randomly
- Repeat until likelihood converges
  - E-step $p(z_{d,w} = j) \propto \pi_{d,j}^{(n)} p^{(n)}(w \mid \theta_j)$

    $\sum_{j=1}^{k} p(z_{d,w} = j) = 1$

    $p(z_{d,w} = B) \propto \lambda_B p(w \mid \theta_B) \leftarrow$ What's the normalizer for this one?
  - M-step

$$\pi_{d,j}^{(n+1)} \propto \sum_{w \in V} c(w,d)(1 - p(z_{d,w} = B)) p(z_{d,w} = j)$$

$\forall d \in C, \sum_{j=1}^{k} \pi_{d,j} = 1$

$$p^{(n+1)}(w \mid \theta_j) \propto \sum_{d \in C} c(w,d)(1 - p(z_{d,w} = B)) p(z_{d,w} = j)$$

$\forall j \in [1,k], \sum_{w \in V} p(w \mid \theta_j) = 1$

In general, accumulate counts, and then normalize

Summary

- PLSA = mixture model with k unigram LMs (k topics)
- Adding a pre-determined background LM helps discover discriminative topics
- ML estimate “discovers” topical knowledge from text data
  - k word distributions (k topics)
  - proportion of each topic in each document
- The output can enable many applications!
  - Clustering of terms and docs (treat each topic as a cluster)
  - Further associate topics with different contexts (e.g., time periods, locations, authors, sources, etc.)