Sentiment Analysis in Information Retrieval

MTP Stage I

Roadmap

- Introduction
- Sentiment Analysis and Approaches
- Information Retrieval
- Joint Modeling of Topic and Sentiment
- Experiments
- Conclusion
- Future Work

Roadmap

- **Introduction**
- Sentiment Analysis and Approaches
- Information Retrieval
- Joint Modeling of Topic and Sentiment
- Experiments
- Conclusion
- Future Work

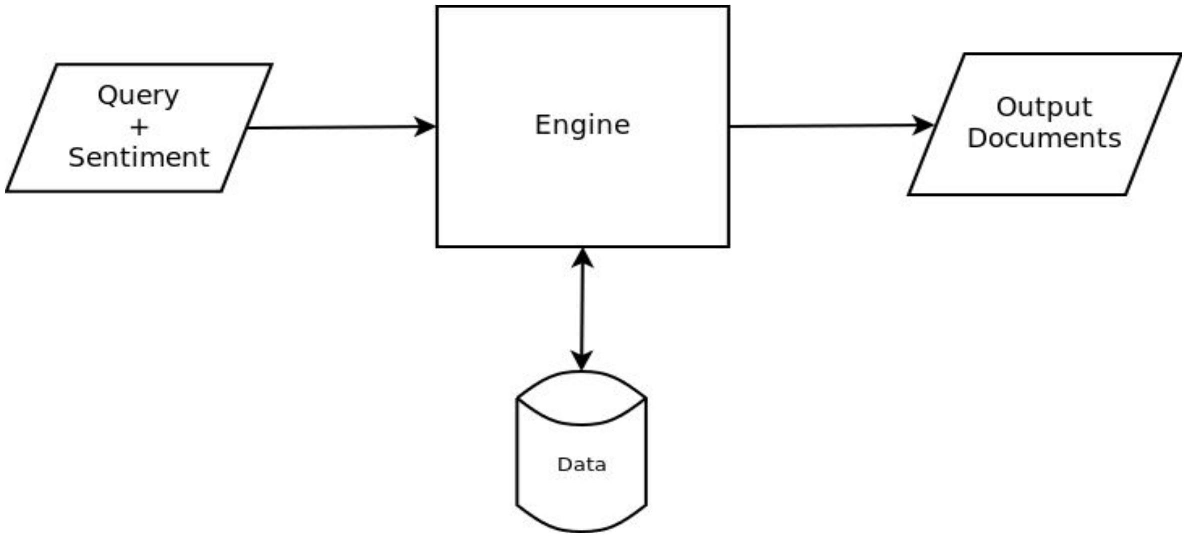
Introduction

- Main message
- Problem Statement
- Motivation
- Contributions

Main message

- Information Retrieval has historically been concerned with retrieval of factual data.
- Nowadays, user's need to fetch data of a desired sentiment
  - "positive reviews about a car", "negative reviews about a phone", etc
- Topic modeling has been used in IR to fetch data relevant to a particular topic.
- Joint modeling of topic and sentiment can address our need. This will be completely unsupervised. Also, this joint model can be used for sentiment classification.

Problem Statement



- To aid retrieval of subjective text of desired sentiment by making use of SA in IR.
- To achieve this, joint model of topic and sentiment has to be designed, implemented and compared with existing approaches.

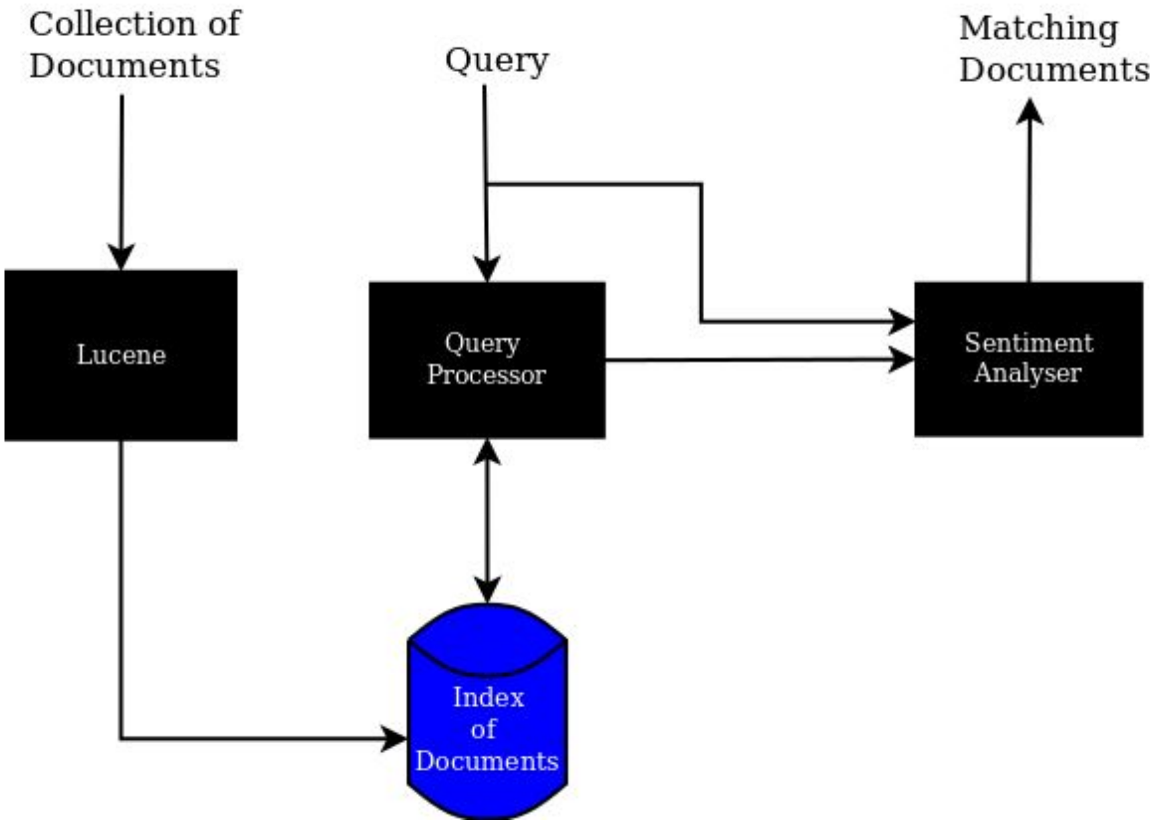
Motivation

- Sentiment Analysis has lot of applications
- Many SA techniques assume availability of subjective data.
  - UNREALISTIC
- Subjective data has to be retrieved.
- Traditional Information Retrieval is concerned with objective data.
- Using SA with IR, we can fetch subjective data of desired sentiment

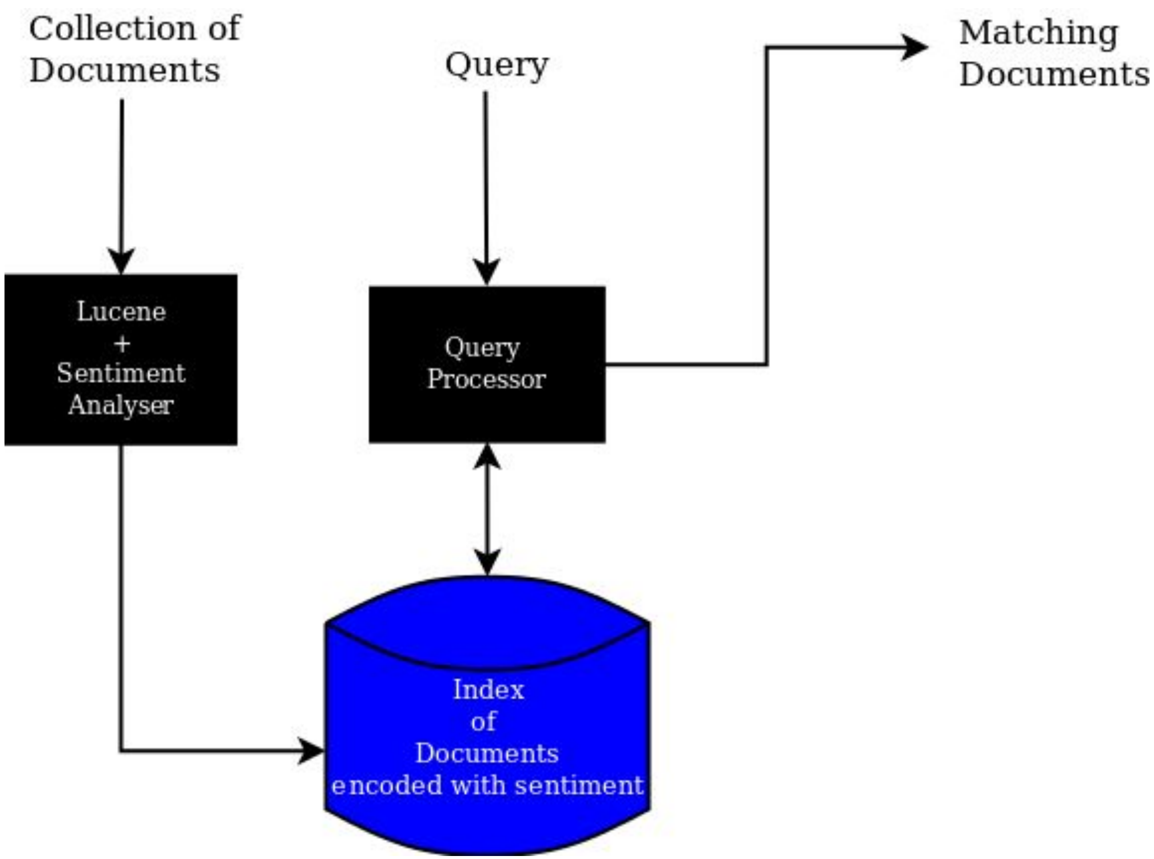
Contributions

- Implemented two POCs using Lucene, SentiWordNet, and Stanford POS tagger
  - Indexing followed by SA
  - Encoding Sentiment in Index

Indexing Followed by SA



Encoding Sentiment in Index



Contributions

- Studied LDA, Gibbs sampling, and Inference using LDA
- Evaluated LDA using implementation in Mallet and downloading documents from DMOZ
- Studied several joint topic-sentiment models

Roadmap

- Introduction
- **Sentiment Analysis and Approaches**
- Information Retrieval
- Joint Modeling of Topic and Sentiment
- Experiments
- Conclusion
- Future Work

Sentiment Analysis

- Sentiment implies some form of emotion expressed in the form of an opinion.
- The emotion can be positive, negative, or neutral towards an entity.
- Sentiment Analysis is the process of detecting whether a given *text* is positive, negative, or neutral.
- Subjectivity detection is a technique to identify whether a given piece of *text* is subjective or objective.

Approaches to SA

- Machine Learning
  - Supervised
  - Unsupervised
  - Semi-supervised
- Feature Vector
- Models used for classification
  - Generative models
  - Discriminative
- Models
  - Naive Bayes
  - MaxEnt
  - SVM

Roadmap

- Introduction
- Sentiment Analysis and Approaches
- **Information Retrieval**
- Joint Modeling of Topic and Sentiment
- Experiments
- Conclusion
- Future Work

Information Retrieval

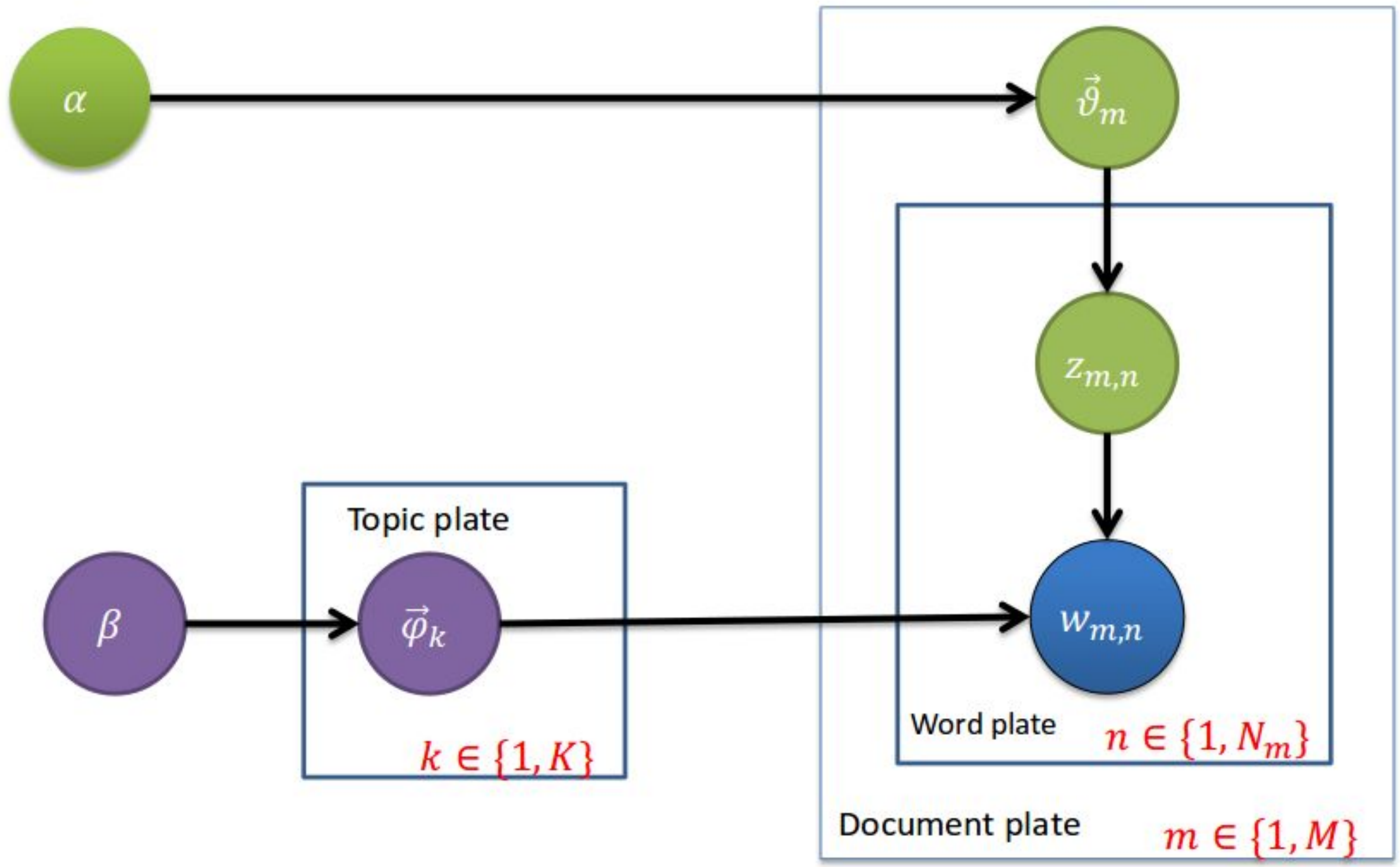
- Information Retrieval
  - retrieving information relevant to the user's need
  - need usually given in the form of a *query*
- Corpus Models
  - Multivariate binary model
  - Poisson model
  - Multinomial model
  - Dirichlet distribution model
  - DCM (Dirichlet Compound Multinomial model)
  - LDA (Latent Dirichlet Allocation)

- We will be discussing *LDA* in detail

Latent Dirichlet Allocation

- Probabilistic generative model
- Completely unsupervised in nature
- The main goal of *LDA* is to to *latent semantic analysis*
  - to find the latent structure of topics or concepts within a text

Bayesian Network of LDA



Generative Process of LDA

☐ Topic Plate :

for all topics $k \in [1, K]$ do

  sample mixture components $\vec{\psi}_k \sim Dir(\vec{\beta})$

end for

☐ Document Plate :

for all documents $m \in [1, M]$ do

  sample mixture proportion $\vec{\vartheta}_m \sim Dir(\vec{\alpha})$

  sample document length $N_m \sim Poiss(\xi)$

  ☐ Word Plate :

  for all words $n \in [1, N_m]$ in document m do

    sample topic index $z_{m,n} \sim Mult(\vec{\vartheta}_m)$

    sample term for word $w_{m,n} \sim Mult(\vec{\psi}_{z_{m,n}})$

  end for

end for

Steps in Gibbs Sampling

- Initialize
  - Assign random topic to each word in the corpus
  - Get the counts
- For a given burn-in period
  - For all documents
    - For all words in the document
      - Decrement counts corresponding to a topic
      - Sample a topic from the multinomial (full conditional)
      - Increment counts corresponding to that topic
- Estimate the parameter sets

Why LDA works?

- Sparsely distributed topics can result in high probability of a document
  - This can be done by making non-overlapping clusters of co-occurring words in different documents
- A sparse topic distribution per document will again increase its probability
  - Achieved by penalizing documents having too many possible topics
- This leads to forming clusters of words which belong to same topic.

Inferencing

- Randomly assign topics to words and then perform a number of loops through the Gibbs sampling update,

$$p(\tilde{z}_i = k|\widetilde{w}_i = t, \tilde{z}_{\neg i}, \widetilde{w}_{\neg i}; MC) = \frac{n_k{}^{(t)} + \tilde{n}_{k,\neg i}{}^{(t)} + \beta_t}{\sum_{t=1}^{V} n_k{}^{(t)} + \tilde{n}_{k,\neg i}{}^{(t)} + \beta_t} \cdot \frac{n_{\widetilde{m},\neg i}{}^{(k)} + \alpha_k}{\sum_{k=1}^{K} n_{\widetilde{m},\neg i}{}^{(k)} + \alpha_k}$$

Topic Distribution

$$\vartheta_{m,k} = \frac{n^{(k)}{}_{\tilde{m}} + \alpha_k}{\sum_{k=1}^{K} n^{(k)}{}_{\tilde{m}} + \alpha_k}$$

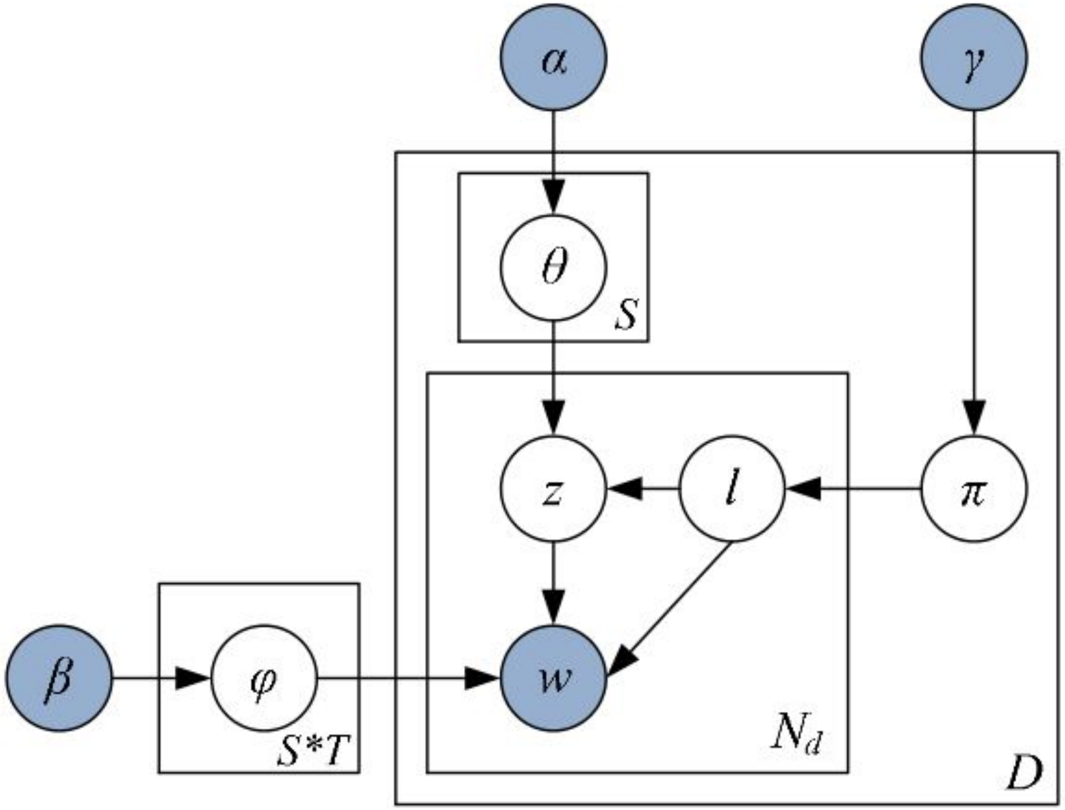
Roadmap

- Introduction
- Sentiment Analysis and Approaches
- Information Retrieval
- **Joint Modeling of Topic and Sentiment**
- Experiments
- Conclusion
- Future Work

Joint Modeling of Topic and Sentiment

- Bayesian network for JST (Joint Sentiment Topic model)
- Generative process of JST

Bayesian Network for JST



Generative Process of JST

For each document d, choose a distribution $\pi_d \sim Dir(\gamma)$.

For each sentiment label l under document d, choose a distribution $\theta_{d,k} \sim Dir(\alpha)$.

For each word w_i in document d

- choose a sentiment label $l_i \sim \pi_d$,
- choose a topic $z_i \sim \theta_{d,l_i}$,
- choose a word w_i from the distribution over words defined by the topic z_i and sentiment label l_i, $\psi_{z_i}\hat{}l_i$

Roadmap

- Introduction
- Sentiment Analysis and Approaches
- Information Retrieval
- Joint Modeling of Topic and Sentiment
- **Experiments**
- Conclusion
- Future Work

Evaluation of LDA

- Introduction
- Evaluation method
- Corpus
- Library used
- Results & Discussion
- Weighted evaluation algorithm
- Results
- Discussion
- High probability words

Introduction

- Topic models can be evaluation indirectly by evaluating performance of the application
- Also, calculating the probability of held-out documents is one way for evaluation
- We propose a new approach to evaluate Topic models in general and LDA in particular

Evaluation method

- Download documents
- Tag every document with a topic to get a tagged corpus
- Held-out some documents for testing before training model
- Train the topic model using the untagged corpus (obtained after removing the tags)
- Now, use the trained model to infer topic distribution for the testing documents
- Check whether the topic having highest proportion matches the tag of the document

Corpus

- 1273 documents were downloaded from DMOZ
- Computers, Films, Real Estate, Cooking, and Sports were the 5 topics chosen
- Each document is a description of a website which falls into one of these topics

Topic	No. of files
Computers	164
Films	384
Real Estate	261
Cooking	251
Sports	213

Library used

- Implementation of LDA in Mallet was used to perform the experiments

Results & Discussion

- Average accuracy after 5-fold cross validation on the corpus was 20.867
  - very low
- Reason
  - Short length of the documents
  - LDA works on the principle of co-occurrence
  - Words appearing in the same document have higher odds of ending up in the same topic
  - Due to the short length of the document, words from the same topic may not always co-occur
  - Also, there is a chance of them co-occurring with words from other topics, which results in bad clustering
  - LDA performs well with large number of topics. In this case, the number of topics is less.
- A more lenient evaluation will lead to better results

Weighted Evaluation

- Weighted evaluation is based on the fact that we get a topic distribution for each testing document
- This topic distribution is arranged in the descending order of topic proportions
- The idea is to assign weights according to the rank given to the original tag of the document

Weighted Evaluation Algorithm

- matches=0, counts=0
- For each document
  - Find topic distribution
  - Switch(tag):
    - case(topic1): matches += 1
    - case(topic2): matches += 0.8
    - case(topic3): matches += 0.6
    - case(topic4): matches += 0.4
    - case(topic5): matches += 0.2
  - counts++
- Accuracy = matches/counts

Results

Fold	Matches	Counts	Accuracy
Fold 1	156	255	61.4
Fold 2	156	255	61.4
Fold 3	170	255	66.9
Fold 4	152	255	59.6
Fold 5	161	253	63.9

Results

Average Accuracy	62.6

Discussion

- Average accuracy has increased
- 62 % accuracy implies
  - given a document, there is 62 % chance that the main topic of the document will be ranked in top 5
- If we consider clustering to be effective only till 3rd rank, we get following average accuracy

Average Accuracy (Till 3rd rank)	51.7

High probability words

Computers	Films	Cooking	Real Estate	Sports
site	film	recipes	services	reviews
software	information	recipe	real	news
free	offers	including	company	interviews
systems	production	collection	estate	information
programming	courses	tips	includes	features
research	links	source	commercial	current
resources	videos	production	based	tennis
code	television	banking	development	running
information	cinema	breakfast	title	tournament

Roadmap

- Introduction
- Sentiment Analysis and Approaches
- Information Retrieval
- Joint Modeling of Topic and Sentiment
- Experiments
- **Conclusion**
- Future Work

Conclusion

- Topic modeling approach can be extended to model the sentiment of text
- Though these two models make strong independence assumptions, their combination helps to model topic dependence of sentiment by encoding co-occurrence of sentiment and topic words
- SA and IR can be combined to achieve sentiment aware IR
- JST can also be used for sentiment classification, the sentiment which has the highest proportion will be the sentiment of the text

Achieved So Far

- A good understanding of topic models
  - Inference and querying using topic models
  - Implementation using Gibbs sampling algorithm
  - Evaluation methods

- Studied various joint sentiment-topic models
  - Methods to evaluate these models

Roadmap

- Introduction
- Sentiment Analysis and Approaches
- Information Retrieval
- Joint Modeling of Topic and Sentiment
- Experiments
- Conclusion
- **Future Work**

Future Work

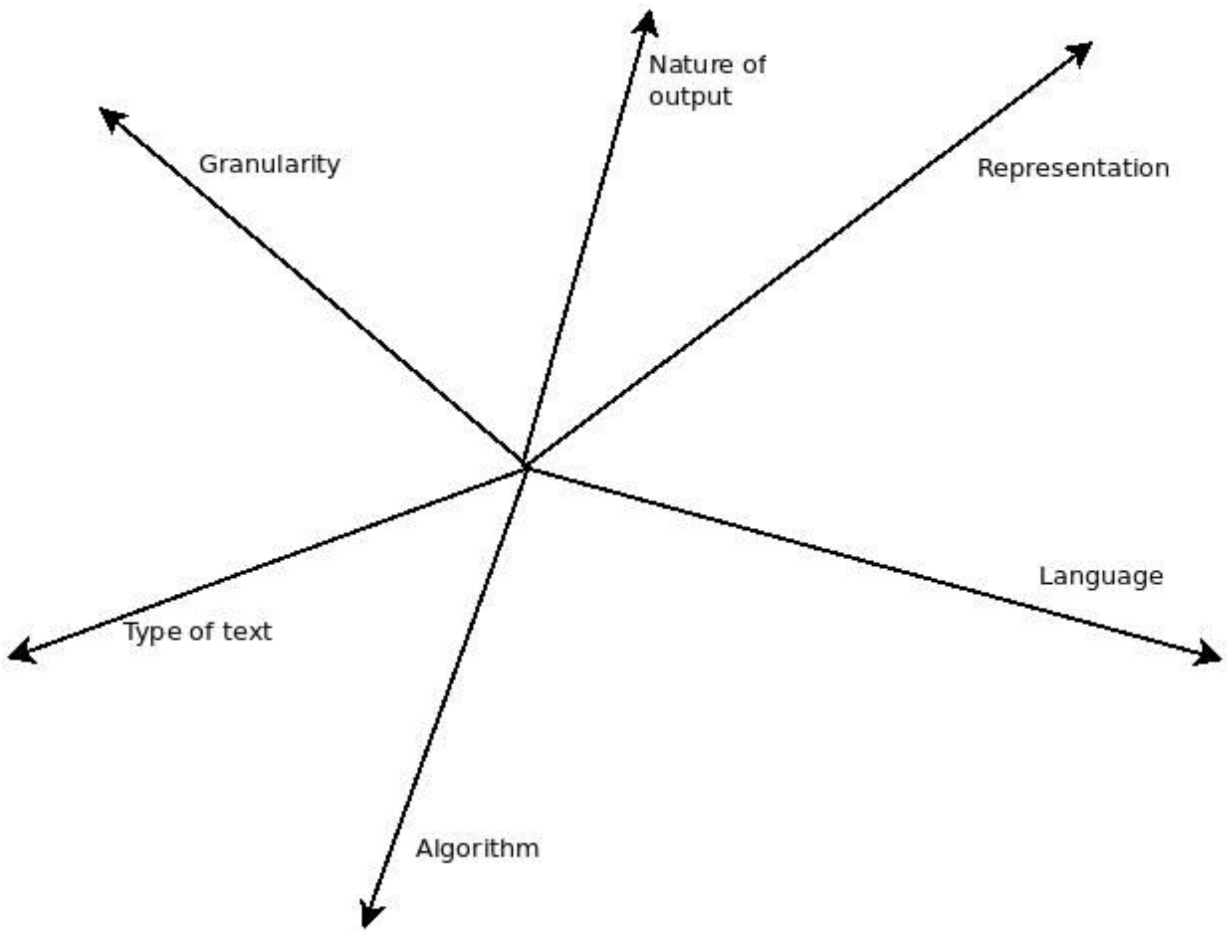
- Short Term
  - Designing a new generative model for sentiment and topics
  - JST doesn't exactly depict the document generation process
  - The fact that the topic chosen depends on the sentiment is not intuitive
  - So, designing a new model which addresses this problem is necessary
- Medium Term
  - Implementing the model
  - Evaluating it by comparison with existing approaches for sentiment classification
- Long Term
  - N-gram modeling
  - Inclusion of discourse and pragmatics in the model

THANKYOU

Problem Definition

- *A direct opinion (opinion about the object) is a quintuple ($o_j, f_{jk}, oo_{ijkl}, h_i, t_l$), where o_j is an object, f_{jk} is a feature of the object o_j, oo_{ijkl} is the orientation or polarity of the opinion on feature f_{jk} of the object o_j , h_i is the opinion holder and t_l is the time when the opinion is expressed by h_i.*

Types of SA


Nature of output
Granularity
Representation
Language
Type of text
Algorithm

Challenges

- Unstructured Text
  - micro-blogs, tweets, comments, and messages
- Sarcasm
  - *"Great! I ate too many chocolates and gained lot of weight :)"*
- Thwarting
  - *"The guy is a chronic drinker, he smokes weed, has drugs but is a good guy"*

Application

- Classification of Tweets
- Classification of Movie Reviews
- Classification of Product Reviews
- Analyzing market trends
- Sentiment Aware Information Retrieval
- Removing subjective sentences to improve IR performance

Example

1) I went to watch the new James Bond ick, Skyfall at IMAX which is the best theater in Mumbai with my brother a month ago. 2) I really liked the seating arrangement over there. 3) The screenplay was superb and kept me guessing till the end. 4) My brother doesn't like the hospitality in the theater even now. 5) The movie is really good and the best bond ick ever

Example

1) I went to watch the new James Bond flick, Skyfall at IMAX which is the best theater in Mumbai with my brother a month ago. 2) I really liked the seating arrangement over there. 3) The screenplay was superb and kept me guessing till the end. 4) My brother doesn't like the hospitality in the theater even now. 5) The movie is really good and the best bond ick ever

Objective sentence

Example

1) I went to watch the new James Bond flick, Skyfall at IMAX which is the best theater in Mumbai with my brother a month ago. 2) I really liked the seating arrangement over there. 3) The screenplay was superb and kept me guessing till the end. 4) My brother doesn't like the hospitality in the theater even now. 5) The movie is really good and the best bond ick ever

Positive word

Example

1) I went to watch the new James Bond flick, Skyfall at IMAX which is the best theater in Mumbai with my brother a month ago. 2) I really liked the seating arrangement over there. 3) The screenplay was superb and kept me guessing till the end. 4) My brother doesn't like the hospitality in the theater even now. 5) The movie is really good and the best bond ick ever

Positive Sentence

Example

1) I went to watch the new James Bond flick, Skyfall at IMAX which is the best theater in Mumbai with my brother a month ago. 2) I really liked the seating arrangement over there. 3) The screenplay was superb and kept me guessing till the end. 4) My brother doesn't like the hospitality in the theater even now. 5) The movie is really good and the best bond ick ever

Positive Word

Example

1) I went to watch the new James Bond flick, Skyfall at IMAX which is the best theater in Mumbai with my brother a month ago. 2) I really liked the seating arrangement over there. 3) The screenplay was superb and kept me guessing till the end. 4) My brother doesn't like the hospitality in the theater even now. 5) The movie is really good and the best bond ick ever

Positive Sentence

Example

1) I went to watch the new James Bond flick, Skyfall at IMAX which is the best theater in Mumbai with my brother a month ago. 2) I really liked the seating arrangement over there. 3) The screenplay was superb and kept me guessing till the end. 4) My brother doesn't like the hospitality in the theater even now. 5) The movie is really good and the best bond ick ever

Negative Word

Example

1) I went to watch the new James Bond flick, Skyfall at IMAX which is the best theater in Mumbai with my brother a month ago. 2) I really liked the seating arrangement over there. 3) The screenplay was superb and kept me guessing till the end. 4) My brother doesn't like the hospitality in the theater even now. 5) The movie is really good and the best bond ick ever

Objective Sentence w.r.t. movie

Example

1) I went to watch the new James Bond flick, Skyfall at IMAX which is the best theater in Mumbai with my brother a month ago. 2) I really liked the seating arrangement over there. 3) The screenplay was superb and kept me guessing till the end. 4) My brother doesn't like the hospitality in the theater even now. 5) The movie is really good and the best bond ick ever

Positive Word

Example

1) I went to watch the new James Bond flick, Skyfall at IMAX which is the best theater in Mumbai with my brother a month ago. 2) I really liked the seating arrangement over there. 3) The screenplay was superb and kept me guessing till the end. 4) My brother doesn't like the hospitality in the theater even now. 5) The movie is really good and the best bond ick ever

Positive Sentence

Likelihoods

- Probability that a particular word $w_{m,n}$ instantiates a particular term t given the LDA parameters is,
- $p\left(w_{m,n} = t \middle| \vec{\vartheta}_m, \underline{\phi}\right) = \sum_{k=1}^{K} p\left(w_{m,n} = t \middle| \vec{\varphi}_k\right) p(z_{m,n} = k | \vec{\vartheta}_m)$
- This corresponds to one iteration on the word plate of the Bayesian network
- Joint distribution of all know and hidden variables given the hyperparameters,
- $p\left(\vec{w}_m, \vec{z}_m, \vec{\vartheta}_m, \underline{\phi} \middle| \vec{\alpha}, \vec{\beta}\right) = \prod_{n=1}^{N_m} \overbrace{\underbrace{p\left(w_{m,n} \middle| \vec{\varphi}_{z_{m,n}}\right). p\left(z_{m,n} \middle| \vec{\vartheta}_m\right)}_{\text{Word plate}}. p\left(\vec{\vartheta}_m \middle| \vec{\alpha}\right)}^{\text{Document plate (1 document)}} . \underbrace{p(\underline{\phi} | \vec{\beta})}_{\text{Topic plate}}$

Likelihoods

- Likelihood of the document,

$$p(\vec{w}_m|\vec{\alpha}|\vec{\beta}) = \iint p(\vec{\vartheta}_m|\vec{\alpha}).\, p\left(\underline{\phi}\middle|\vec{\beta}\right).\prod_{n=1}^{N_m}\sum_{z_{m,n}} p(w_{m,n}|\vec{\varphi}_{z_{m,n}})\, p(z_{m,n}|\vec{\vartheta}_m)\, d\underline{\phi}\, d\vec{\vartheta}_m$$

$$= \iint p(\vec{\vartheta}_m|\vec{\alpha}).\, p\left(\underline{\phi}\middle|\vec{\beta}\right).\prod_{n=1}^{N_m} p(w_{m,n}|\vec{\vartheta}_m, \underline{\phi})\, d\underline{\phi}\, d\vec{\vartheta}_m$$

- Likelihood of the corpus, $\mathcal{W} = \{\vec{w}_m\}_{m=1\ldots M}$,

$$p\left(\mathcal{W}\middle|\vec{\alpha}, \vec{\beta}\right) = \prod_{m=1}^{M} p(\vec{w}_m|\vec{\alpha}, \vec{\beta})$$

Quantities in Likelihood

M number of documents to generate (const scalar).

K number of topics/mixture components (const scalar).

V number of terms t in vocabulary (const scalar).

$\vec{\alpha}$ hyper-parameter on the mixing proportions ($K-vector$ or scalar if symmetric).

$\vec{\beta}$ hyper-parameter on the mixing components ($K-vector$ or scalar if symmetric).

$\vec{\vartheta}_m$ parameter notation for p(z|d=m), the topic mixture proportion for document m.
One proportion for each document, $\underline{\theta} = \{\vec{\vartheta}_m\}\ m=1 \cdots M\ (M \times K\ matrix)$.

$\vec{\psi}_k$ parameter notation for p(t|z=k), the mixture component of topic k.
One component for each topic, $\underline{\phi} = \{\vec{\psi}_k\}\ k=1 \cdots K\ (K \times V\ matrix)$.

N_m document length (document-specific), here modeled with a Poisson distribution with constant parameter xi.

$z_{m,n}$ mixture indicator that chooses the topic for the nth word in document m.

$w_{m,n}$ term indicator for the nth word in document m.

Gibbs Sampling

- Makov-chain Monte Carlo (MCMC) can emulate high-dimensional probability distributions $p(\vec{x})$ by the stationary distribution of a Markov chain
- Each sample is generated for each transition in the chain
  - After a stationary state of the chain has been reached
  - This happens after so-called "burn-in period" which eliminates the effect of initialization parameters
- Gibbs sampling is a special case of MCMC where
  - The dimensions x_i of the distribution are sampled alternately one at a time, conditioned on the values of all other dimensions, denoted by $\vec{x}_{\neg i}$

Bivariate Case

- Consider a bivariate random variable (x, y), and suppose we wish to compute the marginals, $p(x)$ and $p(y)$
- The idea behind the sampler
  - Easier to consider a sequence of distributions, $p(x|y)$ and $p(y|x)$
  - Than obtaining marginal by integration, $p(x) = \int p(x,y)dy$
- Steps
  - Start with some initial value y_0 for y
  - Obtain x_0 by generating a random variable from the conditional distribution, $p(x|y = y_0)$
  - Use x_0 to generate a new value of y_1, drawing from the conditional distribution, $p(y|x = x_0)$

Bivariate Case

- The sampler proceeds as follows:
  - $x_i \sim p(x|y = y_{i-1})$
  - $y_i \sim p(y|x = x_i)$
- Repeating this process k times, generates a Gibbs sequence of length k, where
  - a subset of points $(x_{j,} y_j)\ for\ 1 \leq j \leq m < k$ are taken as the simulated draws from the full joint distribution

Multivariate Case

- The value of the k^{th} variable is drawn from the distribution, $p(\theta^{(k)}|\boldsymbol{\theta}^{(\neg k)})$ where $\boldsymbol{\theta}^{(\neg k)}$ denotes a vector containing all of the variable but k
- We draw from the distribution,

$$\theta_i^{(k)} \sim p\left(\theta^{(k)}\middle|\theta^{(1)} = \theta_i^{(1)}, \ldots, \theta^{(k-1)} = \theta_i^{(k-1)}, \theta^{(k+1)} = \theta_{i-1}^{(k+1)}\right), \ldots, \theta^{(n)} = \theta_{i-1}^{(n)}$$

- For example, if there are four variables, (w, x, y, z) the sampler becomes,
  - $w_i \sim p(w \mid x = x_{i-1}, y = y_{i-1}, z = z_{i-1})$
  - $x_i \sim p(w \mid w = w_i, y = y_{i-1}, z = z_{i-1})$
  - $y_i \sim p(w \mid w = w_i, x = x_i, z = z_{i-1})$
  - $z_i \sim p(w \mid w = w_i, x = x_i, y = y_i)$

Gibbs Sampling Algorithm

To get a sample from $p(x)$

1. Choose dimension i (random or by permutation)
2. Sample x_i from $p(x_i \mid \vec{x}_{\neg i})$

- $p(x_i \mid \vec{x}_{\neg i}) = \frac{p(\vec{x})}{p(\vec{x}_{\neg i})}$ with $\vec{x} = \{x_i, \vec{x}_{\neg i}\}$

Gibbs Sampling for models with hidden variables

- For models containing hidden variables $\vec{z}$, their posterior given the evidence, $p(\vec{z}|\vec{x})$ is a distribution commonly wanted

- The general formulation of a Gibbs sampler for such latent-variable models becomes:

$$p(z_i \mid \vec{z}_{\neg i}, \vec{x}) = \frac{p(\vec{z}, \vec{x})}{p(\vec{z}_{\neg i}, \vec{x})}$$

LDA Gibbs Sampler

- Target of inference is the distribution, $p(\vec{z}|\vec{w})$

$$p(\vec{z}|\vec{w}) = \frac{p(\vec{z},\vec{w})}{p(\vec{w})} = \frac{\prod_{i=1}^{W} p(z_i, w_i)}{\prod_{i=1}^{W} \sum_{k=1}^{K} p(z_i = k, w_i)}$$

- Full conditional, $p(z_i|\vec{z}_{\neg i}, \vec{w})$ is used to simulate $p(\vec{z}|\vec{w})$

- This requires the joint distribution,

$$p(\vec{w},\vec{z}|\vec{\alpha},\vec{\beta}) = p(\vec{w}|\vec{z},\vec{\beta})p(\vec{z}|\vec{\alpha})$$

$$p(\vec{w}|\vec{z},\vec{\beta})$$

- W words of the corpus are observed according to the independent multinomial trials

$$p\left(\vec{w}\middle|\vec{z},\underline{\phi}\right) = \prod_{t=1}^{W} p(w_i|z_i) = \prod_{t=1}^{W} \varphi_{z_i,w_i}$$

- Splitting the product over words into product over topics and one over vocabulary,

$$p\left(\vec{w}\middle|\vec{z},\underline{\phi}\right) = \prod_{k=1}^{K}\prod_{t=1}^{V} p(w_i = t|z_i = k) = \prod_{k=1}^{K}\prod_{t=1}^{V} {\varphi_{k,t}}^{{n_k}^{(t)}}$$

where, $n_k^{(t)}$ denotes the number of times that the term t has been observed with topic k.

$$p(\vec{w}|\vec{z},\vec{\beta})$$

- Integrating over $\underline{\phi}$, we get

$$p(\vec{w}|\vec{z},\vec{\beta}) = \int p\left(\vec{w}\middle|\vec{z},\underline{\phi}\right) p(\underline{\phi}|\vec{\beta})\, \mathrm{d}\underline{\phi}$$

$$= \int \prod_{z=1}^{K} \frac{1}{\Delta(\vec{\beta})} \prod_{t=1}^{V} {\varphi_{z,t}}^{{n_z}^{(t)}+\beta_t-1} d\vec{\varphi}_z$$

$$= \prod_{z=1}^{K} \frac{\Delta(\vec{n}_z+\vec{\beta})}{\Delta(\vec{\beta})}\ ,\vec{n}_z = \{{n^{(t)}}_z\}_{t=1\ldots N}$$

$$p(\vec{z}|\vec{\alpha})$$

$$\begin{aligned} p(\vec{z}|\underline{\theta}) &= \prod_{i=1}^{W} p(z_i|d_i) \\ &= \prod_{m=1}^{M}\prod_{k=1}^{K} p(z_i = k|d_i = m) \\ &= \prod_{m=1}^{M}\prod_{k=1}^{K} \vartheta_{m,k}{}^{n_m{}^{(k)}} \end{aligned}$$

where, d_i refers to the document a word i belongs to and $n_m^{(k)}$ refers to the number of times that topic k has been observed with a word of document m.

$$p(\vec{z}|\vec{\alpha})$$

$$p(\vec{z}|\vec{\alpha}) = \int p\left(\vec{z}|\underline{\theta}\right)p(\underline{\theta}|\vec{\alpha})\, d\underline{\theta}$$

$$= \int \prod_{m=1}^{M} \frac{1}{\Delta(\vec{\alpha})} \prod_{k=1}^{K} \vartheta_{m,k}{}^{n_m{}^{(k)}+\alpha_k-1} d\vec{\vartheta}_m$$

$$= \prod_{m=1}^{M} \frac{\Delta(\vec{n}_m + \vec{\alpha})}{\Delta(\vec{\alpha})}\ , \vec{n}_m = \{n^{(k)}{}_m\}_{k=1\ldots K}$$

Joint distribution

$$p\left(\vec{z},\vec{w}\middle|\vec{\alpha},\vec{\beta}\right) = \prod_{z=1}^{K}\frac{\Delta(\vec{n}_z+\vec{\beta})}{\Delta(\vec{\beta})}\prod_{m=1}^{M}\frac{\Delta(\vec{n}_m+\vec{\alpha})}{\Delta(\vec{\alpha})}$$

Full conditional

$$p(z_i = k|\vec{z}_{\neg i}, \vec{w}) = \frac{p(\vec{w}, \vec{z})}{p(\vec{w}, \vec{z}_{\neg i})} = \frac{p(\vec{w}|\vec{z})p(\vec{z})}{p(\vec{w}|\vec{z}_{\neg i})p(\vec{z}_{\neg i})}$$

$$p(z_i = k|\vec{z}_{\neg i}, \vec{w}) \propto \frac{\Delta\left(\vec{n}_z + \vec{\beta}\right)}{\Delta\left(\vec{n}_{z,\neg i} + \vec{\beta}\right)} \cdot \frac{\Delta(\vec{n}_m + \vec{\alpha})}{\Delta\left(\vec{n}_{m,\neg i} + \vec{\alpha}\right)}$$

$$\propto \frac{\Gamma\left({n_k}^{(t)} + \beta_t\right)\Gamma\left(\sum_{t=1}^{V} {n_{k,\neg i}}^{(t)} + \beta_t\right)}{\Gamma\left({n_{k,\neg i}}^{(t)} + \beta_t\right)\Gamma(\sum_{t=1}^{V} {n_k}^{(t)} + \beta_t)} \cdot \frac{\Gamma({n_m}^{(k)} + \alpha_k)\Gamma(\sum_{k=1}^{K} {n_{m,\neg i}}^{(k)} + \alpha_k)}{\Gamma({n_{m,\neg i}}^{(k)} + \alpha_k)\Gamma(\sum_{k=1}^{K} {n_m}^{(k)} + \alpha_k)}$$

$$\propto \frac{{n_{k,\neg i}}^{(t)} + \beta_t}{\sum_{t=1}^{V} {n_{k,\neg i}}^{(t)} + \beta_t} \cdot \frac{{n_{m,\neg i}}^{(k)} + \alpha_k}{\sum_{k=1}^{K} {n_{m,\neg i}}^{(k)} + \alpha_k}$$

Multinomial parameter sets $\underline{\theta}$ and $\underline{\phi}$

$$p(\vec{\vartheta}_m | \mathrm{M}C, \vec{\alpha}) = \frac{1}{Z_{\vartheta_m}} \prod_{n=1}^{N_m} p(z_{m,n} | \vec{\vartheta}_m) p(\vec{\vartheta}_m | \vec{\alpha})$$
$$= \mathrm{Dir}(\vec{\vartheta}_m | \vec{n}_m + \vec{\alpha})$$

$$p(\vec{\varphi}_m | \mathrm{M}C, \vec{\beta}) = \frac{1}{Z_{\varphi_k}} \prod_{[i:z_i=k]}^{N_m} p(w_i | \vec{\varphi}_k) p(\vec{\varphi}_k | \vec{\beta})$$
$$= \mathrm{Dir}(\vec{\varphi}_k | \vec{n}_k + \vec{\beta})$$

Multinomial parameter sets $\underline{\theta}$ *and* $\underline{\phi}$

- Using the expectation of the Dirichlet distribution, $Dir(\vec{\alpha}) = a_i / \sum_i a_i$,

$$\varphi_{k,t} = \frac{n^{(t)}{}_k + \beta_t}{\sum_{t=1}^{V} n^{(t)}{}_k + \beta_t}$$

$$\vartheta_{m,k} = \frac{n^{(k)}{}_m + \alpha_k}{\sum_{k=1}^{K} n^{(k)}{}_m + \alpha_k}$$

References

- *Bing Liu. Sentiment analysis and subjectivity. Handbook of natural language processing, 2:568, 2010.*
- *David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. the Journal of machine Learning research, 3:993–1022, 2003*
- *Gregor Heinrich. Parameter estimation for text analysis. Web: http://www.arbylon.net/publications/text-est.pdf, 2005.*
- *Brian Walsh. Markov chain monte carlo and gibbs sampling. 2004.*
- *Chenghua Lin and Yulan He. Joint sentiment/topic model for sentiment analysis. In Proceedings of the 18th ACM conference on Information and knowledge management, pages 375–384. ACM, 2009.*
- *Hanna M Wallach, Iain Murray, Ruslan Salakhutdinov, and David Mimno. Evaluation methods for topic models. In Proceedings of the 26th Annual International Conference on Machine Learning, pages 1105–1112. ACM, 2009.*