Leveraging Social Context for Modeling Topic Evolution

Janani Kalyanam*, Amin Mantrach+, Diego Saez-Trumper+, Hossein Vahabi+ and Gert Lanckriet*

*University of California, San Diego, +Yahoo Labs

Text corpora today

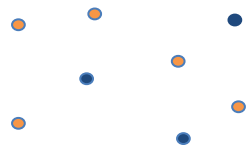
- Social Media posts every instant
- Constantly evolving corpora
- Free user vocabulary
- Volatile

Our idea

- Traditional approaches: Bayesian (like LDA), or NMF-based
- Use textual content
- Our idea is to use side-information

Our goal

When does side information in the form of community help in modeling topic discovery and evolution?



Our approach

X^t is documents-by-words

U^t is documents-by-users

$X^t \approx W^t H^t$ $\quad$ $U^t \approx W^t G^t$

$H^t\ G^t$ each row is a topic/community represented as distribution over words/users

Key assumption

W^t explains each document in terms of the underlying latent topics/community.

Temporal evolution

$X^t \approx W^t \underbrace{M_T^t H^{t-1}}_{H^t}$ $\quad$ H^{t-1} Communities in the previous time step (considered known at time-t)

$U^t \approx W^t \underbrace{M_C^t G^{t-1}}_{G^t}$ $\quad$ M_T^t Evolution matrix

Objective function

$$\underbrace{\|X^t - W^t H^t\|^2 + \|X^t - W^t M_T^t H^{t-1}\|^2}_{L_T} + \underbrace{\|U^t - W^t G^t\|^2 + \|U^t - W^t M_C^t G^{t-1}\|^2}_{L_C}$$

Variables are $W^t\ H^t\ G^t\ M_T^t\ M_C^t$

$$L = \mu L_T + (1-\mu) L_C + R$$

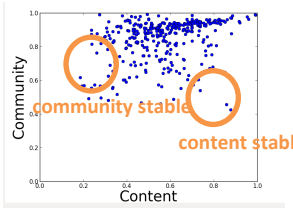
μ importance parameter

R regularization

How to validate?

- Content stable topics; Community stable topics; Mixed stable topics
- Obtain centroids of representations at time-t
- Cosine similarities for representation at time-t and time-(t+1) for all time
- Average all similarities

Types of topics



Experiments

Community Stable

	K = 5	K = 10	K = 15	K = 20
NDCG	0.4081	0.4800	0.5029	0.5129
MAP	0.2653	0.3637	0.4007	0.4173
μ				
NDCG	0.3699	0.4496	0.4608	0.4138
MAP	0.2191	0.3596	0.3462	0.3420

Content Stable

NDCG	0.6888	0.6055	0.6317	0.6623
MAP	0.5655	0.4784	0.5115	0.5559
μ	1	1	0.75	0.75
NDCG	0.3699	0.4496	0.4608	0.4138
MAP	0.2191	0.3596	0.3462	0.3420

Stability of M

$$stability(M) := 1/n \bullet \Sigma_i abs(\gamma_i)$$