Joint Author Sentiment Topic Model

Subhabrata Mukherjee
Max Planck Institute for Informatics

Gaurab Basu and Sachindra Joshi
IBM India Research Lab

April 25, 2014

Aspect Rating and Review Rating

“ [This film is based on a true-life incident. It sounds like a <u>great</u> plot and the director makes a <u>decent</u> attempt in narrating a <u>powerful</u> story.] [However, the film <u>does not quite make the mark</u> due to <u>sloppy</u> acting.] ”

Aspect Rating and Review Rating

" [This film is based on a true-life incident. It sounds like a <u>great</u> plot and the director makes a <u>decent</u> attempt in narrating a <u>powerful</u> story.] [However, the film <u>does not quite make the mark</u> due to <u>sloppy</u> acting.] "

- Identify topics - *direction, story* and *acting*
  - Story has facets - *plot* and *narration*

Aspect Rating and Review Rating

“ [This film is based on a true-life incident. It sounds like a <u>great</u> plot and the director makes a <u>decent</u> attempt in narrating a <u>powerful</u> story.] [However, the film <u>does not quite make the mark</u> due to <u>sloppy</u> acting.] ”

- Identify topics - *direction, story* and *acting*
  - Story has facets - *plot* and *narration*

- Identify facet sentiments – *great* (plot)*, powerful* (story)*, sloppy* (acting) *etc.*

Aspect Rating and Review Rating

“ [This film is based on a true-life incident. It sounds like a great plot and the director makes a decent attempt in narrating a powerful story.] [However, the film does not quite make the mark due to sloppy acting.] ”

- Identify topics - *direction, story* and *acting*
  - Story has facets - *plot* and *narration*

- Identify facet sentiments – *great* (plot), *powerful* (story), *sloppy* (acting) *etc.*

- Overall review rating - aggregation of facet-specific sentiments

Aspect Rating and Review Rating

" [This film is based on a true-life incident. It sounds like a <u>great</u> plot and the director makes a <u>decent</u> attempt in narrating a <u>powerful</u> story.] [However, the film <u>does not quite make the mark</u> due to <u>sloppy</u> acting.] "

- Identify topics - *direction, story* and *acting*
  - Story has facets - *plot* and *narration*

- Identify facet sentiments – *great* (plot), *powerful* (story), *sloppy* (acting) *etc.*

- Overall review rating - aggregation of facet-specific sentiments

- Why joint modeling ?
  - Sentiment words help locating topic words and vice-versa
  - Neighboring words establish semantics / sentiment of terms

Why Author-Specificity ?

" [This film is based on a true-life incident. It sounds like a <u>great</u> plot and the director makes a <u>decent</u> attempt in narrating a <u>powerful</u> story.] [However, the film <u>does not quite make the mark</u> due to <u>sloppy</u> acting.] "

Why Author-Specificity ?

“ [This film is based on a true-life incident. It sounds like a great plot and the director makes a decent attempt in narrating a powerful story.] [However, the film does not quite make the mark due to sloppy acting.] ”

- Overall rating varies for authors with different topic preferences
  - *Positive* for those with greater preference for *acting* and *narration*
  - *Negative* for *acting*

Why Author-Specificity ?

“ [This film is based on a true-life incident. It sounds like a <u>great</u> plot and the director makes a <u>decent</u> attempt in narrating a <u>powerful</u> story.] [However, the film <u>does not quite make the mark</u> due to <u>sloppy</u> acting.] ”

- Overall rating varies for authors with different topic preferences
  - *Positive* for those with greater preference for *acting* and *narration*
  - *Negative* for *acting*

- Affective sentiment value varies for authors
  - How much negative is “*does not quite make the mark*” for me ?

Why Author-Specificity ?

“ [This film is based on a true-life incident. It sounds like a great plot and the director makes a decent attempt in narrating a powerful story.] [However, the film does not quite make the mark due to sloppy acting.] ”

- Overall rating varies for authors with different topic preferences
  - *Positive* for those with greater preference for *acting* and *narration*
  - *Negative* for *acting*
- Affective sentiment value varies for authors
  - How much negative is “*does not quite make the mark*” for me ?
- Author-writing style helps in locating / associating facets and sentiments
  - E.g. *topic switch*, *verbosity*, *use of content* and *function words etc.*
  - The author makes a topic switch in above review using the function word *however*

Why Author-Specificity ?

“ [This film is based on a true-life incident. It sounds like a great plot and the director makes a decent attempt in narrating a powerful story.] [However, the film does not quite make the mark due to sloppy acting.] ”

- Overall rating varies for authors with different topic preferences
  - *Positive* for those with greater preference for *acting* and *narration*
  - *Negative* for *acting*

- Affective sentiment value varies for authors
  - How much negative is “*does not quite make the mark*” for me ?

- Author-writing style helps in locating / associating facets and sentiments
  - E.g. *topic switch*, *verbosity*, *use of content* and *function words etc.*
  - The author makes a topic switch in above review using the function word *however*

- *Traditional works learn a global model independent of the review author*

Why care about writing style or coherence?

➢ Better association of facets to topics by detecting *semantic-syntactic class transitions* and *topic switch*

➢ *semantic dependencies* - association between facets to topics

➢ *syntactic dependencies* - connection between facets and background words required to make the review coherent and grammatically correct

Contributions

- Show that author identity helps in rating prediction

Contributions

- Show that author identity helps in rating prediction

- Author-specific generative model of a review that incorporates author-specific
    - *topic and facet preferences*

Contributions

- Show that author identity helps in rating prediction

- Author-specific generative model of a review that incorporates author-specific
  - *topic and facet preferences*
  - *grading style*

Contributions

- Show that author identity helps in rating prediction

- Author-specific generative model of a review that incorporates author-specific
  - *topic and facet preferences*
  - *grading style*
  - *writing style*

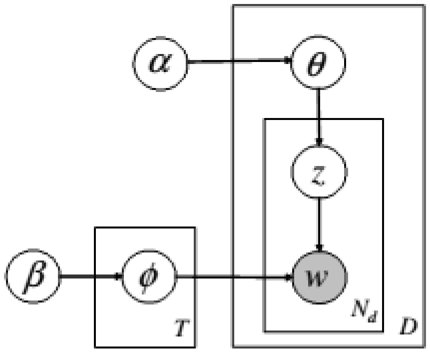
Contributions

- Show that author identity helps in rating prediction

- Author-specific generative model of a review that incorporates author-specific
  - *topic and facet preferences*
  - *grading style*
  - *writing style*
  - maintain *coherence* in reviews

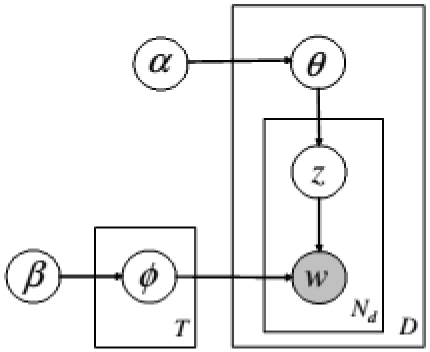
Topic Models

Topic Models

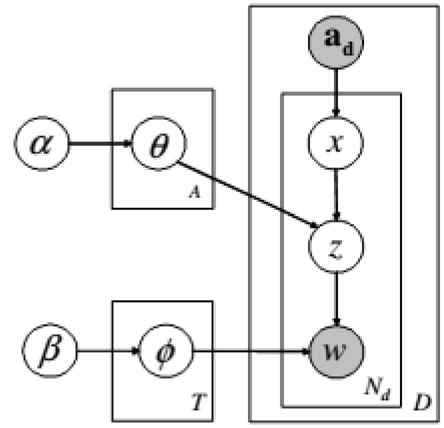


1. LDA Model

Topic Models

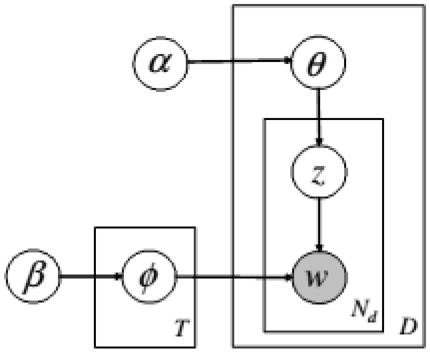


1. LDA Model

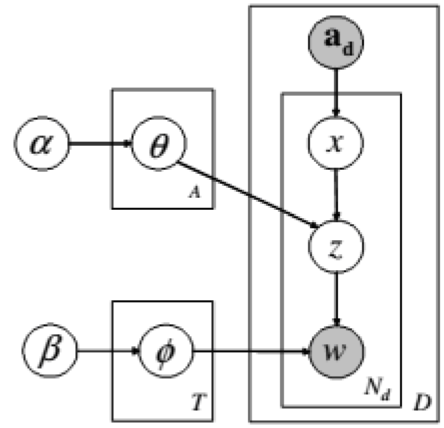


2. Author-Topic Model

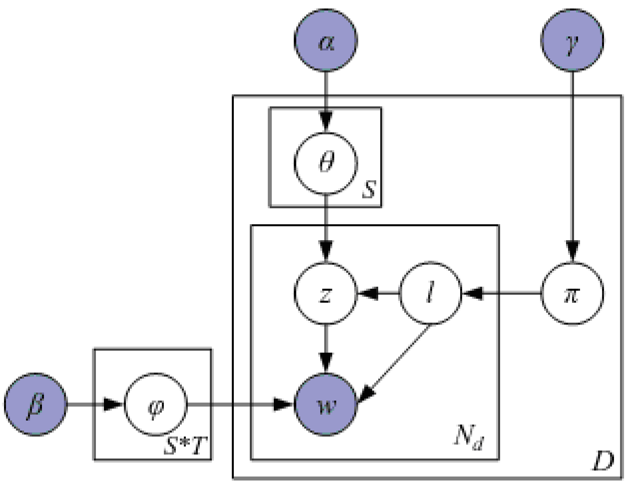
Topic Models



1. LDA Model



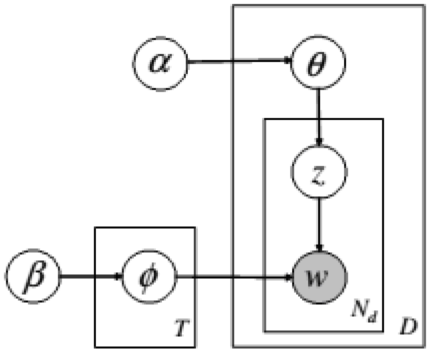
2. Author-Topic Model



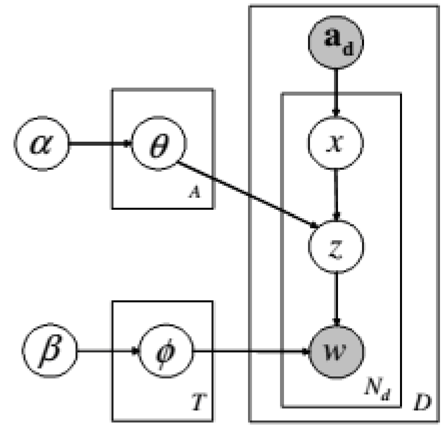
3. Joint Sentiment Topic Model

max planck institut
informatik

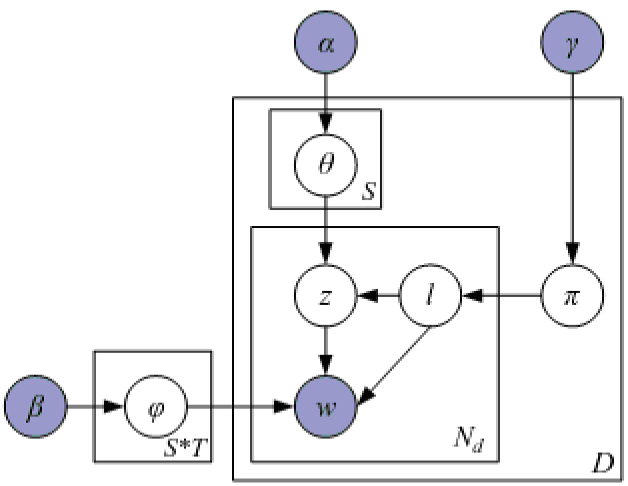
Topic Models



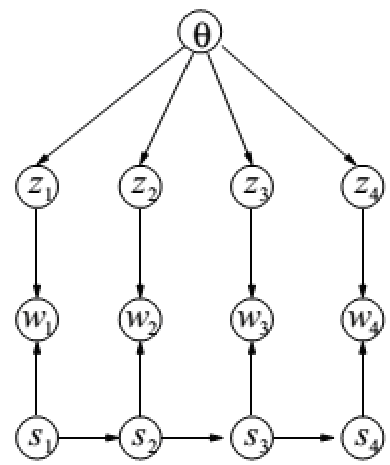
1. LDA Model



2. Author-Topic Model



3. Joint Sentiment Topic Model



4. Topic Syntax Model

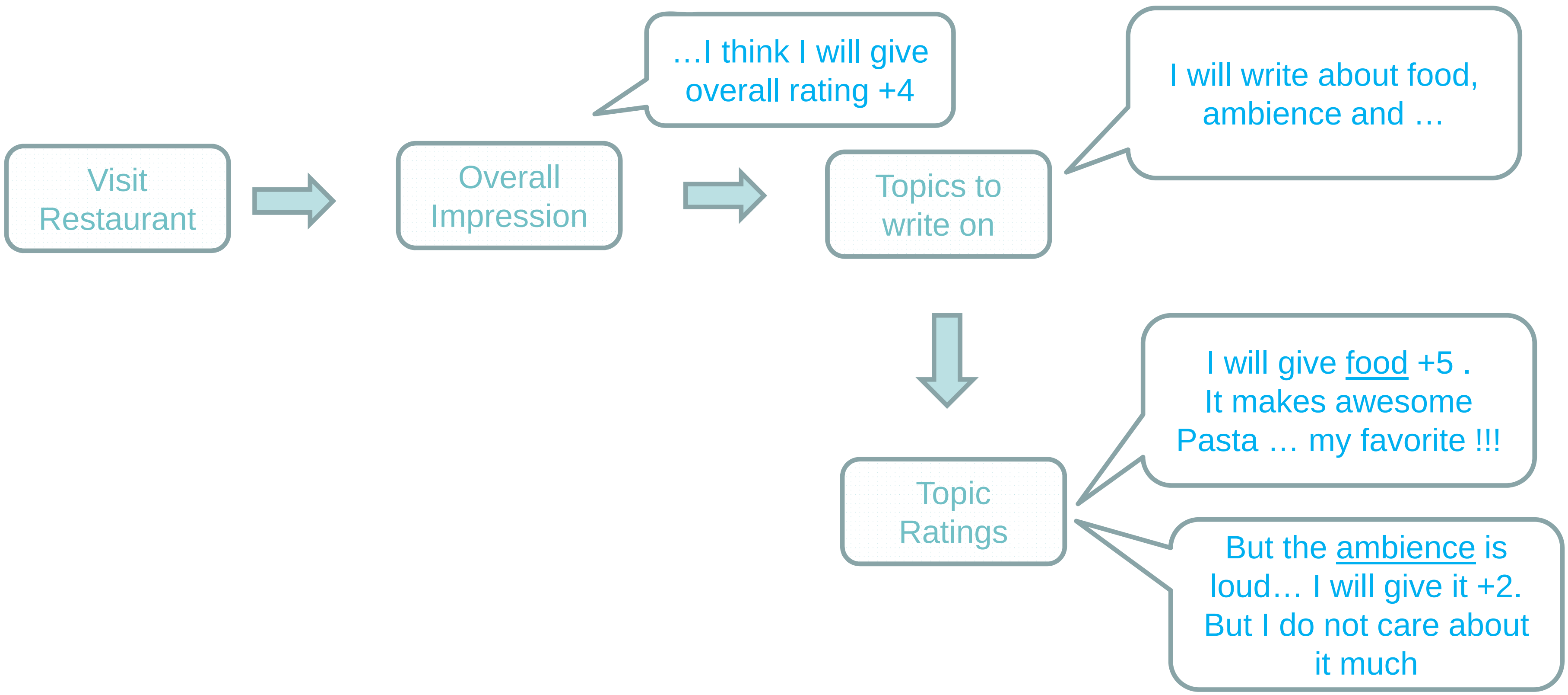
Generative Process for a Review

Visit
Restaurant

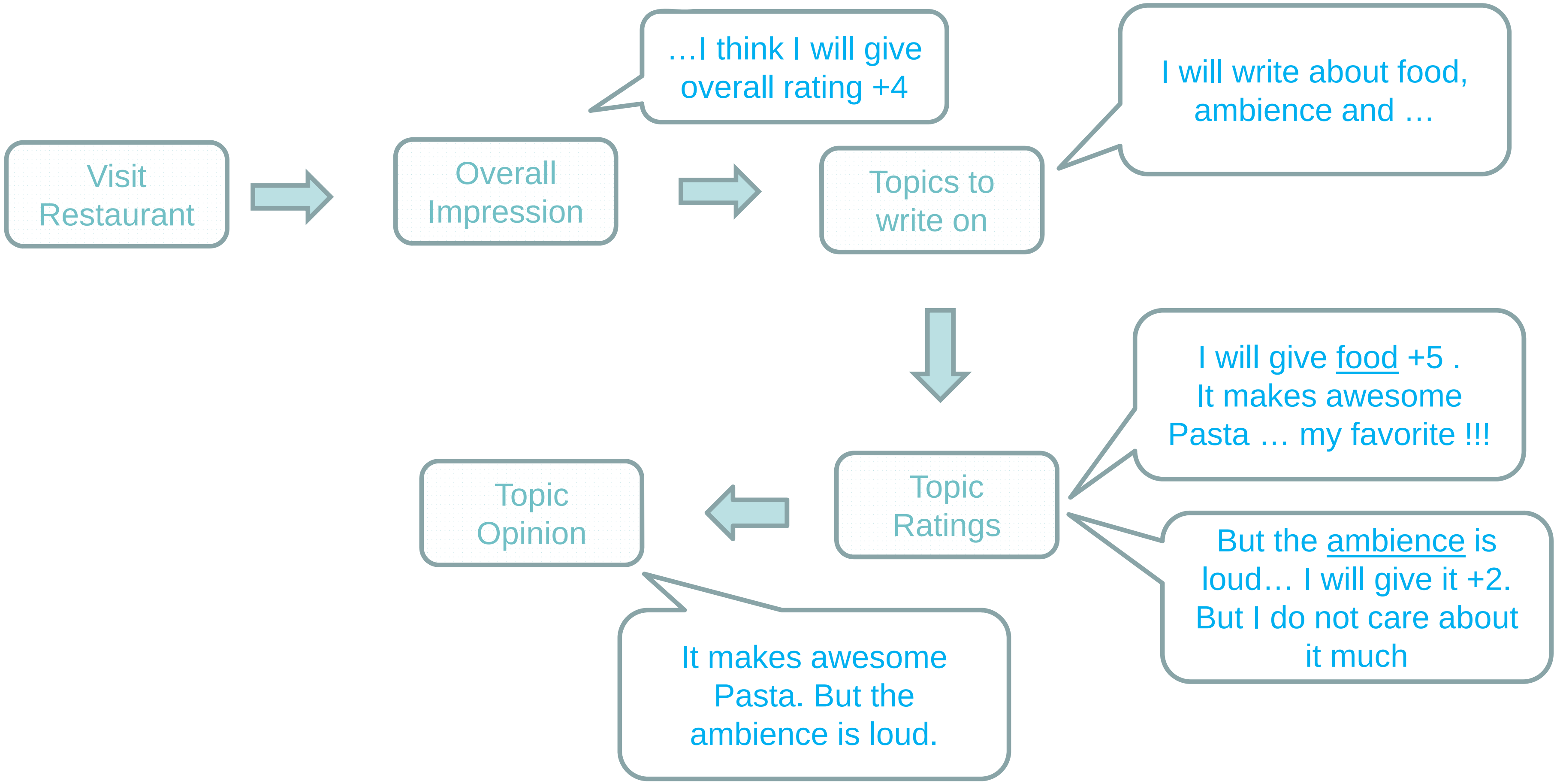
Generative Process for a Review

Generative Process for a Review

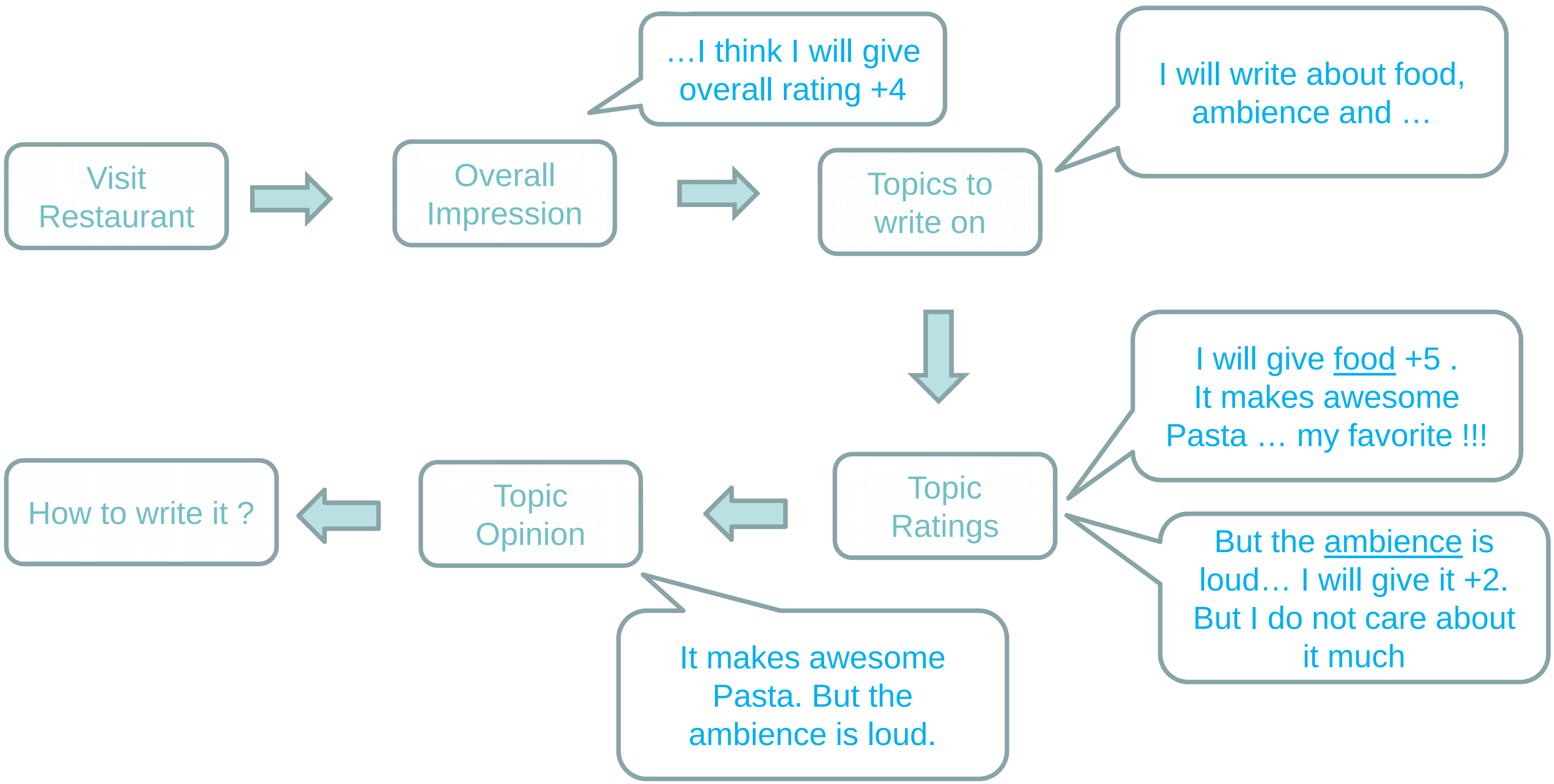
Generative Process for a Review



Generative Process for a Review



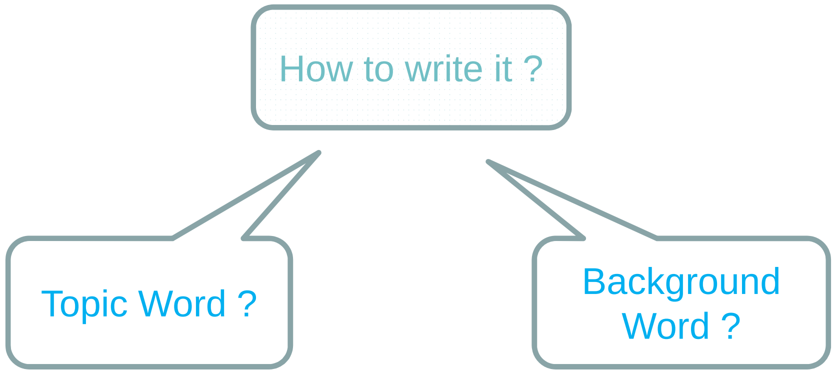
Generative Process for a Review



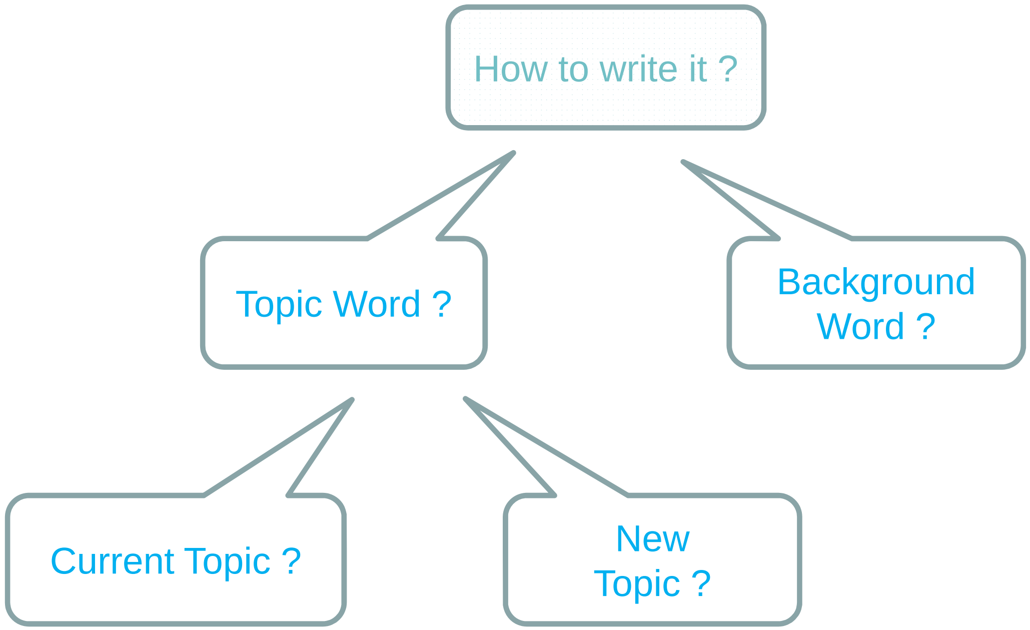
Generative Process for a Review

How to write it ?

Generative Process for a Review

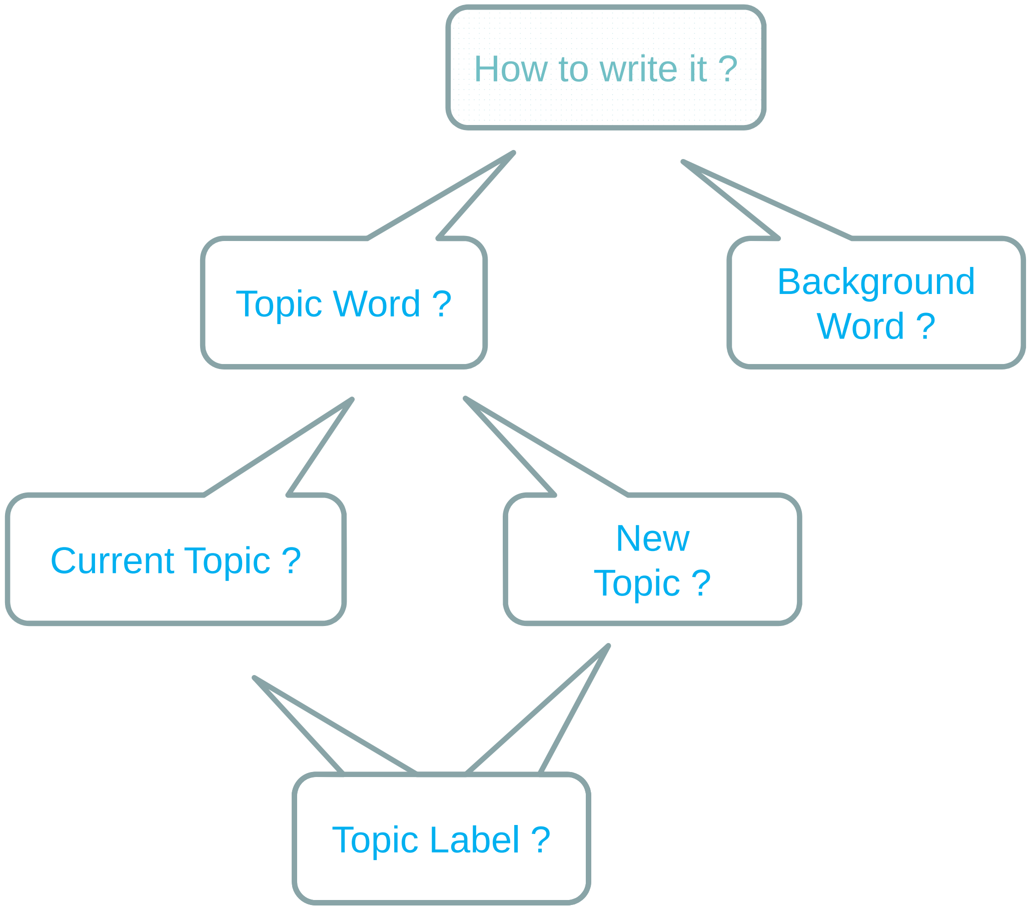


Generative Process for a Review

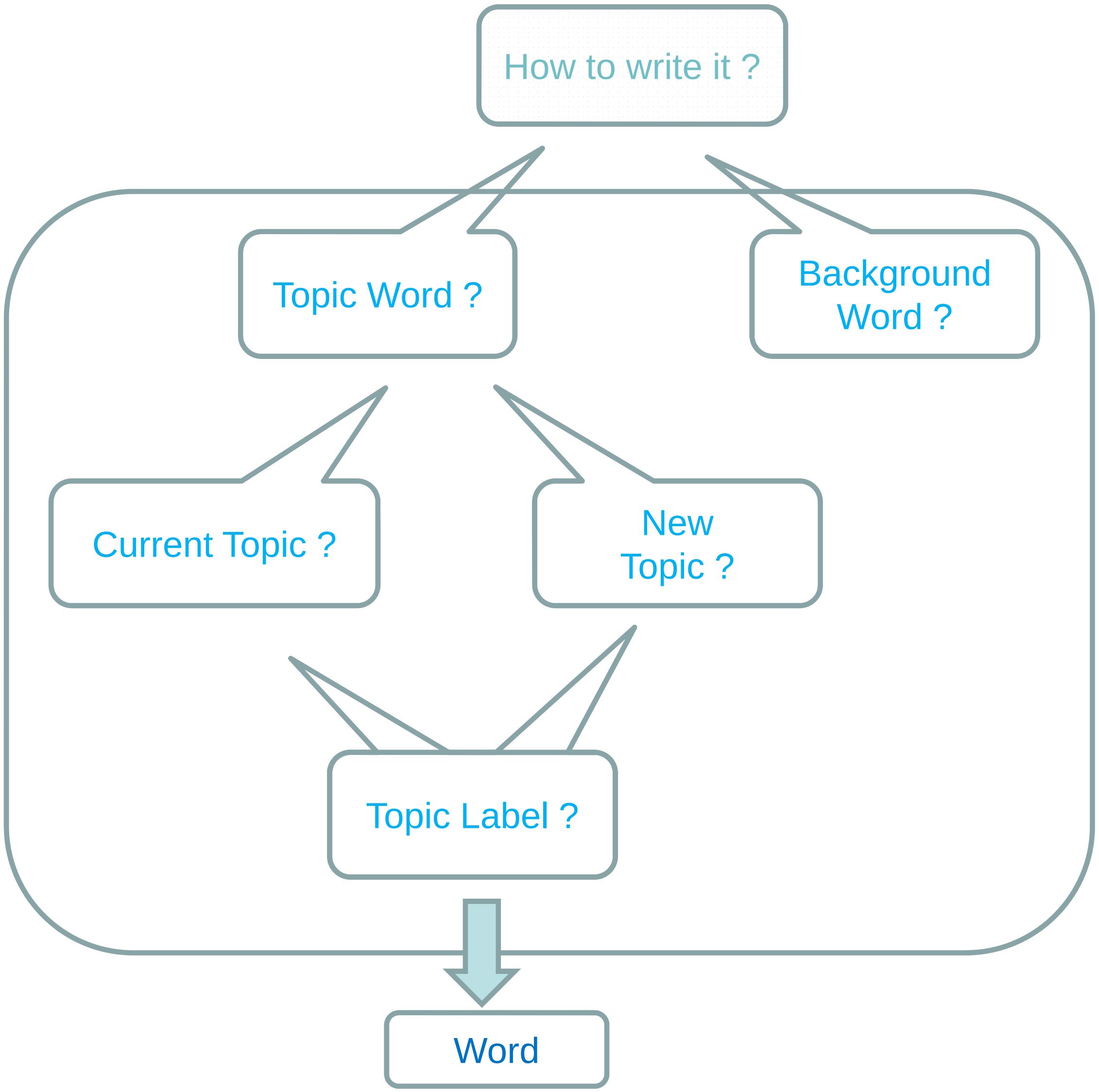


max planck institut informatik

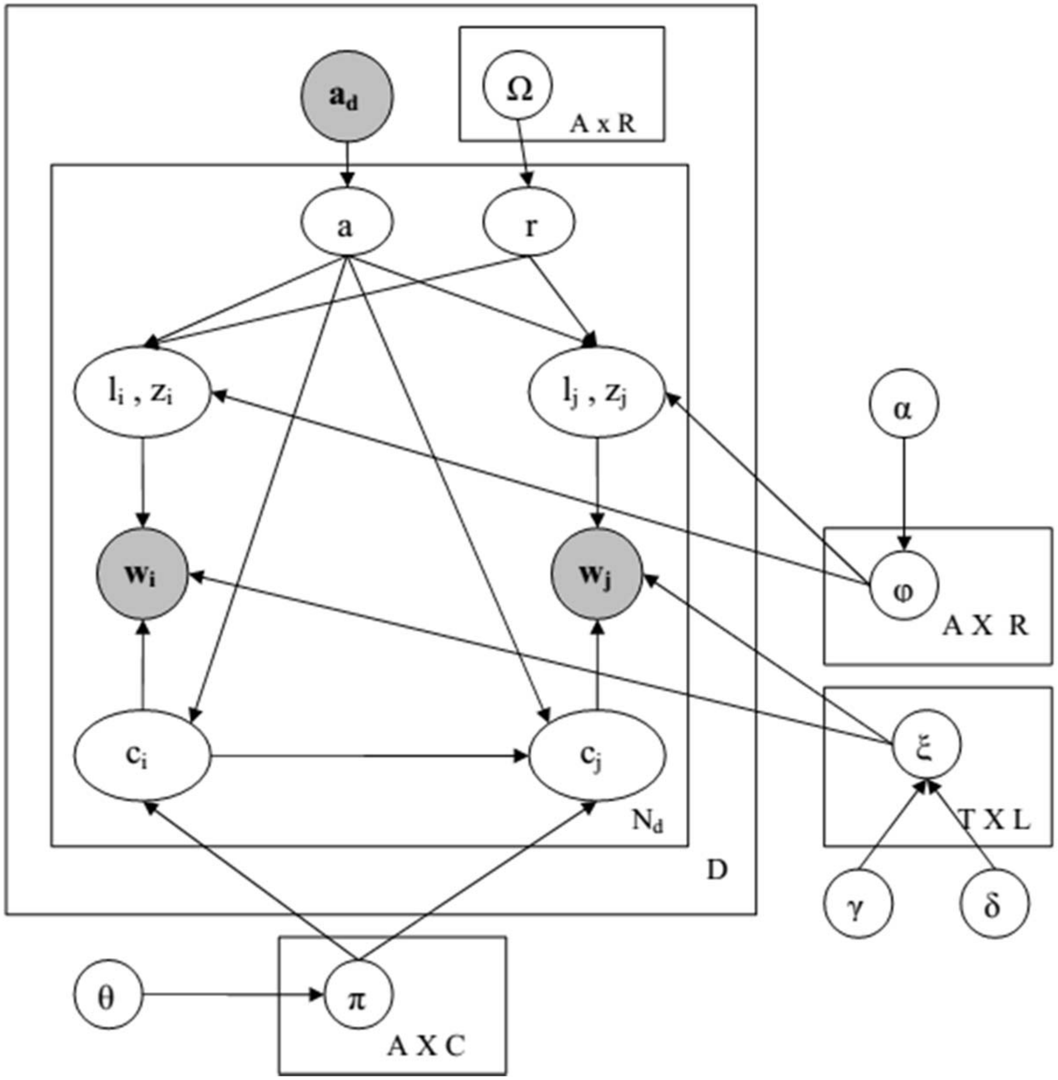
Generative Process for a Review



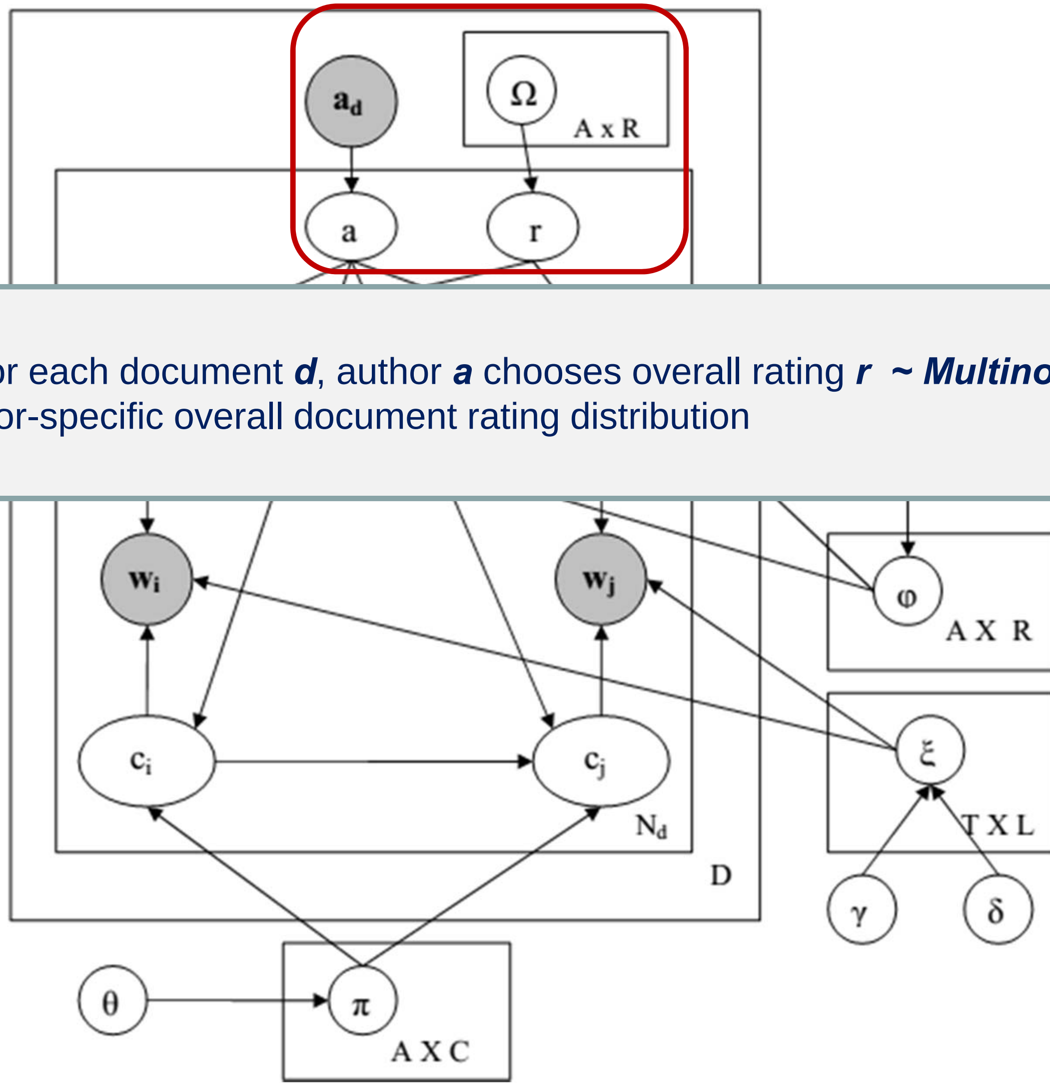
Generative Process for a Review



JAST Model



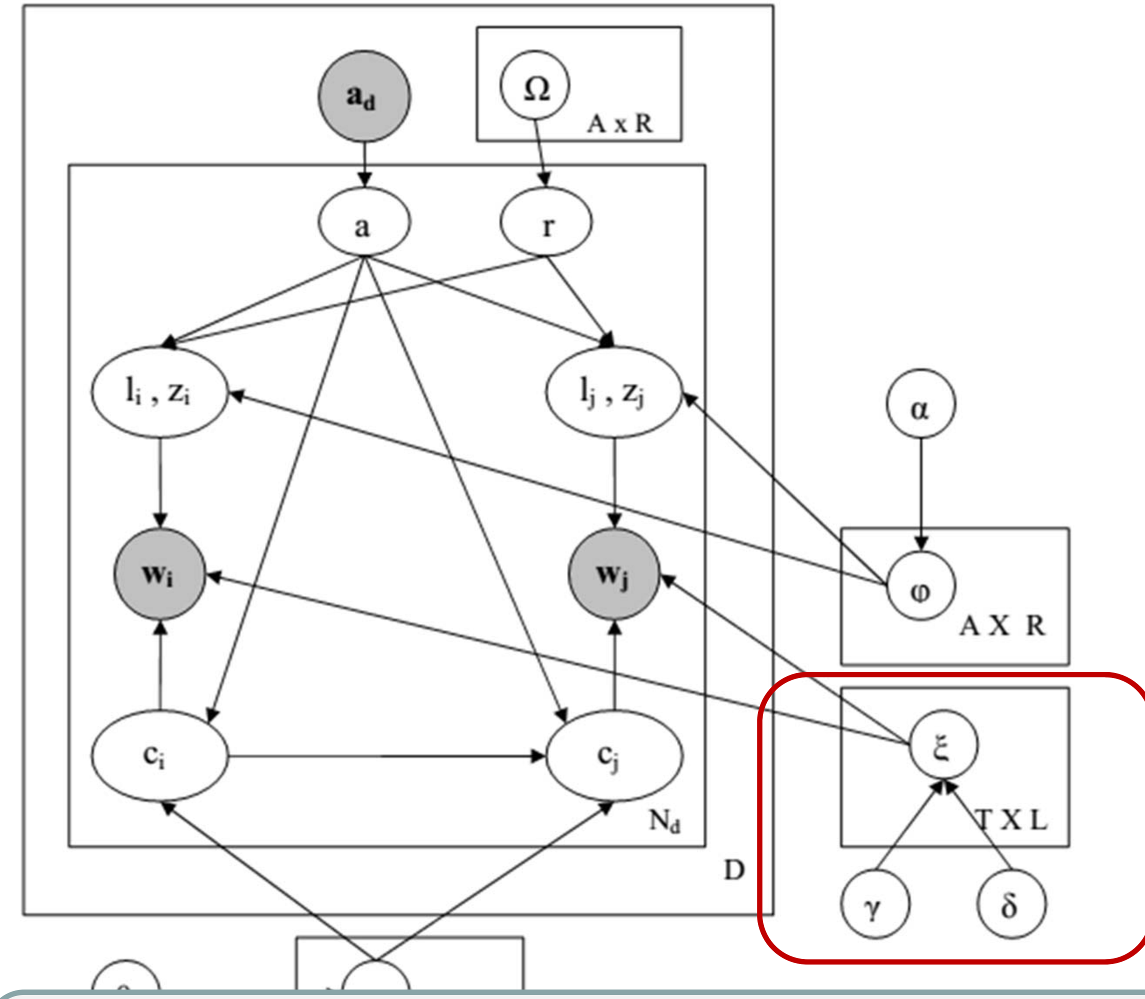
JAST Model



1. For each document ***d***, author ***a*** chooses overall rating ***r ~ Multinomial(Ω)*** from author-specific overall document rating distribution

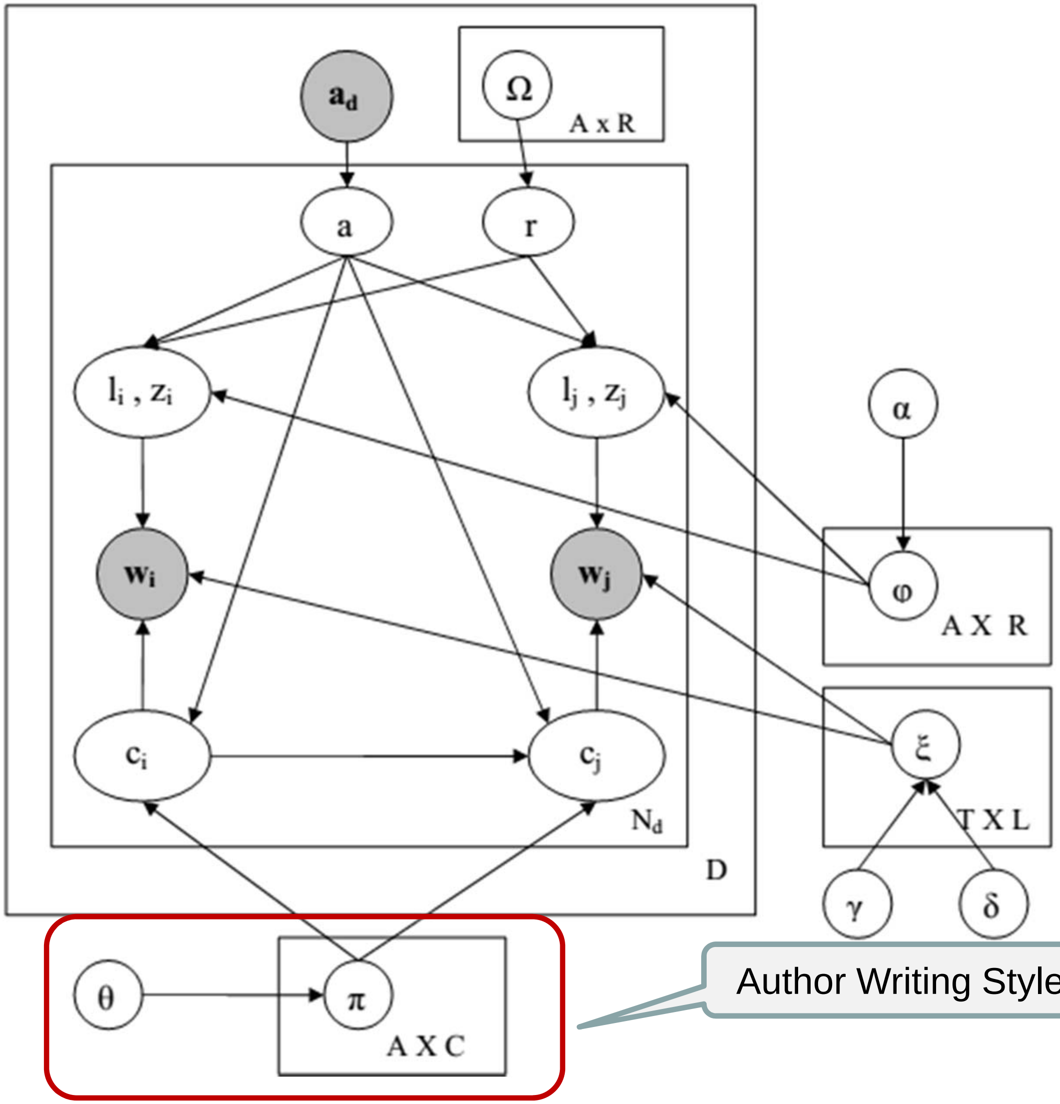
max planck institut
informatik

JAST Model



2. For each topic ***z*** and each sentiment label ***l***, draw ***$\xi_{z,l}$ ~ Dirichlet(γ)***
3. For each class ***c*** and each sentiment label ***l*** = 0, draw ***$\xi_{c,l}$ ~ Dirichlet(δ)***

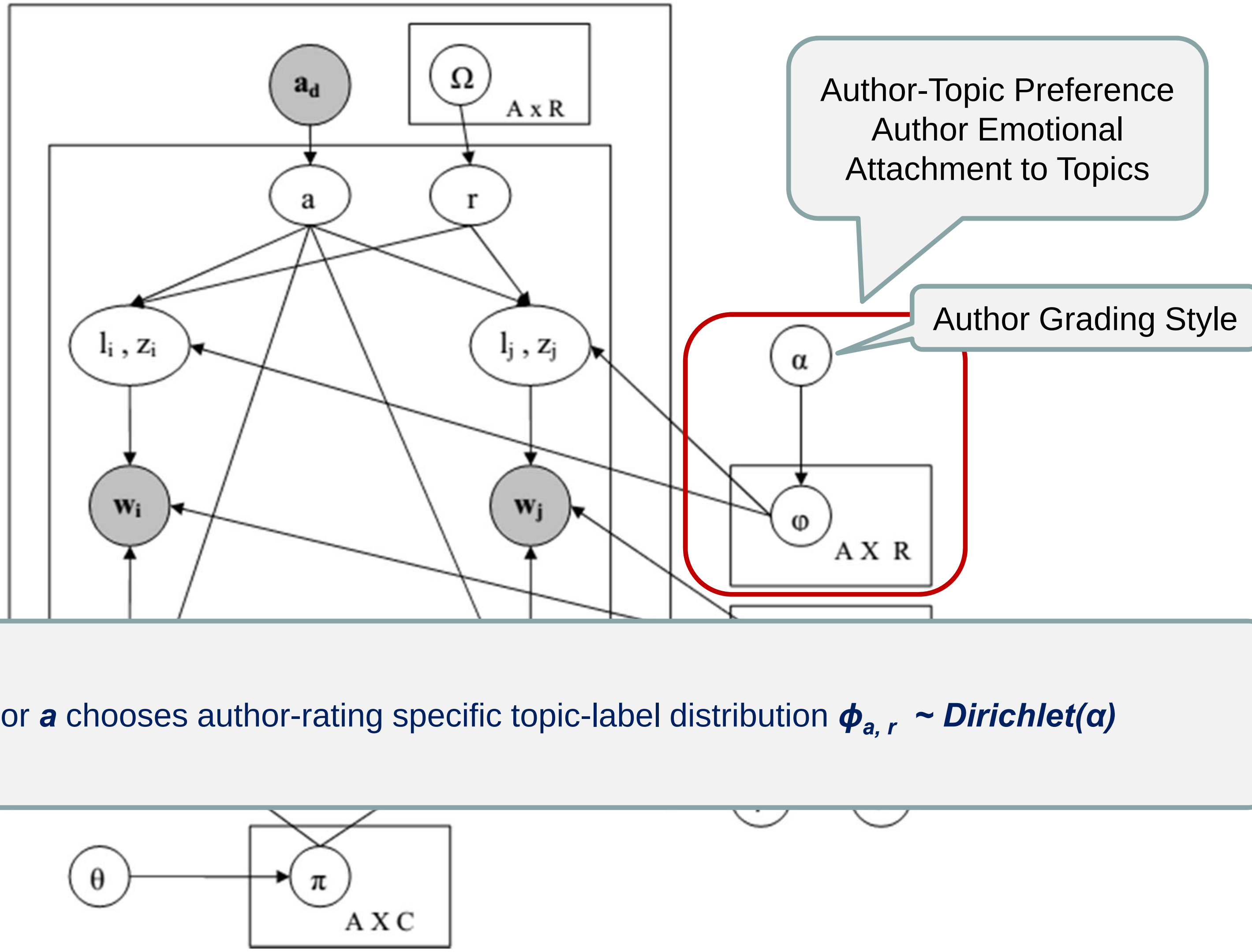
JAST Model



Author Writing Style

4. Choose author-specific class transition distribution ***π***

JAST Model

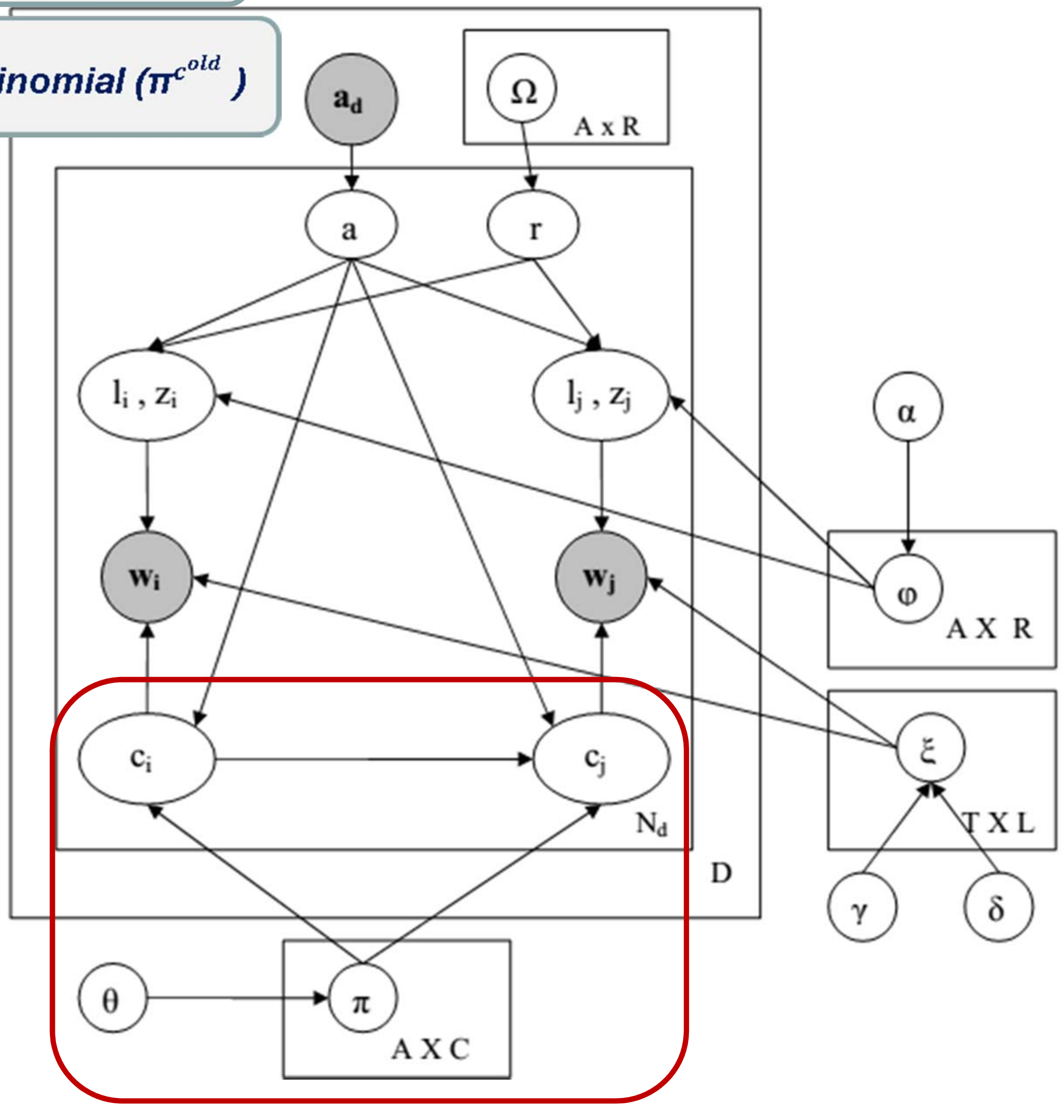


5. Author ***a*** chooses author-rating specific topic-label distribution $\boldsymbol{\phi_{a,\,r}} \sim \textbf{\textit{Dirichlet}}(\boldsymbol{\alpha})$

JAST Model

5. For each word ***w*** in the document

a. Draw ***c* ~ *Multinomial* ($\pi^{c^{old}}$)**

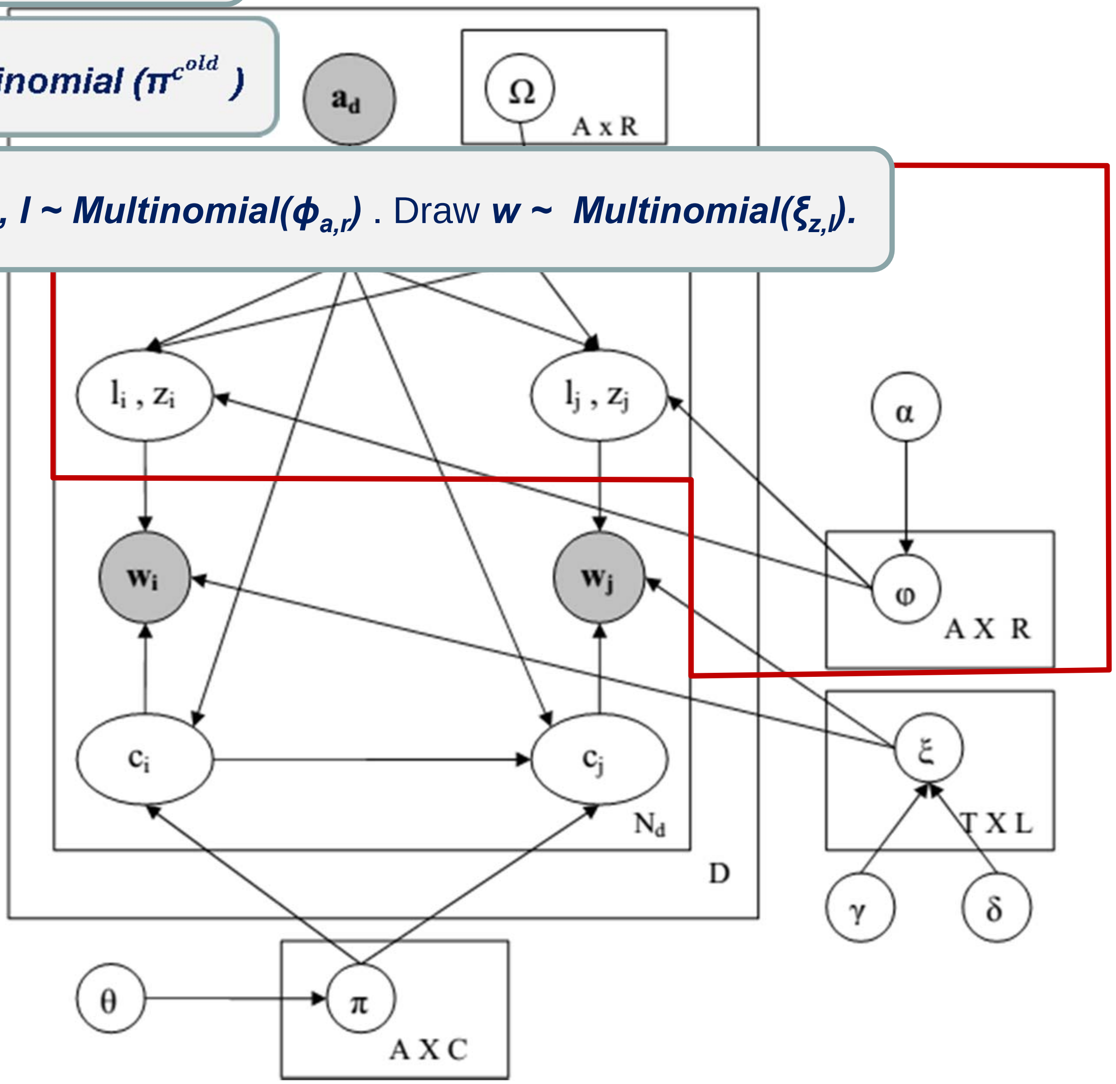


JAST Model

5. For each word ***w*** in the document

a. Draw ***c* ~ *Multinomial*** $(\pi^{c^{old}})$

b. If $c = 1$, Draw ***z, l* ~ *Multinomial***$(\phi_{a,r})$. Draw ***w* ~ *Multinomial***$(\xi_{z,l})$.



max planck institut
informatik

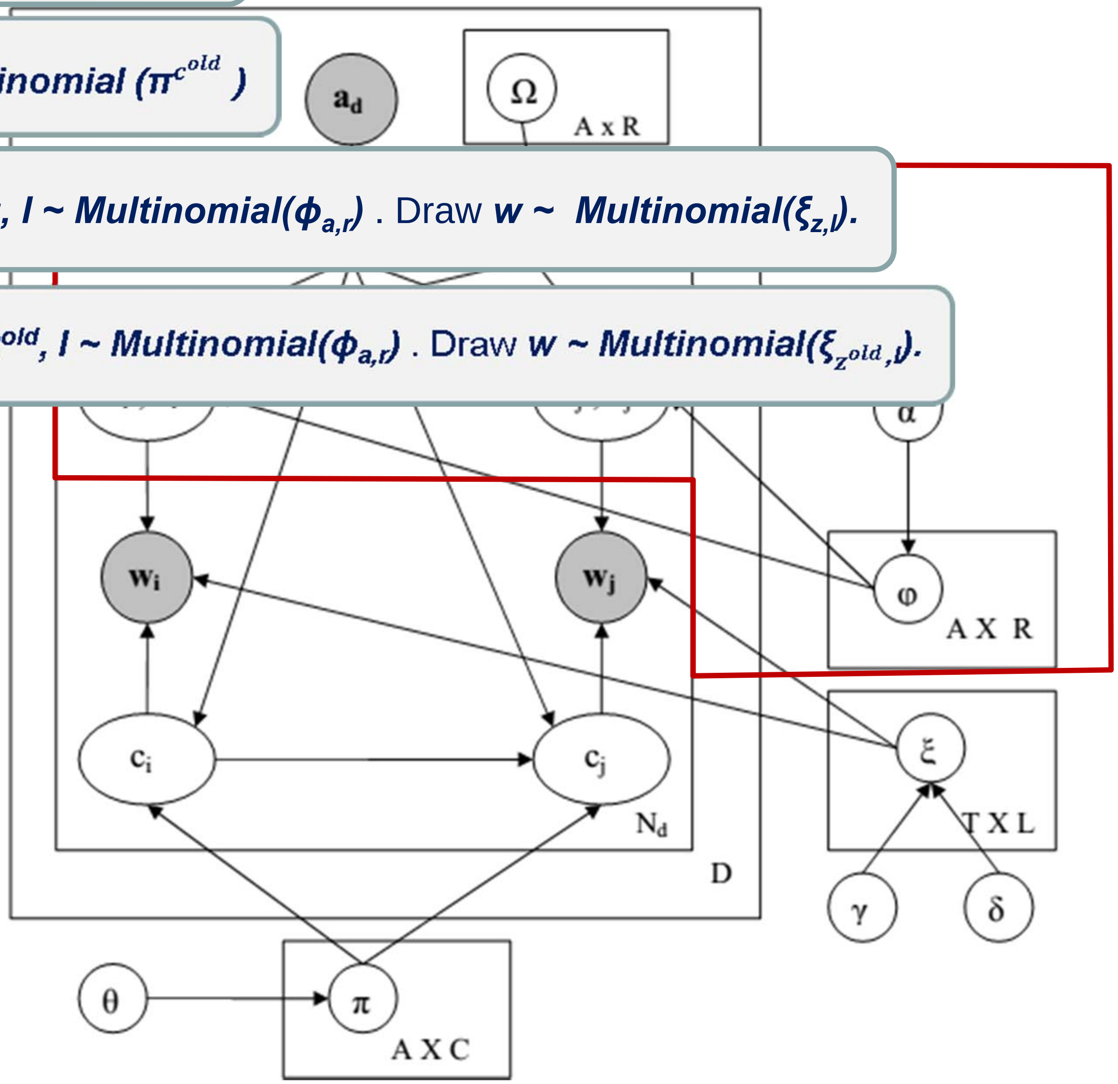
JAST Model

5. For each word ***w*** in the document

a. Draw $\boldsymbol{c} \sim \boldsymbol{Multinomial}(\pi^{c^{old}})$

b. If $c = 1$, Draw $\boldsymbol{z}, \boldsymbol{l} \sim \boldsymbol{Multinomial}(\phi_{a,r})$. Draw $\boldsymbol{w} \sim \boldsymbol{Multinomial}(\xi_{z,l})$.

c. If $c = 2$, Draw $\boldsymbol{z^{old}}, \boldsymbol{l} \sim \boldsymbol{Multinomial}(\phi_{a,r})$. Draw $\boldsymbol{w} \sim \boldsymbol{Multinomial}(\xi_{z^{old},l})$.



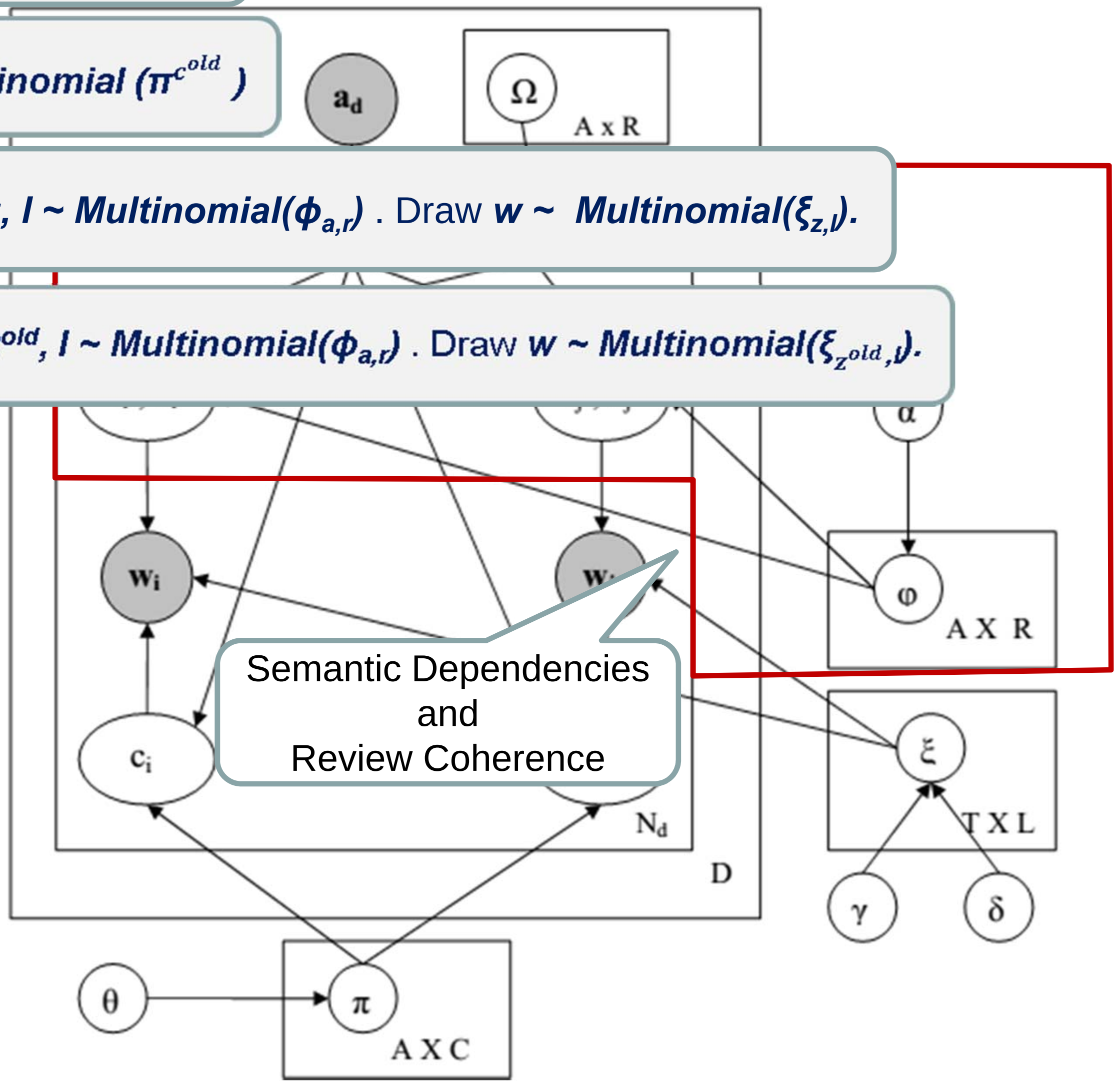
JAST Model

5. For each word ***w*** in the document

a. Draw ***c* ~ *Multinomial*** ($\pi^{c^{old}}$)

b. If $c = 1$, Draw ***z, l* ~ *Multinomial***($\phi_{a,r}$) . Draw ***w* ~ *Multinomial***($\xi_{z,l}$).

c. If $c = 2$, Draw z^{old}, l ~ ***Multinomial***($\phi_{a,r}$) . Draw ***w* ~ *Multinomial***($\xi_{z^{old},l}$).



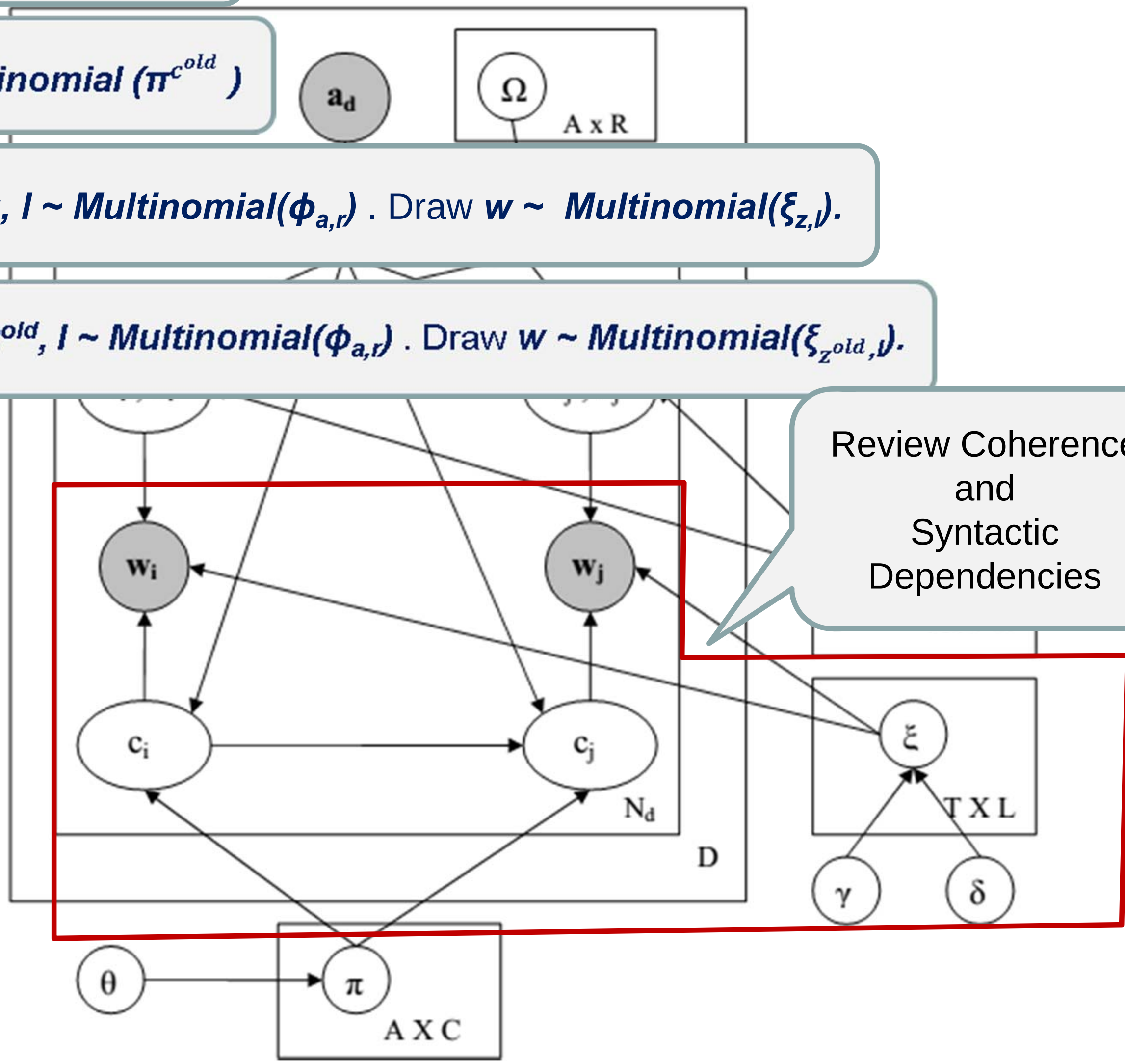
JAST Model

5. For each word ***w*** in the document

a. Draw ***c*** ~ ***Multinomial*** $(\pi^{c^{old}})$

b. If $c = 1$, Draw ***z, l*** ~ ***Multinomial***$(\phi_{a,r})$. Draw ***w*** ~ ***Multinomial***$(\xi_{z,l})$.

c. If $c = 2$, Draw z^{old}, l ~ ***Multinomial***$(\phi_{a,r})$. Draw ***w*** ~ ***Multinomial***$(\xi_{z^{old},l})$.



Review Coherence
and
Syntactic
Dependencies

JAST Model

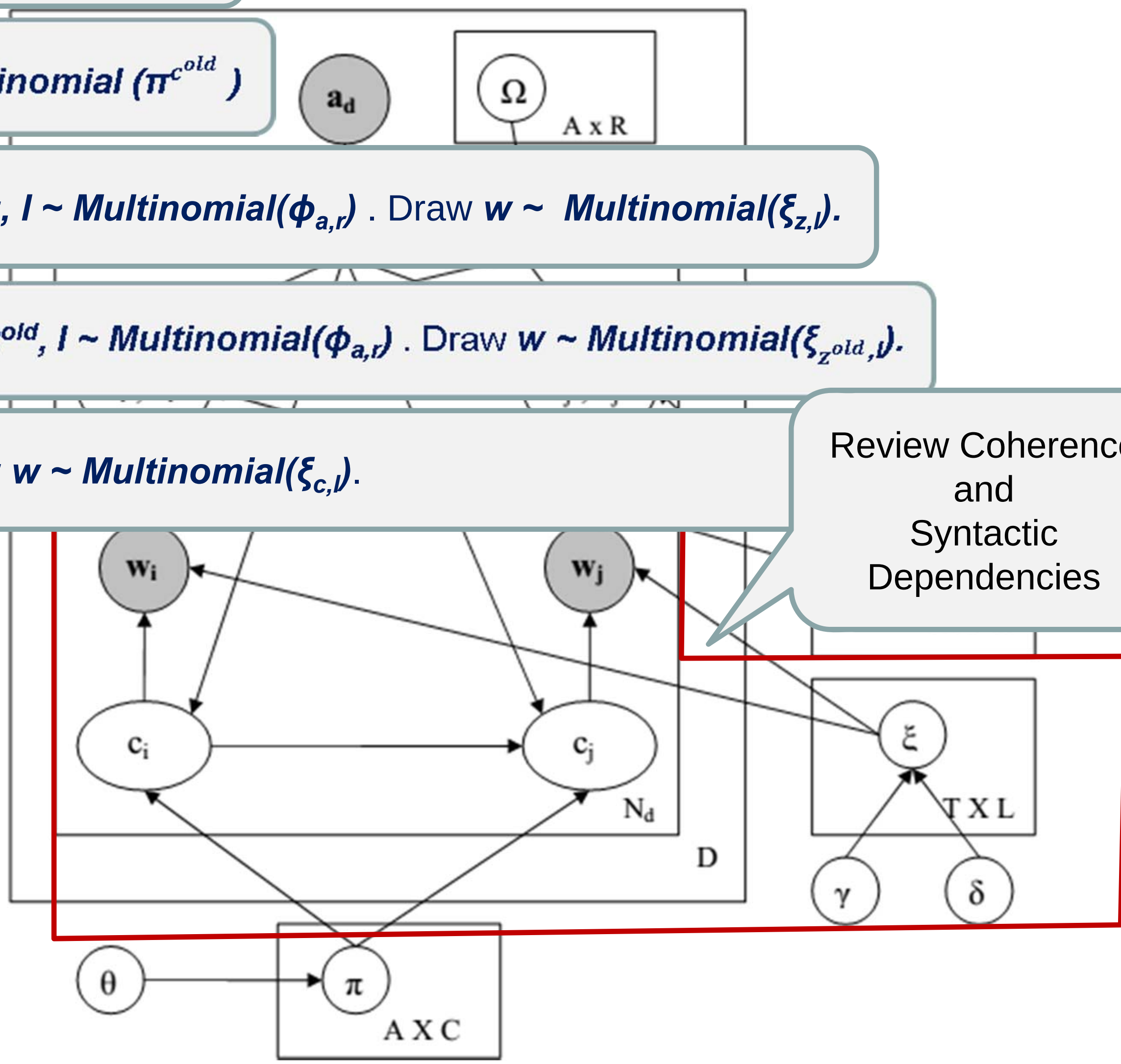
5. For each word ***w*** in the document

a. Draw ***c* ~ *Multinomial*** $(\pi^{c^{old}})$

b. If $c = 1$, Draw ***z, l* ~ *Multinomial*** $(\phi_{a,r})$. Draw ***w* ~ *Multinomial*** $(\xi_{z,l})$.

c. If $c = 2$, Draw z^{old}, ***l* ~ *Multinomial*** $(\phi_{a,r})$. Draw ***w* ~ *Multinomial*** $(\xi_{z^{old},l})$.

d. If $c \neq 1, 2$, Draw ***w* ~ *Multinomial*** $(\xi_{c,l})$.



Review Coherence and Syntactic Dependencies

max planck institut informatik

Inferencing

$A \rightarrow Authors$
$R \rightarrow Ratings$
$T \rightarrow Topics$
$L \rightarrow Topic\ Labels$
$C \rightarrow Classes$
$W \rightarrow Words$

$\Omega \rightarrow Author\ Rating\ Distribution$
$\phi \rightarrow Author\ Topic\ Label\ Distribution$
$\xi \rightarrow Topic\ Label\ Word\ Distribution$
$\pi \rightarrow Author\ Class\ Distribution$

Inferencing

$A \rightarrow Authors$
$R \rightarrow Ratings$
$T \rightarrow Topics$
$L \rightarrow Topic\ Labels$
$C \rightarrow Classes$
$W \rightarrow Words$

$\Omega \rightarrow Author\ Rating\ Distribution$
$\phi \rightarrow Author\ Topic\ Label\ Distribution$
$\xi \rightarrow Topic\ Label\ Word\ Distribution$
$\pi \rightarrow Author\ Class\ Distribution$

$$
\begin{gathered}
P(A, R, T, L, C, W, \Omega, \phi, \xi, \pi; \alpha, \beta, \gamma, \delta) = \\
P(\Omega) \times P(r|\Omega) \times \\
P(\phi; \alpha) \times P(z, l|\phi_r) \times \\
P(\pi; \beta) \times \mathrm{P}(\mathrm{c}|\pi) \times \\
P(\xi; \gamma, \delta) \times P(w|\xi_{z,l}, \xi_{c,l}, \pi_c)
\end{gathered}
$$

Inferencing

$A \rightarrow Authors$
$R \rightarrow Ratings$
$T \rightarrow Topics$
$L \rightarrow Topic\ Labels$
$C \rightarrow Classes$
$W \rightarrow Words$

$\Omega \rightarrow Author\ Rating\ Distribution$
$\phi \rightarrow Author\ Topic\ Label\ Distribution$
$\xi \rightarrow Topic\ Label\ Word\ Distribution$
$\pi \rightarrow Author\ Class\ Distribution$

$$
\begin{gathered}
P(A, R, T, L, C, W, \Omega, \phi, \xi, \pi; \alpha, \beta, \gamma, \delta) = \\
P(\Omega) \times P(r|\Omega) \times \\
P(\phi; \alpha) \times P(z, l|\phi_r) \times \\
P(\pi; \beta) \times \mathrm{P}(\mathrm{c}|\pi) \times \\
P(\xi; \gamma, \delta) \times P(w|\xi_{z\ l}, \xi_{c\ l}, \pi_c)
\end{gathered}
$$

Inferencing

$A \rightarrow Authors$
$R \rightarrow Ratings$
$T \rightarrow Topics$
$L \rightarrow Topic\ Labels$
$C \rightarrow Classes$
$W \rightarrow Words$

$\Omega \rightarrow Author\ Rating\ Distribution$
$\phi \rightarrow Author\ Topic\ Label\ Distribution$
$\xi \rightarrow Topic\ Label\ Word\ Distribution$
$\pi \rightarrow Author\ Class\ Distribution$

$$
\begin{gathered}
P(A, R, T, L, C, W, \Omega, \phi, \xi, \pi; \alpha, \beta, \gamma, \delta) = \\
\boxed{P(\Omega) \times P(r|\Omega) \times} \\
P(\phi; \alpha) \times P(z, l|\phi_r) \times \\
P(\pi; \beta) \times \mathrm{P}(\mathrm{c}|\pi) \times \\
P(\xi; \gamma, \delta) \times P(w|\xi_{z\_l}, \xi_{c\_l}, \pi_c)
\end{gathered}
$$

Inferencing

$A \rightarrow Authors$
$R \rightarrow Ratings$
$T \rightarrow Topics$
$L \rightarrow Topic\ Labels$
$C \rightarrow Classes$
$W \rightarrow Words$

$\Omega \rightarrow Author\ Rating\ Distribution$
$\phi \rightarrow Author\ Topic\ Label\ Distribution$
$\xi \rightarrow Topic\ Label\ Word\ Distribution$
$\pi \rightarrow Author\ Class\ Distribution$

$$
\begin{gathered}
P(A, R, T, L, C, W, \Omega, \phi, \xi, \pi; \alpha, \beta, \gamma, \delta) = \\
P(\Omega) \times P(r|\Omega) \times \\
\boxed{P(\phi; \alpha) \times P(z, l|\phi_r) \times} \\
P(\pi; \beta) \times \mathrm{P}(\mathrm{c}|\pi) \times \\
P(\xi; \gamma, \delta) \times P(w|\xi_{z\ l}, \xi_{c\ l}, \pi_c)
\end{gathered}
$$

Inferencing

$A \rightarrow Authors$
$R \rightarrow Ratings$
$T \rightarrow Topics$
$L \rightarrow Topic\ Labels$
$C \rightarrow Classes$
$W \rightarrow Words$

$\Omega \rightarrow Author\ Rating\ Distribution$
$\phi \rightarrow Author\ Topic\ Label\ Distribution$
$\xi \rightarrow Topic\ Label\ Word\ Distribution$
$\pi \rightarrow Author\ Class\ Distribution$

$$
\begin{gathered}
P(A, R, T, L, C, W, \Omega, \phi, \xi, \pi; \alpha, \beta, \gamma, \delta) = \\
P(\Omega) \times P(r|\Omega) \times \\
P(\phi; \alpha) \times P(z, l|\phi_r) \times \\
\boxed{P(\pi; \beta) \times \mathrm{P}(\mathrm{c}|\pi) \times} \\
P(\xi; \gamma, \delta) \times P(w|\xi_{z\ l}, \xi_{c\ l}, \pi_c)
\end{gathered}
$$

Inferencing

$A \rightarrow Authors$
$R \rightarrow Ratings$
$T \rightarrow Topics$
$L \rightarrow Topic\ Labels$
$C \rightarrow Classes$
$W \rightarrow Words$

$\Omega \rightarrow Author\ Rating\ Distribution$
$\phi \rightarrow Author\ Topic\ Label\ Distribution$
$\xi \rightarrow Topic\ Label\ Word\ Distribution$
$\pi \rightarrow Author\ Class\ Distribution$

$$P(A, R, T, L, C, W, \Omega, \phi, \xi, \pi; \alpha, \beta, \gamma, \delta) =$$
$$P(\Omega) \times P(r|\Omega) \times$$
$$P(\phi; \alpha) \times P(z, l|\phi_r) \times$$
$$P(\pi; \beta) \times \mathrm{P}(\mathrm{c}|\pi) \times$$
$$\boxed{P(\xi; \gamma, \delta) \times P(w|\xi_{z_l}, \xi_{c_l}, \pi_c)}$$

Inferencing

- We use collapsed Gibb's sampling for estimating the parameters
- Conditional distribution for joint updation of the latent variables is given by :

Inferencing

- We use collapsed Gibb's sampling for estimating the parameters
- Conditional distribution for joint updation of the latent variables is given by :

$$P(z = k, l = u, c = c \mid a, r, z_{-i}, l_{-i}, c_{-i}, w) \propto$$

$$\left(\#words_{a,r,k,u} + \alpha\right) \times \left(\#words_{w,k,u,1} + \gamma\right) \times \Omega_{a,r} \qquad if\ c = 1$$

$$\left(\#words_{a,r,k,u} + \alpha\right) \times \left(\#words_{w,k=k^{old},u,2} + \gamma\right) \times \Omega_{a,r} \qquad if\ c = 2$$

$$(\#words_{a,r,k,u} + \alpha) \times (\#words_{w,\,u=0,c} + \delta) \times \Omega_{a,r} \qquad if\ c \neq 1,2$$

Inferencing

- We use collapsed Gibb's sampling for estimating the parameters
- Conditional distribution for joint updation of the latent variables is given by :

$$P(z = k, l = u, c = c \mid a, r, z_{-i}, l_{-i}, c_{-i}, w) \propto$$

$$\left(\#words_{a,r,k,u} + \alpha\right) \times \left(\#words_{w,k,u,1} + \gamma\right) \times \Omega_{a,r} \quad if\ c = 1$$

$$\left(\#words_{a,r,k,u} + \alpha\right) \times \left(\#words_{w,k=k^{old},u,2} + \gamma\right) \times \Omega_{a,r} \quad if\ c = 2$$

$$(\#words_{a,r,k,u} + \alpha) \times (\#words_{w,\,u=0,c} + \delta) \times \Omega_{a,r} \quad if\ c \neq 1,2$$

Inferencing

- We use collapsed Gibb's sampling for estimating the parameters
- Conditional distribution for joint updation of the latent variables is given by :

$$P(z = k, l = u, c = c \mid a, r, z_{-i}, l_{-i}, c_{-i}, w) \propto$$

$$\left(\#words_{a,r,k,u} + \alpha\right) \times \left(\#words_{w,k,u,1} + \gamma\right) \times \Omega_{a,r} \qquad if\ c = 1$$

$$\left(\#words_{a,r,k,u} + \alpha\right) \times \left(\#words_{w,k=k^{old},u,2} + \gamma\right) \times \Omega_{a,r} \qquad if\ c = 2$$

$$(\#words_{a,r,k,u} + \alpha) \times (\#words_{w,\, u=0,c} + \delta) \times \Omega_{a,r} \qquad if\ c \neq 1,2$$

$\phi \rightarrow Author\ Topic\ Label\ Distribution$

Inferencing

- We use collapsed Gibb's sampling for estimating the parameters
- Conditional distribution for joint updation of the latent variables is given by :

$$P(z = k, l = u, c = c \mid a, r, z_{-i}, l_{-i}, c_{-i}, w) \propto$$

$$\boxed{(\#words_{a,r,k,u} + \alpha)} \times (\#words_{w,k,u,1} + \gamma) \times \Omega_{a,r} \quad if\ c = 1$$

$$\boxed{(\#words_{a,r,k,u} + \alpha)} \times (\#words_{w,k=k^{old},u,2} + \gamma) \times \Omega_{a,r} \quad if\ c = 2$$

$$\boxed{(\#words_{a,r,k,u} + \alpha)} \times (\#words_{w,\, u=0,c} + \delta) \times \Omega_{a,r} \quad if\ c \neq 1,2$$

$\phi \rightarrow Author\ Topic\ Label\ Distribution$

Inferencing

- We use collapsed Gibb's sampling for estimating the parameters
- Conditional distribution for joint updation of the latent variables is given by :

$$P(z = k, l = u, c = c \mid a, r, z_{\_i}, l_{\_i}, c_{\_i}, w) \propto$$

$$\boxed{\begin{aligned}
&(\#words_{a,r,k,u} + \alpha) \times (\#words_{w,k,u,1} + \gamma) \times \Omega_{a,r} && if\ c = 1 \\
&(\#words_{a,r,k,u} + \alpha) \times (\#words_{w,k=k^{old},u,2} + \gamma) \times \Omega_{a,r} && if\ c = 2 \\
&(\#words_{a,r,k,u} + \alpha) \times (\#words_{w,\,u=0,c} + \delta) \times \Omega_{a,r} && if\ c \neq 1,2
\end{aligned}}$$

$$\xi \rightarrow Topic\ Label\ Word\ Distribution$$

Inferencing

- We use collapsed Gibb's sampling for estimating the parameters
- Conditional distribution for joint updation of the latent variables is given by :

$$P(z = k, l = u, c = c \mid a, r, z_{\_i}, l_{\_i}, c_{\_i}, w) \propto$$

$$\boxed{\left(\#words_{a,r,k,u} + \alpha\right) \times \left(\#words_{w,k,u,1} + \gamma\right) \times \Omega_{a,r} \quad if\ c = 1}$$

$$\left(\#words_{a,r,k,u} + \alpha\right) \times \left(\#words_{w,k=k^{old},u,2} + \gamma\right) \times \Omega_{a,r} \quad if\ c = 2$$

$$(\#words_{a,r,k,u} + \alpha) \times (\#words_{w,\,u=0,c} + \delta) \times \Omega_{a,r} \quad if\ c \neq 1{,}2$$

Inferencing

- We use collapsed Gibb's sampling for estimating the parameters
- Conditional distribution for joint updation of the latent variables is given by :

$$P(z = k, l = u, c = c \mid a, r, z_{\_i}, l_{\_i}, c_{\_i}, w) \propto$$

$$\left(\#words_{a,r,k,u} + \alpha\right) \times \left(\#words_{w,k,u,1} + \gamma\right) \times \Omega_{a,r} \qquad if\ c = 1$$

$$\left(\#words_{a,r,k,u} + \alpha\right) \times \left(\#words_{w,k=k^{old},u,2} + \gamma\right) \times \Omega_{a,r} \qquad if\ c = 2$$

$$\boxed{(\#words_{a,r,k,u} + \alpha) \times (\#words_{w,\,u=0,c} + \delta) \times \Omega_{a,r} \qquad if\ c \neq 1{,}2}$$

Inferencing

$$P(c = c|a, z = k, l = u, c_{\_i}, w) \propto$$
$$f_1(\xi) \times P(c|c^{old}, a) \; if \; c = 1$$
$$f_2(\xi) \times P(c|c^{old}, a) \; if \; c = 2$$
$$f_3(\xi) \times P(c|c^{old}, a) \; if \; c \neq 1,2$$

Inferencing

$$P(c = c | a, z = k, l = u, c_{\_i}, w) \propto$$
$$f_1(\xi) \times P(c|c^{old}, a)\ if\ c = 1$$
$$f_2(\xi) \times P(c|c^{old}, a)\ if\ c = 2$$
$$f_3(\xi) \times P(c|c^{old}, a)\ if\ c \neq 1{,}2$$

given an unseen review ***r*** and its author ***a***

for each word ***w*** in the review

its topic and topic-rating ***(k, u)*** are extracted from $\boldsymbol{\xi_{T \times L}[w]}$

review rating is given by $argmax_r \Omega_{a,r}$

$$\Omega_{a,r} = \frac{\Sigma_{k,u} \mathrm{I}(r = argmax_{r^*} \phi_{a,r^*}[k,u]) \times \phi_{a,r}[k,u]}{K}$$

Inferencing

$$P(c = c | a, z = k, l = u, c_{\_i}, w) \propto$$
$$f_1(\xi) \times P(c|c^{old}, a) \ if \ c = 1$$
$$f_2(\xi) \times P(c|c^{old}, a) \ if \ c = 2$$
$$f_3(\xi) \times P(c|c^{old}, a) \ if \ c \neq 1{,}2$$

given an unseen review ***r*** and its author ***a***

for each word ***w*** in the review

its topic and topic-rating ***(k, u)*** are extracted from $\boldsymbol{\xi_{T \times L}[w]}$

review rating is given by $argmax_r \Omega_{a,r}$

$$\Omega_{a,r} = \frac{\Sigma_{k,u} \mathrm{I}(r{=}argmax_{r^*} \phi_{a,r^*}[k,u]) \times \phi_{a,r}[k,u]}{K}$$

Dataset for Evaluation

- IMDB movie review dataset
- TripAdvisor restaurant review dataset

Dataset	Authors	Avg Rev/ Author	Rev/ Rating						Avg Rev Length	Avg Words/ Rev
Movie Review*	312	7	**Pos** 1000		**Neg** 1000			**Total** 2000	32	746
Movie Review⊥	65	23	**Pos** 705		**Neg** 762			**Total** 1467	32	711
Restaurant Review*	9	170	**R 1** 43	**R2** 134	**R 3** 501	**R 4** 612	**R 5** 237	**Total** 1526	16	71
Restaurant Review⊥	9	340	**R 1** 514	**R 2** 532	**R 3** 680	**R 4** 700	**R 5** 626	**Total** 3052	20	81

Baselines

- Lexical classification using majority voting
- Joint Sentiment Topic Model[1]
- Author-Topic LR Model[2]

- Model Prior
  - A sentiment lexicon is used to initialize the prior polarity of words in $\xi_{T \times L}[w]$

1. Chenghua Lin and Yulan He, Joint sentiment/topic model for sentiment analysis, CIKM '09, pp. 375-384.
2. Subhabrata Mukherjee, Gaurab Basu, and Sachindra Joshi, Incorporating author preference in sentiment rating prediction of reviews, WWW 2013.

Model Initialization Parameters

Model Parameters	Movie Review	Restaurant Review
A	65	9
R	2	5
T	50	25
L	3	3
C	20	15
$\alpha = 1/T \times L$	0.007	0.013
$\gamma = 1/T \times L$	0.007	0.013
$\delta = 1/C \times L$	0.017	0.022
$\theta = 1/A \times C$	0.0007	0.007

Model Initialization Parameters

Model Parameters	Movie Review	Restaurant Review
A	65	9
R	2	5
T	50	25
L	3	3
C	20	15
$\alpha = 1/T \times L$	0.007	0.013
$\gamma = 1/T \times L$	0.007	0.013
$\delta = 1/C \times L$	0.017	0.022
$\theta = 1/A \times C$	0.0007	0.007

Model Initialization Parameters

Model Parameters	Movie Review	Restaurant Review
A	65	9
R	2	5
T	50	25
L	3	3
C	20	15
$\alpha = 1/T \times L$	0.007	0.013
$\gamma = 1/T \times L$	0.007	0.013
$\delta = 1/C \times L$	0.017	0.022
$\theta = 1/A \times C$	0.0007	0.007

Minimize Model Perplexity

Model Comparison with Baselines

Model Comparison with Baselines

Models	**Accuracy**
Lexical Baseline	65
JST [9]	82.8
Mukherjee *et al.* (2013) [12]	84.39
JAST	**87.69**

IMDB Movie Review Dataset

Model Comparison with Baselines

Models	Accuracy
Lexical Baseline	65
JST [9]	82.8
Mukherjee *et al.* (2013) [12]	84.39
JAST	**87.69**

IMDB Movie Review Dataset

Models	MAE
Lexical Baseline (Hu et. al 2004)	1.24
JST [9]	1.01
Facet Specific General Author Preference [12]	0.75
Facet and Author Specific Preference [12]	0.71
JAST	**0.61**

TripAdvisor Restaurant Review Dataset

Comparison with Top Performing Models in IMDB Dataset

Models	**Acc.**
Eigen Vector Clustering [2]	70.9
Semi Supervised, 40% doc. Label [8]	73.5
LSM Unsupervised with prior info [10]	74.1
SO-CAL Full Lexicon [21]	76.37
RAE Semi Supervised Recursive Auto Encoders with random word initialization [20]	76.8
WikiSent: Extractive Summarization with Wikipedia + Lexicon [13]	76.85
Supervised Tree-CRF [14]	77.3
RAE: Supervised Recursive Auto Encoders with 10% cross-validation [20]	77.7
JST: Without Subjectivity Detection using LDA [9]	82.8
JST: With Subjectivity Detection [9]	84.6
Pang *et al.* (2002): Supervised SVM [16]	82.9
Supervised Subjective MR, SVM [15]	87.2
Kennedy *et al.* (2006): Supervised SVM [6]	86.2
Appraisal Group: Supervised [25]	90.2
JAST: Unsupervised HMM-LDA	**87.69**

Comparison with Top Performing Models in IMDB Dataset

Models	Acc.
Eigen Vector Clustering [2]	70.9
Semi Supervised, 40% doc. Label [8]	73.5
LSM Unsupervised with prior info [10]	74.1
SO-CAL Full Lexicon [21]	76.37
RAE Semi Supervised Recursive Auto Encoders with random word initialization [20]	76.8
WikiSent: Extractive Summarization with Wikipedia + Lexicon [13]	76.85
Supervised Tree-CRF [14]	77.3
RAE: Supervised Recursive Auto Encoders with 10% cross-validation [20]	77.7
JST: Without Subjectivity Detection using LDA [9]	82.8
JST: With Subjectivity Detection [9]	84.6
Pang *et al.* (2002): Supervised SVM [16]	82.9
Supervised Subjective MR, SVM [15]	87.2
Kennedy *et al.* (2006): Supervised SVM [6]	86.2
Appraisal Group: Supervised [25]	90.2
JAST: Unsupervised HMM-LDA	**87.69**

Comparison with Top Performing Models in IMDB Dataset

Models	Acc.
Eigen Vector Clustering [2]	70.9
Semi Supervised, 40% doc. Label [8]	73.5
LSM Unsupervised with prior info [10]	74.1
SO-CAL Full Lexicon [21]	76.37
RAE Semi Supervised Recursive Auto Encoders with random word initialization [20]	76.8
WikiSent: Extractive Summarization with Wikipedia + Lexicon [13]	76.85
Supervised Tree-CRF [14]	77.3
RAE: Supervised Recursive Auto Encoders with 10% cross-validation [20]	77.7
JST: Without Subjectivity Detection using LDA [9]	82.8
JST: With Subjectivity Detection [9]	84.6
Pang *et al.* (2002): Supervised SVM [16]	82.9
Supervised Subjective MR, SVM [15]	87.2
Kennedy *et al.* (2006): Supervised SVM [6]	86.2
Appraisal Group: Supervised [25]	90.2
JAST: Unsupervised HMM-LDA	**87.69**

Snapshot of Topic-Label-Word Extraction by JAST

Movie Review Dataset					Restaurant Review Dataset				
T=bad	T=good	T=actor	T=actor	T= actor	T=food	T=food	T=food	T=service	T=bad
L=neg	L=pos	L=neg	L=pos	L=obj	L=obj	L=neg	L=pos	L=pos	L=neg
bad	good	kevin	funny	cruise	food	bad	dish	ambience	average
suppose	great	violence	comedy	name	diner	awful	price	face	noth
bore	sometimes	comic	laugh	run	customer	seem	din	hearty	wasn
unfortunate	different	early	joke	ship	sweet	just	first	pretty	bad
stupid	hunt	someth	fun	group	kitchen	cheap	beautiful	exceptional	basic
waste	truman	not	eye	patch	feel	wasn	chicken	diner	nor
ridiculous	sean	long	talk	creature	meal	stop	quality	friendly	didn
half	excellent	every	hour	tribe	front	cold	recommend	perfection	don
terrible	relationship	support	act	big	home	quite	lovely	help	last
lame	amaze	type	moment	rise	serve	small	taste	worth	probably
dull	damon	somewhat	close	board	warm	loud	fun	extra	slow
poorly	martin	question	scene	studio	waitress	no	available	effort	sometimes
attempt	chemistry	fall	picture	sink	treat	common	definitely	warm	serious

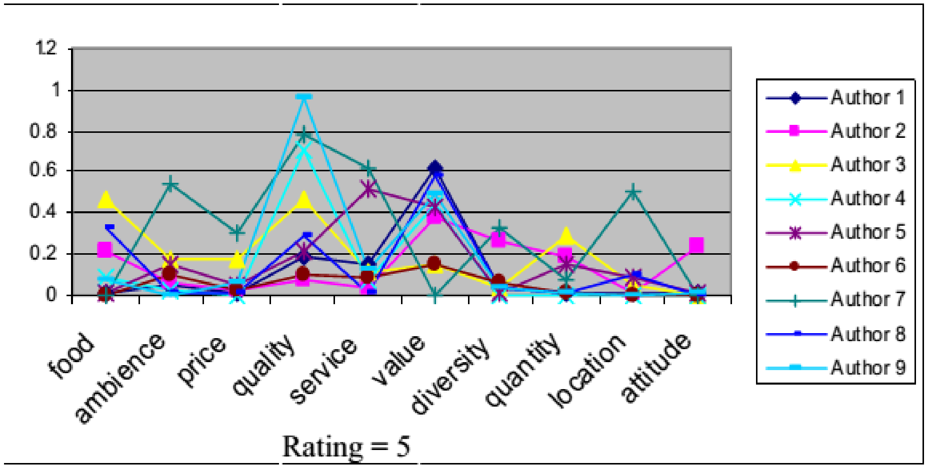
Snapshot of Topic-Label-Word Extraction by JAST

Movie Review Dataset					Restaurant Review Dataset				
T=bad	T=good	T=actor	T=actor	T= actor	T=food	T=food	T=food	T=service	T=bad
L=neg	L=pos	L=neg	L=pos	L=obj	L=obj	L=neg	L=pos	L=pos	L=neg
bad	good	kevin	funny	cruise	food	bad	dish	ambience	average
suppose	great	violence	comedy	name	diner	awful	price	face	noth
bore	sometimes	comic	laugh	run	customer	seem	din	hearty	wasn
unfortunate	different	early	joke	ship	sweet	just	first	pretty	bad
stupid	hunt	someth	fun	group	kitchen	cheap	beautiful	exceptional	basic
waste	truman	not	eye	patch	feel	wasn	chicken	diner	nor
ridiculous	sean	long	talk	creature	meal	stop	quality	friendly	didn
half	excellent	every	hour	tribe	front	cold	recommend	perfection	don
terrible	relationship	support	act	big	home	quite	lovely	help	last
lame	amaze	type	moment	rise	serve	small	taste	worth	probably
dull	damon	somewhat	close	board	warm	loud	fun	extra	slow
poorly	martin	question	scene	studio	waitress	no	available	effort	sometimes
attempt	chemistry	fall	picture	sink	treat	common	definitely	warm	serious

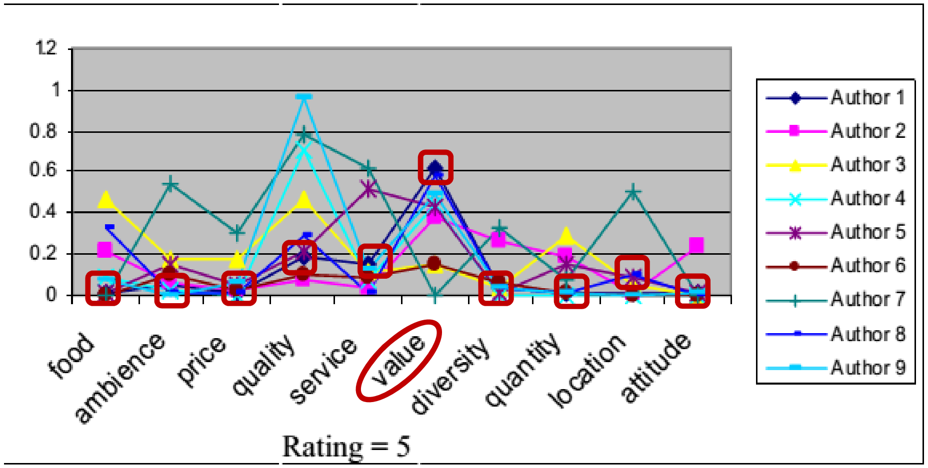
Snapshot of Topic-Label-Word Extraction by JAST

Movie Review Dataset						Restaurant Review Dataset			
T=bad	T=good	T=actor	T=actor	T= actor	T=food	T=food	T=food	T=service	T=bad
L=neg	L=pos	L=neg	L=pos	L=obj	L=obj	L=neg	L=pos	L=pos	L=neg
bad	good	kevin	funny	cruise	food	bad	dish	ambience	average
suppose	great	violence	comedy	name	diner	awful	price	face	noth
bore	sometimes	comic	laugh	run	customer	seem	din	hearty	wasn
unfortunate	different	early	joke	ship	sweet	just	first	pretty	bad
stupid	hunt	someth	fun	group	kitchen	cheap	beautiful	exceptional	basic
waste	truman	not	eye	patch	feel	wasn	chicken	diner	nor
ridiculous	sean	long	talk	creature	meal	stop	quality	friendly	didn
half	excellent	every	hour	tribe	front	cold	recommend	perfection	don
terrible	relationship	support	act	big	home	quite	lovely	help	last
lame	amaze	type	moment	rise	serve	small	taste	worth	probably
dull	damon	somewhat	close	board	warm	loud	fun	extra	slow
poorly	martin	question	scene	studio	waitress	no	available	effort	sometimes
attempt	chemistry	fall	picture	sink	treat	common	definitely	warm	serious

Snapshot of Author-Rating-Topic-Label Distribution Extracted by JAST - TripAdvisor

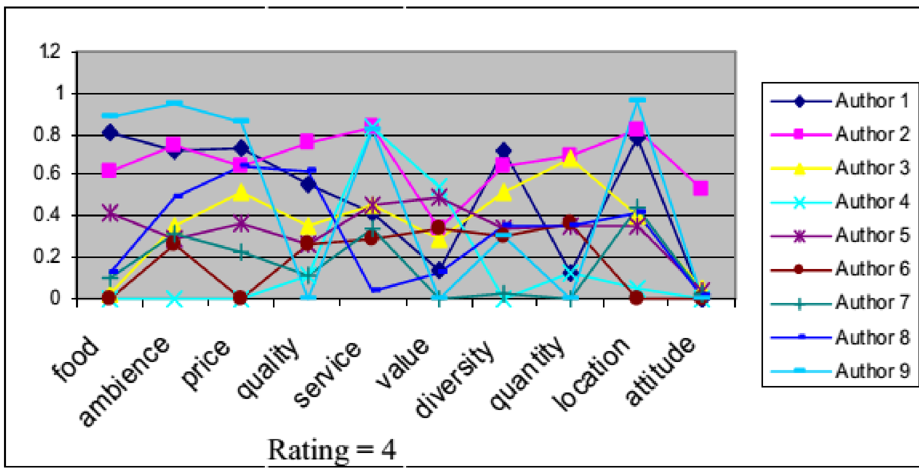


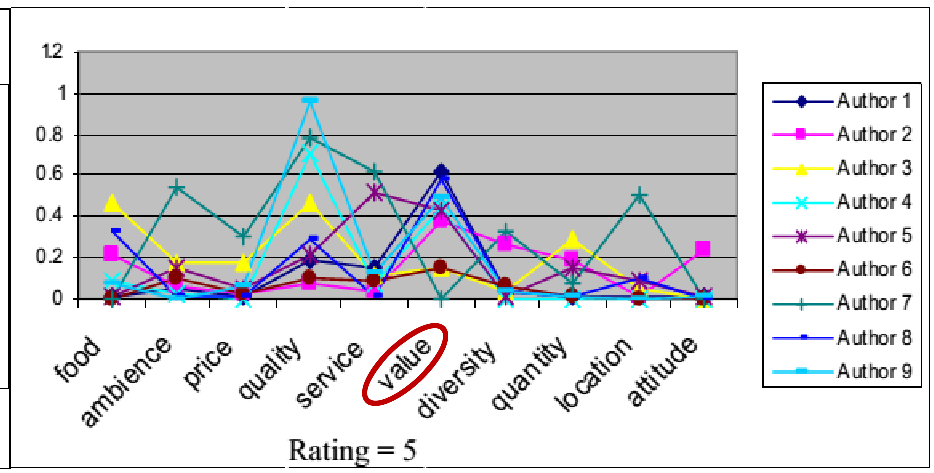
Snapshot of Author-Rating-Topic-Label Distribution Extracted by JAST - TripAdvisor



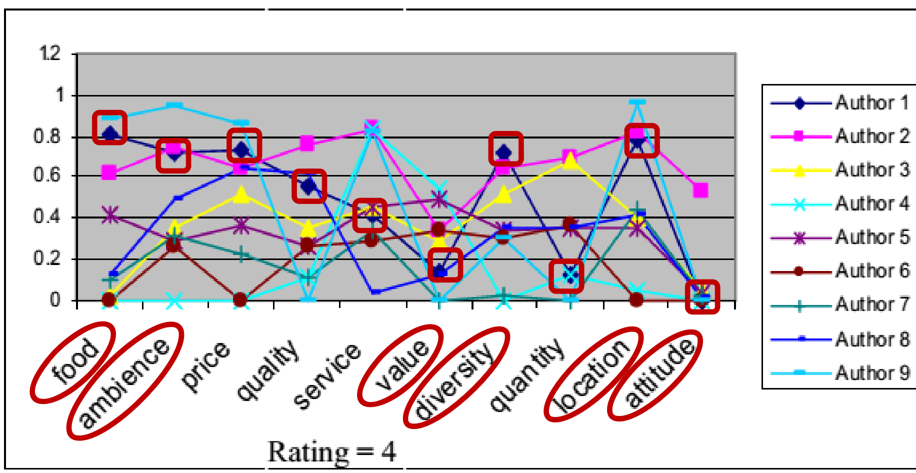
max planck institut
informatik

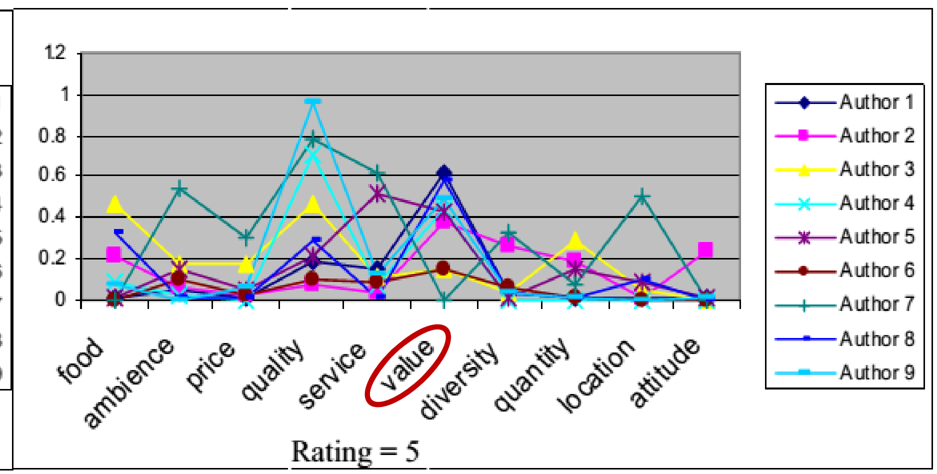
Snapshot of Author-Rating-Topic-Label Distribution Extracted by JAST - TripAdvisor





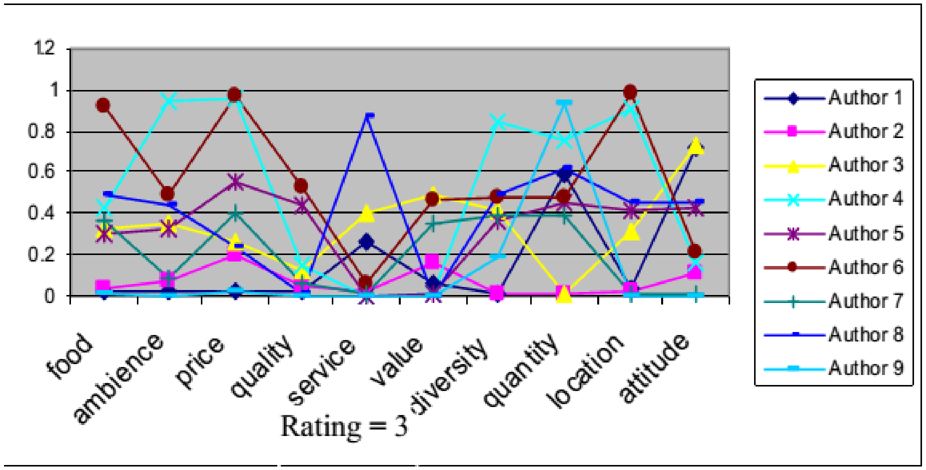
Snapshot of Author-Rating-Topic-Label Distribution Extracted by JAST - TripAdvisor

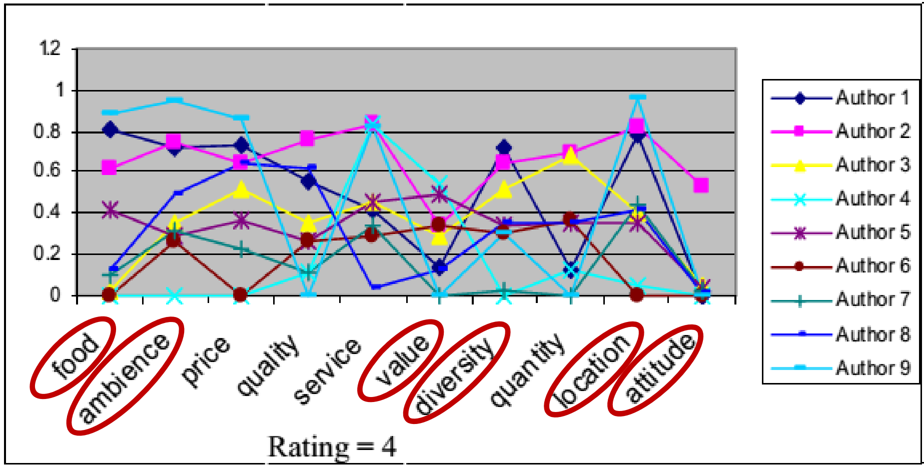


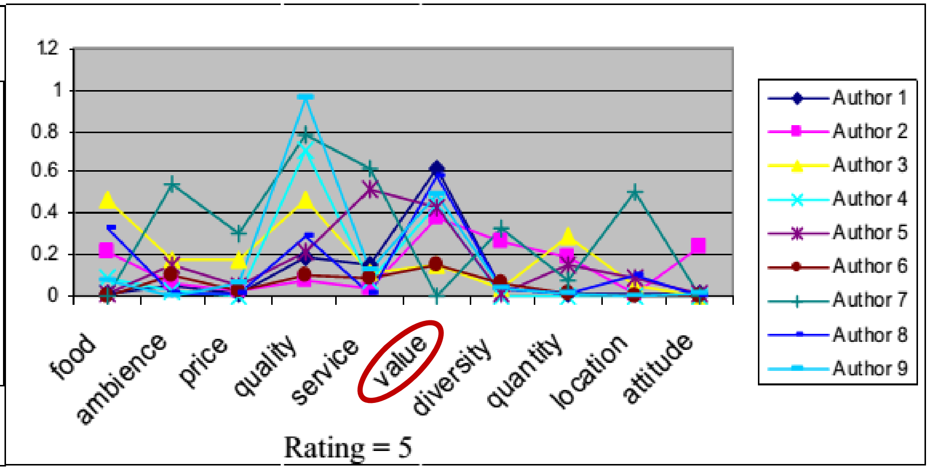


max planck institut informatik

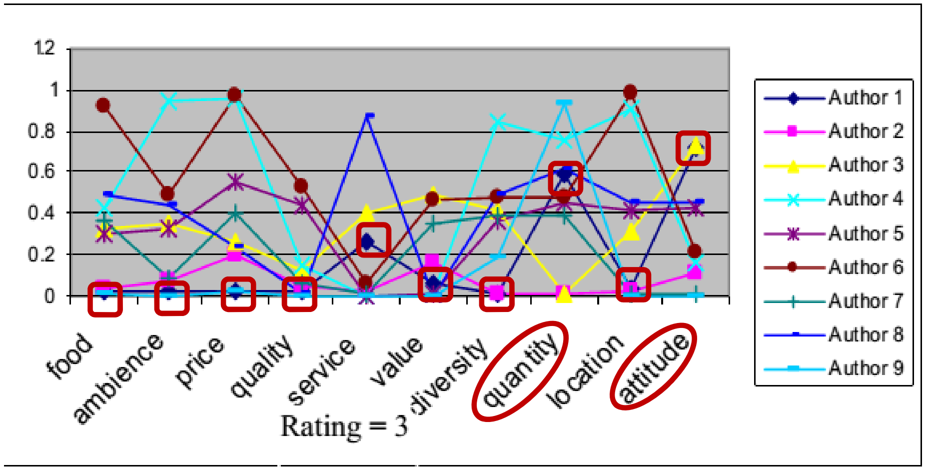
Snapshot of Author-Rating-Topic-Label Distribution Extracted by JAST - TripAdvisor

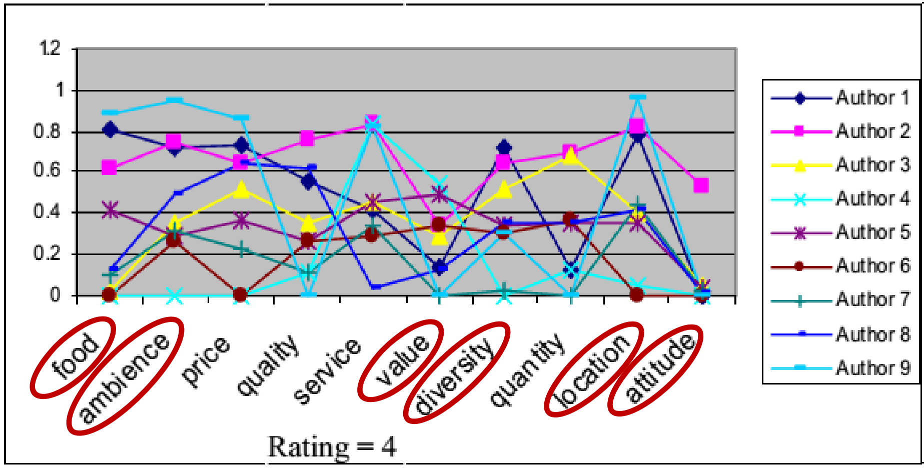


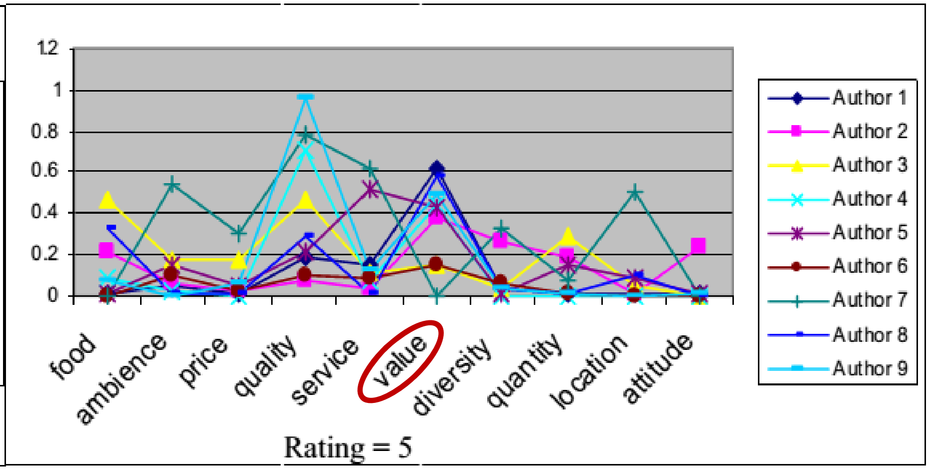




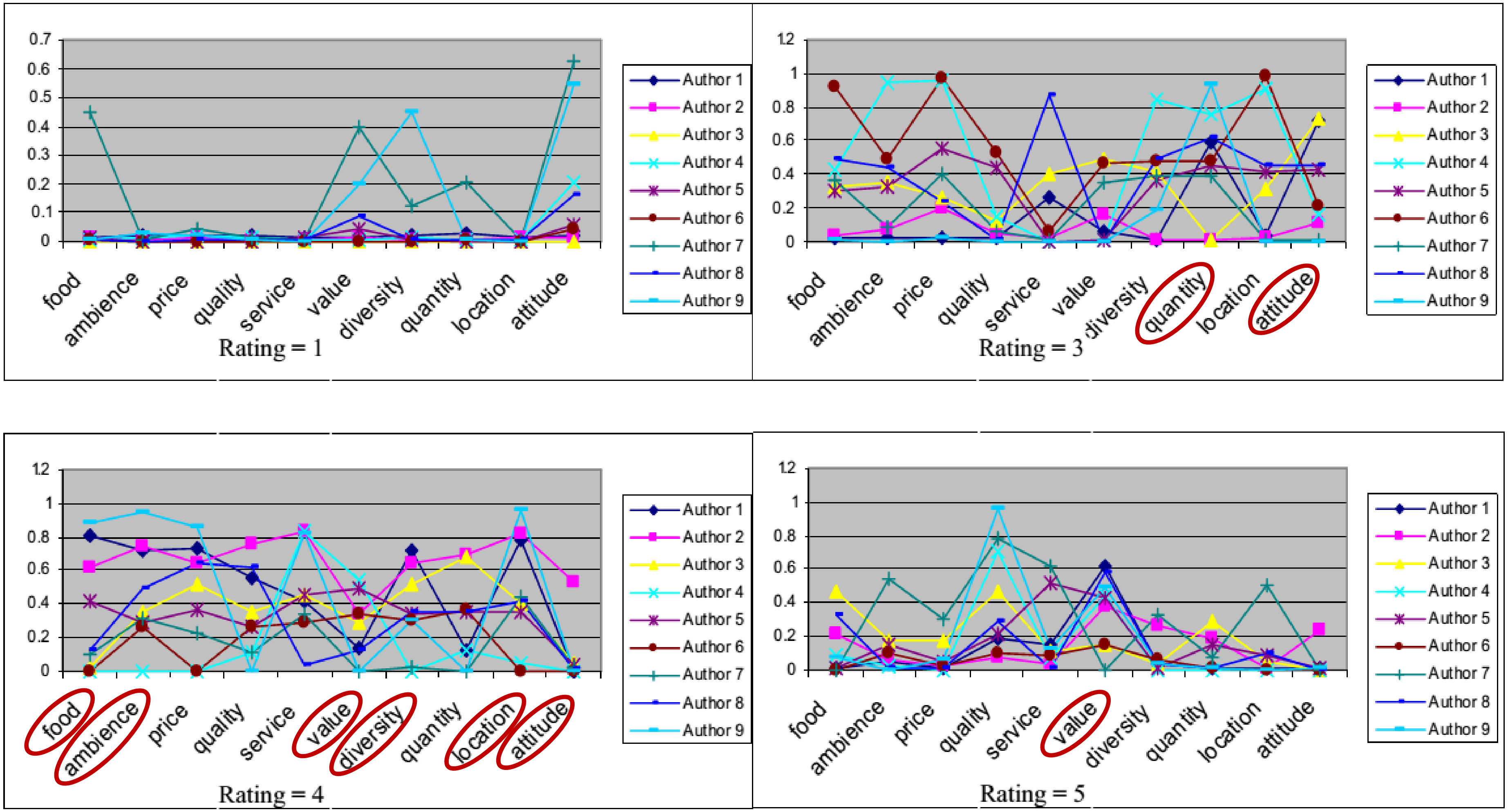
Snapshot of Author-Rating-Topic-Label Distribution Extracted by JAST - TripAdvisor



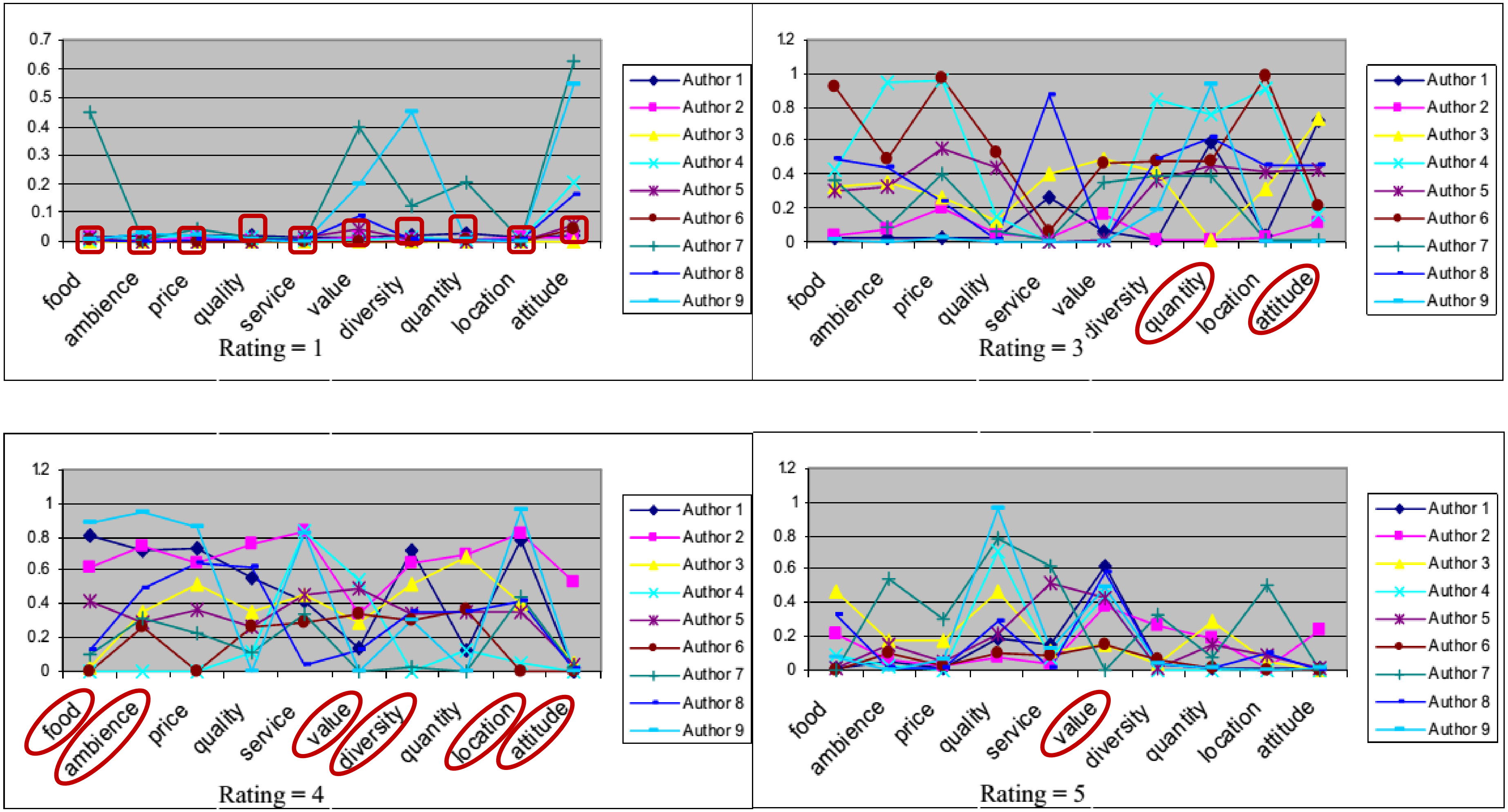




Snapshot of Author-Rating-Topic-Label Distribution Extracted by JAST - TripAdvisor

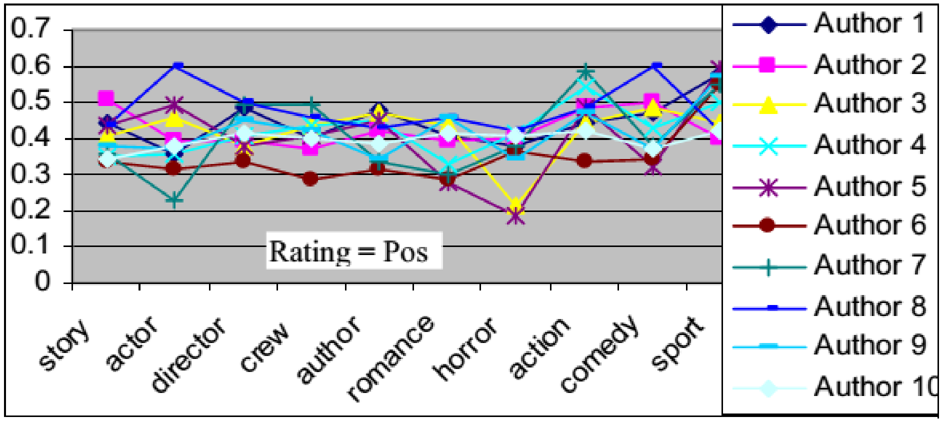


Snapshot of Author-Rating-Topic-Label Distribution Extracted by JAST - TripAdvisor

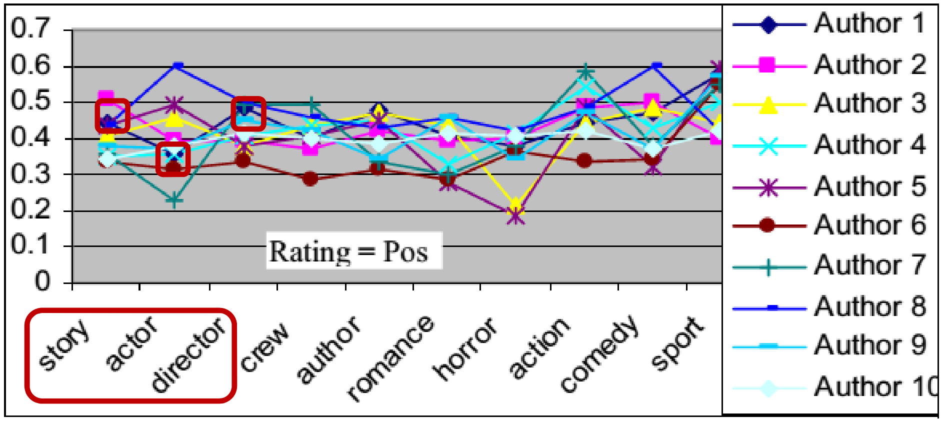


max planck institut informatik

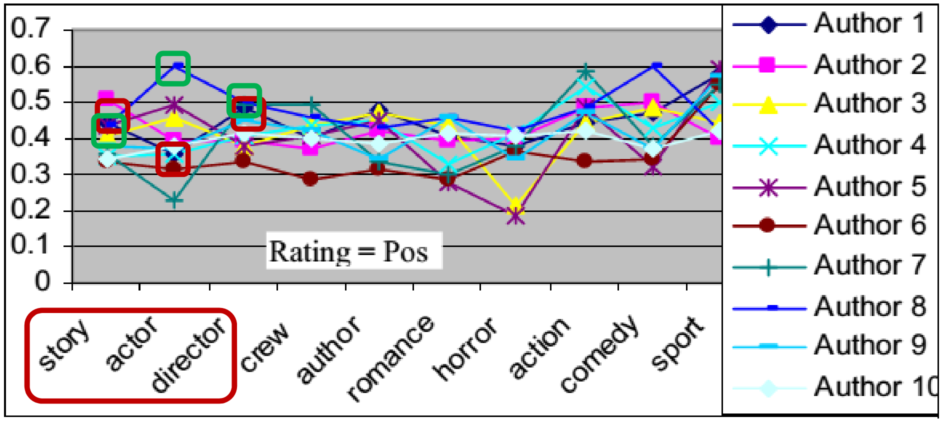
Snapshot of Author-Rating-Topic-Label Distribution Extracted by JAST - IMDB



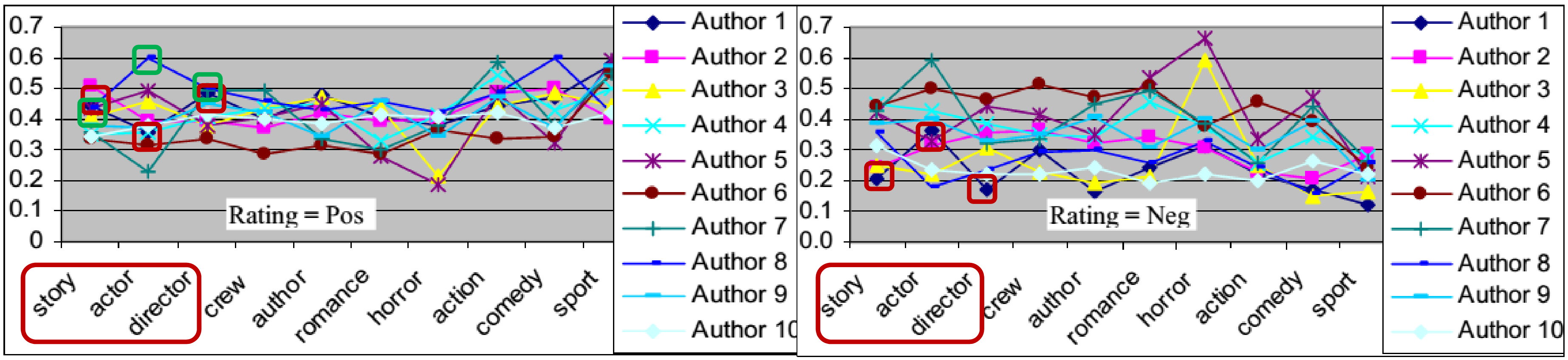
Snapshot of Author-Rating-Topic-Label Distribution Extracted by JAST - IMDB



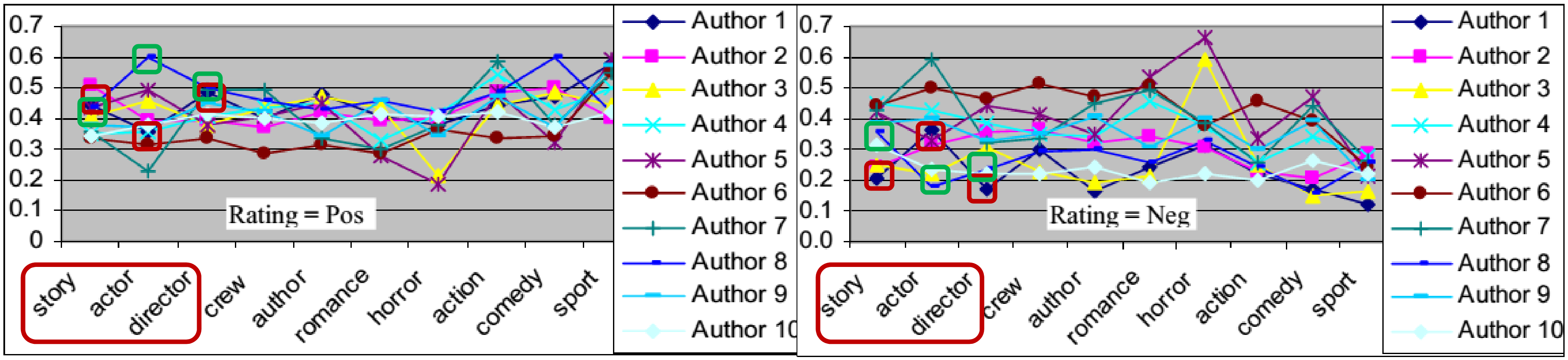
Snapshot of Author-Rating-Topic-Label Distribution Extracted by JAST - IMDB



Snapshot of Author-Rating-Topic-Label Distribution Extracted by JAST - IMDB



Snapshot of Author-Rating-Topic-Label Distribution Extracted by JAST - IMDB



Conclusions

- Sentiment classification and aspect rating prediction models can be improved if author is *known*

Conclusions

- Sentiment classification and aspect rating prediction models can be improved if author is *known*
- Authorship information helps in identifying author topic preferences, and author writing style to maintain review coherence
  - Semantic-syntactic class transition and topic switch

Conclusions

- Sentiment classification and aspect rating prediction models can be improved if author is *known*
- Authorship information helps in identifying author topic preferences, and author writing style to maintain review coherence
  - Semantic-syntactic class transition and topic switch
- JAST is unsupervised, with overhead of knowing author identity
- Performs better than all unsupervised/semi-supervised models and some supervised models

Conclusions

- Sentiment classification and aspect rating prediction models can be improved if author is *known*
- Authorship information helps in identifying author topic preferences, and author writing style to maintain review coherence
  - Semantic-syntactic class transition and topic switch
- JAST is unsupervised, with overhead of knowing author identity
- Performs better than all unsupervised/semi-supervised models and some supervised models
- It will be interesting to use JAST for authorship attribution task

QUESTIONS ???