School of
Information Systems

SMU
SINGAPORE MANAGEMENT UNIVERSITY

Predicting Outcome of Collaborative Featured Article Nomination in Wikipedia

Meiqun Hu, Ee-Peng Lim and Ramayya Krishnan

Outline

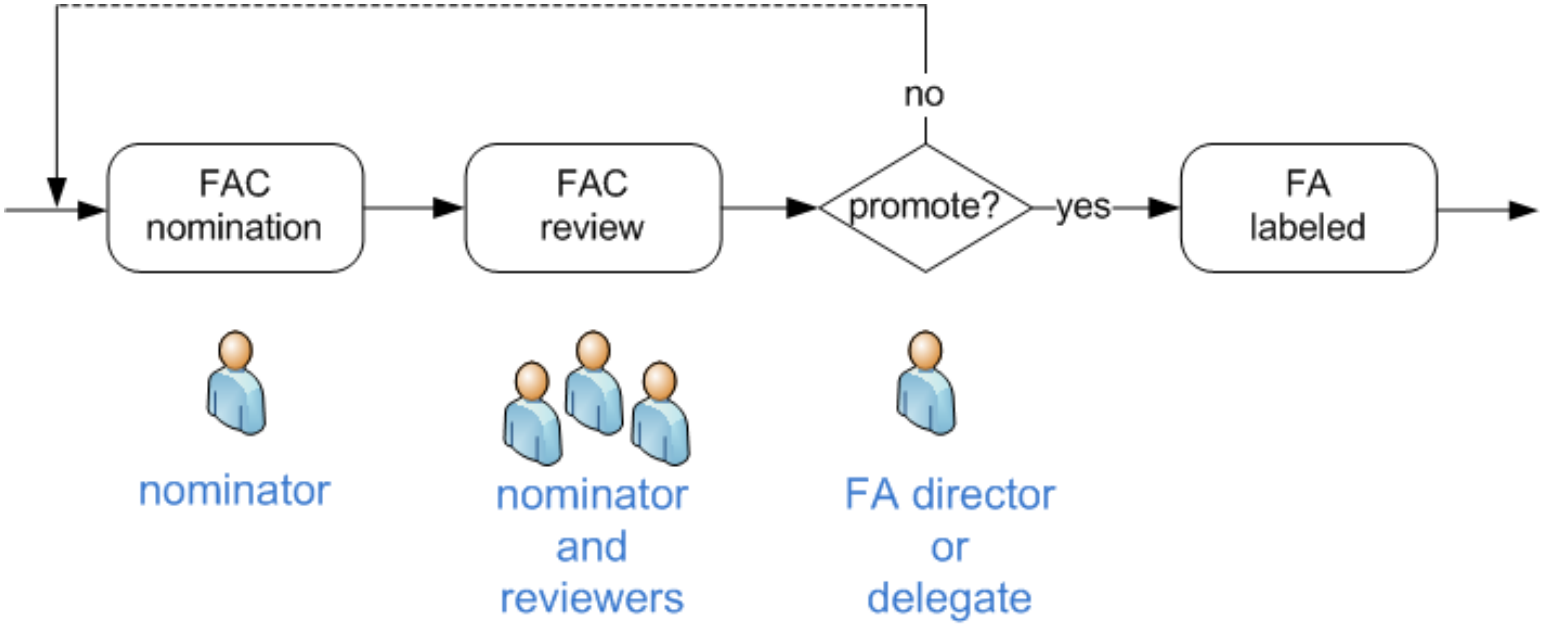
- Featured articles in Wikipedia
  - Nomination of featured articles
  - Featured article candidates dataset
- Prediction on FAC outcome
  - Discussion features
  - User features
  - Collaborator features
- Conclusion

Featured article in Wikipedia

- Wikipedia
  - the largest online collaborative authoring site
  - anyone can edit
  - uneven quality in articles

- Featured articles (FA)
  - represent the best articles in Wikipedia
  - Featured article criteria
  - Featured article nomination/review

Key steps in acquiring FA label



FAC : featured article candidate
FAC session : [time of nomination, time of decision]

nomination justification

Mass Rapid Transit (Singapore)

[edit]

Semi-**self nomination**. Took Singapore Wikipedians a month to summarise and cleanup the article to its current form. Compare with before version Peer review suggests no significiant ideas/changes, so I think it should be ready by now. This is the first Singapore-related article going up for Featured Article Candidate. - Mailer Diablo 04:03, 18 December 2005 (UTC)

- **Support**, must admit it's a semi-self nom for me to vote too, but it's been a long way and I think it is up to standard. -- Natalinasmpf 04:22, 18 December 2005 (UTC)
- **Support**:Good article. I haven't read the whole of it in detail, but overall, through the titles, pictures, and some portions I read, it looks comprehensive. Great visual impact, and I noticed that everything is properly referenced. Can't see any reasons why it shouldn't be featured. deeptrivia (talk) 04:26, 18 December 2005 (UTC)
- **Weak support** haven't delved into it yet but looks good. NSLE (T+C+CVU) 04:29, 18 December 2005 (UTC)
- ~~**Object**. Needs a good copy-edit. Overlinked (see WP's policy on trivial chronological links and common noun links, and the following pages Wikipedia:Make only links relevant to the context, Wikipedia:Manual of Style (links)#Internal links, Wikipedia:Manual of Style (dates and numbers)#Date formatting and Wikipedia_talk:Manual_of_Style_(dates_and_numbers)#Dates_linking_convention_currently_ludicrous. Please use lower case for headings consistently. Tony 06:49, 18 December 2005 (UTC)~~

  depth of 1

  - De-linked sections, date now consistent. ~~Are you sure there is Wikipedia_talk:Manual_of_Style_(dates_and_numbers)#Dates_linking_convention_currently_ludicrous? I cannot find it.~~ For headings, only names that are not in lower case are official names given by the authorities, including the Standard Ticket. - Mailer Diablo 08:18, 18 December 2005 (UTC)

    depth of 2

    - AFAIK Wikipedia_talk:Manual_of_Style_(dates_and_numbers)/archive28#Dates_linking_convention_ is not official policy. - Mailer Diablo 09:04, 18 December 2005 (UTC)

    depth of 3

    That's right—I wrote "see WP's policy on trivial chronological links and common noun links, *and* the following pages". The linking problem has been fixed: well done! I'll have a look at the prose later—it needs work. Tony 09:20, 18 December 2005 (UTC)

    depth of 3

- **Support** of course! Great job done. Article deserves what it really deserves. --Terence Ong | talk 17:00, 18 December 2005 (UTC)
- **Object**. [Now strong object, see below for revised discussion]. I agree with the need for a thorough copyedit, with particular attention to lengthy sentences which really have next to no content, like this one: "Numerous measures have been taken by operators and authorities to ensure the safety of passengers travelling on the system." ("passengers travelling on the system" should just be "passengers"; and the sentence would be better in the form

Motivation and research objectives

- Motivation
  - In Wikipedia, good articles are wanted.
  - Wikipedia has been growing exponentially, however,
    - number of featured articles is growing linearly
    - FA selection process is laborious
    - decision making is only shouldered by the FA director and his delegate

- We aim to aid in decision making
  - to collect FAC review data, and analyze user interaction during review process
  - to predict nomination outcomes using feature derived from interaction analysis

Featured article candidates dataset

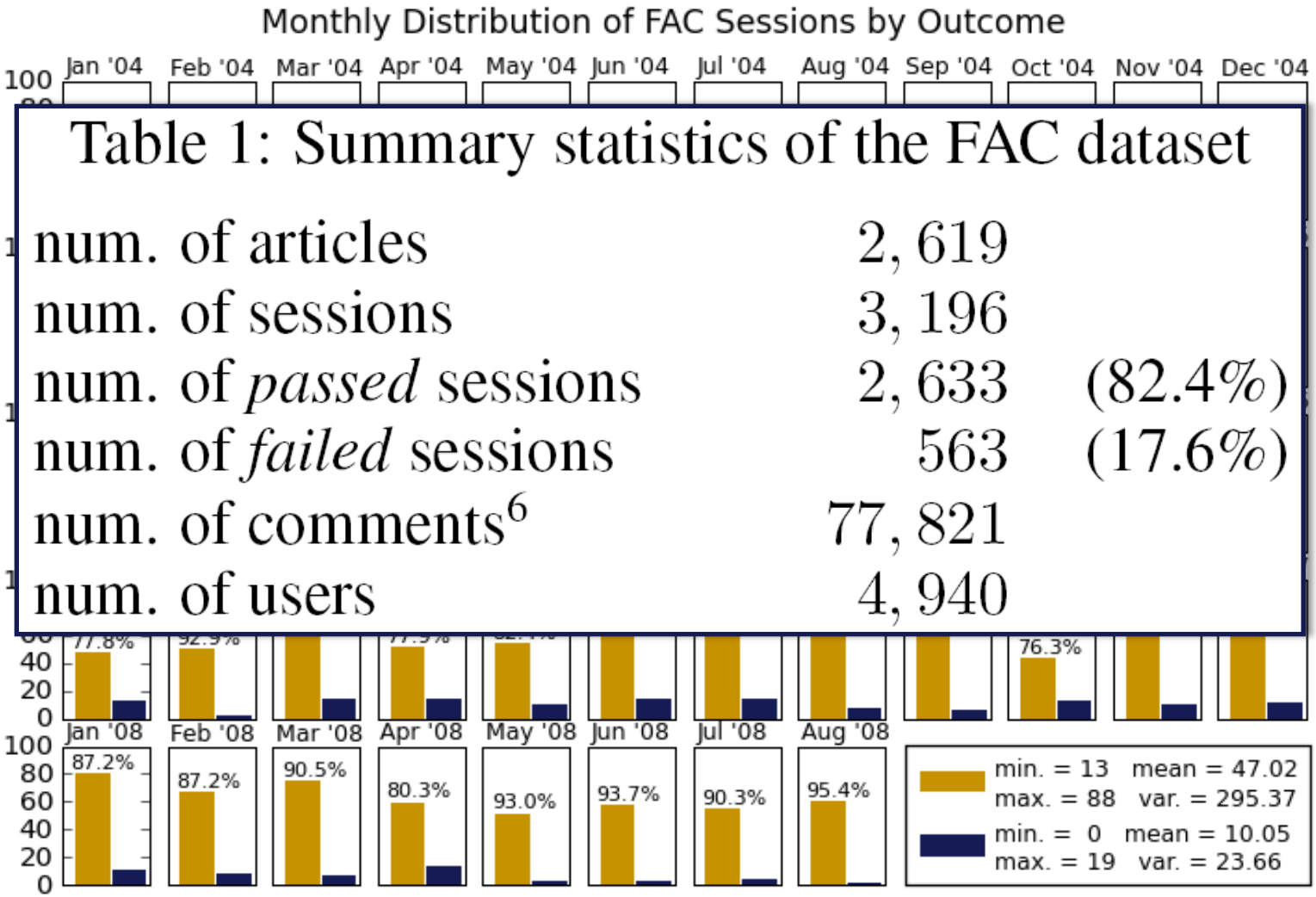
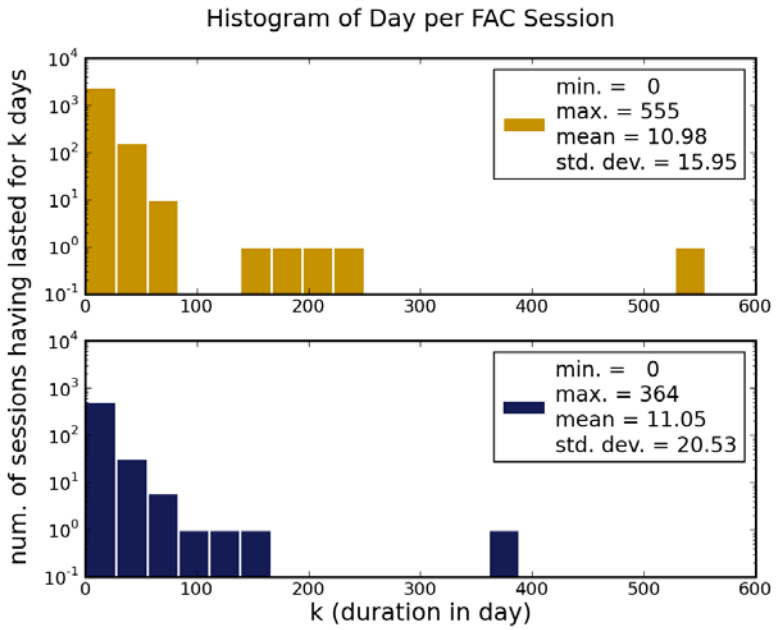
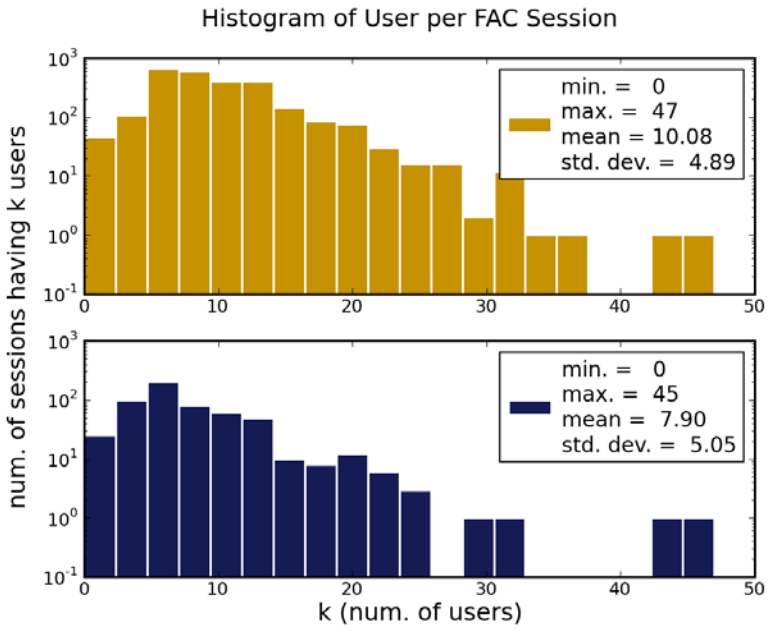


Table 1: Summary statistics of the FAC dataset

num. of articles	2,619	
num. of sessions	3,196	
num. of *passed* sessions	2,633	(82.4%)
num. of *failed* sessions	563	(17.6%)
num. of comments[6]	77,821	
num. of users	4,940	

Statistics per FAC session (1)

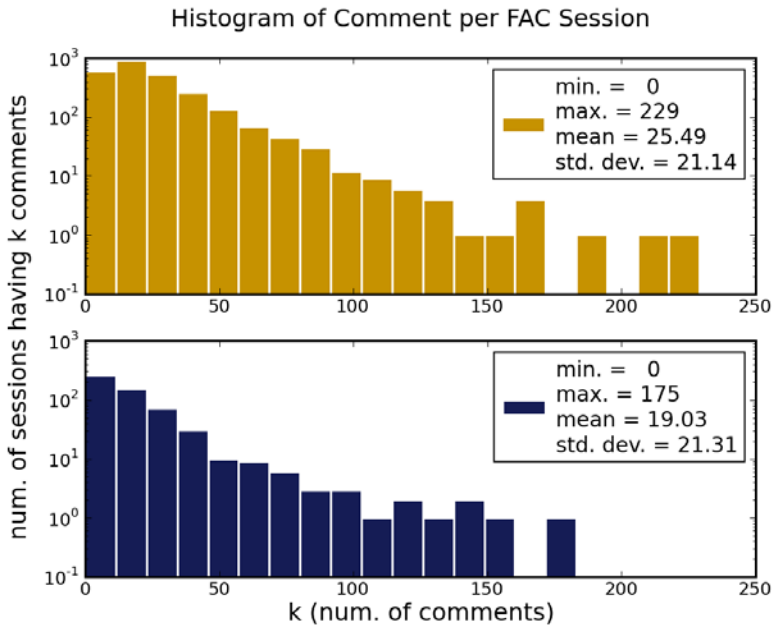


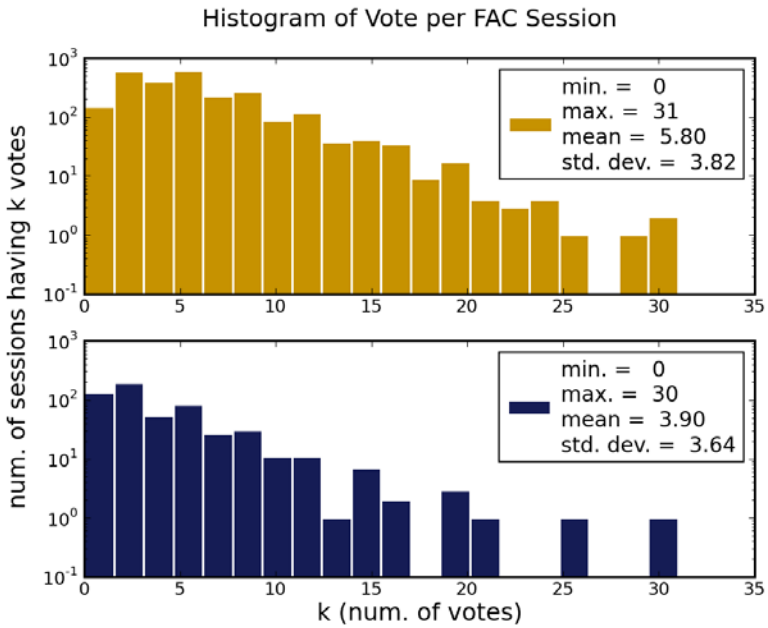


duration (in days)

distinct users

Statistics per FAC session (2)

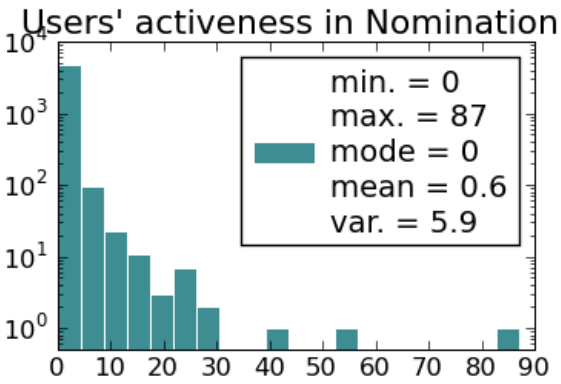


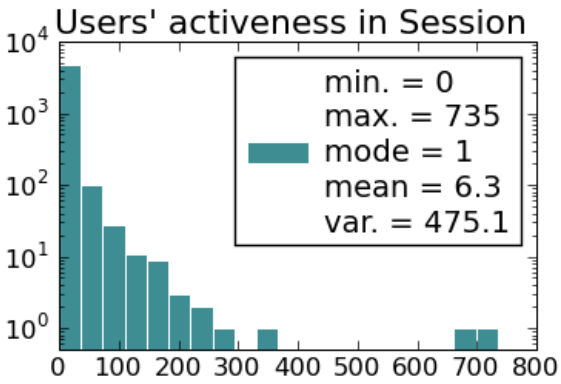


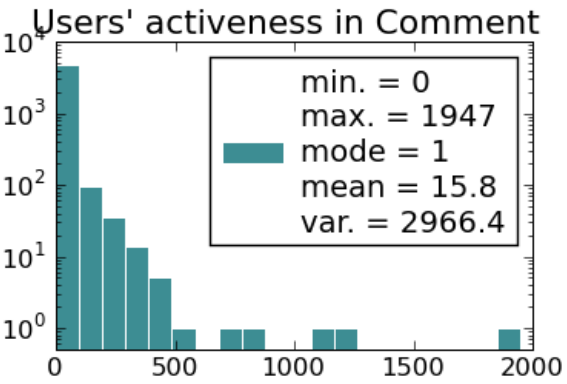
number of comments

number of votes

Users' activeness in FAC sessions



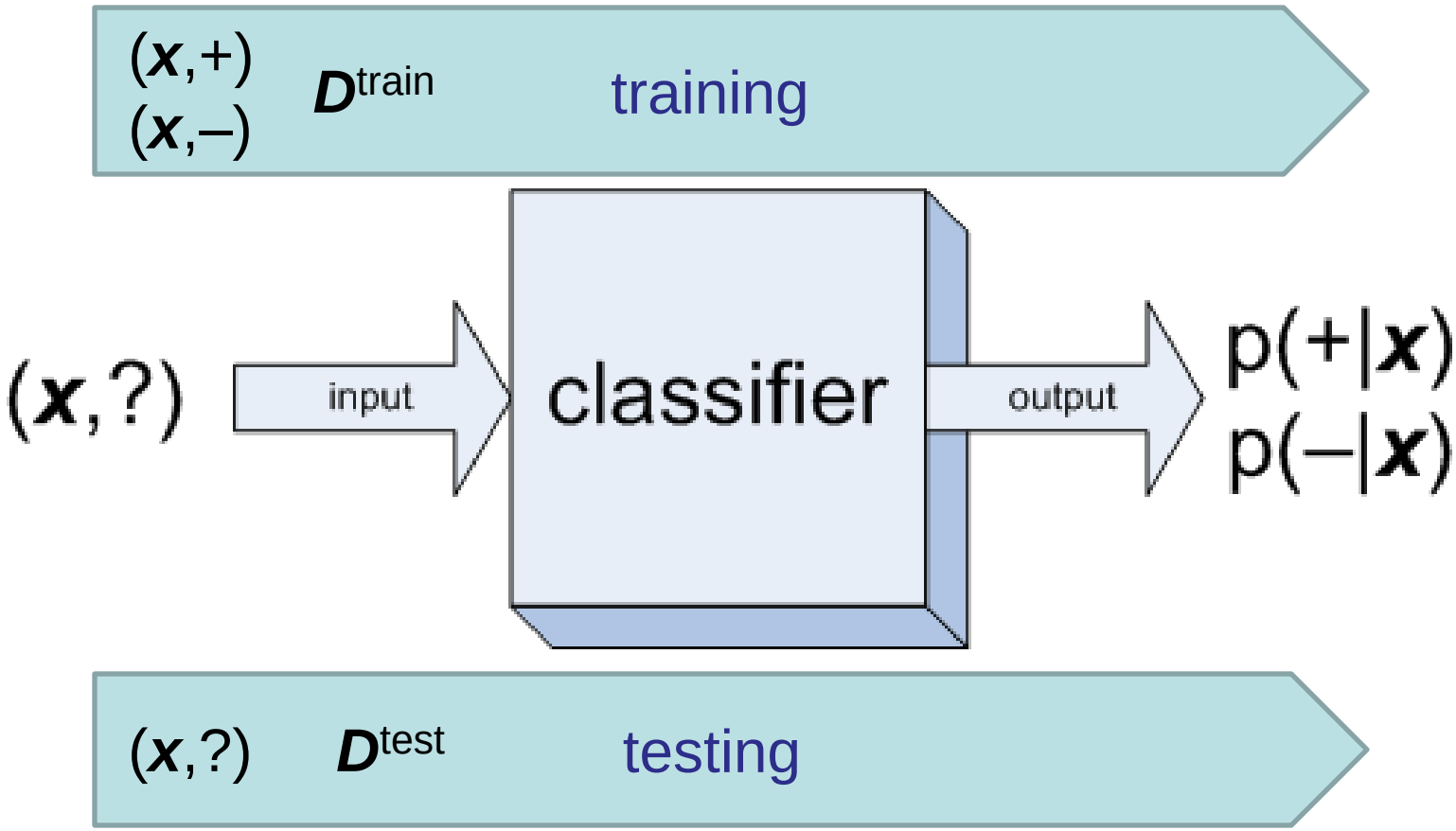




Outline

- Featured articles in Wikipedia
  - Nomination of featured articles
  - Featured article candidates dataset
- Prediction on FAC outcome
  - Discussion features
  - User features
  - Collaborator features
- Conclusion

Classification in a nutshell



x : features of the data instance, often multiple dimensions

Feature engineering for prediction

$$\langle D_x, U_y(F_u), P_z(F_p) \rangle$$

- D_x
  - discussion features, $x \subseteq \{g + c, v\}$
- $U_y(F_u)$
  - user features, $y \in \{N, S, C, L\}$
  - user weighting options, $F_u \in \{e_u, c_u, p_u, s_u\}$
- $P_z(F_p)$
  - collaborator (pair) features, $z \in \{Co, Ag, Dg\}$
  - collaborator weighting options, $F_p \in \{e_p, c_p, p1_p, p2_p\}$

Discussion features, $D_{\{g,c,v\}}$

- **General discussion features**
  1. duration (in days)
  2. total number of comments
  3. total number of distinct users
  4. average number of comments per user
- **Comment-specific discussion features**
  5-6. max. and avg. length of comments
  7-8. max. and avg. depth of comments
  9. self nomination (b)
  10. FA director participation (b)
  11. FA director's delegate participation (b)
- **Voting-specific discussion features**
  12. number of comments at depth 1
  13. number of comments at depth 1 that also votes
  14. fraction of comments that vote for support
  15. fraction of comments that vote for objection

User features, $U_{\{N,S,C,L\}}(F_u)$

- Features defined on the dimension of individual users
- Selecting top 50 active users
  - N, number of FAC nomination
  - S, number of FAC participation
  - C, number of comments given in FAC sessions
  - L, number of distinct FAC co-reviewers
- Assigning feature values
  - e_u, existence, {1,0}
  - p_u, polarity, {+1,-1,0}
  - c_u, comment, {0,1,…}
  - s_u, signed comment, {…,-1,0,+1,…}

Collaborator features, $U_{\{Co,Ag,Dg\}}(F_p)$

- Features defined on the dimension of pairs of users
- Selecting top 100 collaborative user pairs
  - Co, number of FAC sessions co-reviewed
  - Ag, degree of agreement

$$\frac{\text{number of sessions the pair agree in their votes}}{\text{number of sessions the pair both voted}}$$

  - Dg, degree of disagreement
- Assigning feature values
  - e_p, pair existence, {1,0}
  - c_p, sum of comments, {0,1,…}
  - $p1_p$, paired polarity, option 1, {-2,-1,0,1,2}
  - $p2_p$, paired polarity, option 2, {-2,-1,-0.5,0,1,2}

Experiment setup

- Training vs. test dataset
  - 10 folds cross-validation
  - stratified sampling based on the outcome
- Classifier
  - Linear SVM, with cost factor 0.2
  - Platt's calibration, SVM decision values to class posterior probabilities
- Evaluation
  - area under the curve (AUC) on precision-recall (PR) curve
  - precision and recall for the '–' class

AUC using discussion features

$\langle \mathsf{D}_{\{\mathsf{g+c}\}}, \emptyset, \emptyset \rangle$	$0.402\ (\pm 0.063)$
$\langle \mathsf{D}_{\{\mathsf{v}\}}, \emptyset, \emptyset \rangle$	$0.816\ (\pm 0.057)$
$\langle \mathsf{D}_{\{\mathsf{g+c,v}\}}, \emptyset, \emptyset \rangle$	$0.822\ (\pm 0.052)$
baseline	0.176

baseline : the maximum prior classifier

1. using voting specific discussion features performs better than non-voting discussion features;
2. using both voting and non-voting features outperforms the latter;
3. all proposed feature settings perform better than the baseline.

AUC using user features

$\langle D_{\{g+c\}}, U_N(e_u), \emptyset \rangle$	0.438^* (± 0.060)
$\langle D_{\{g+c\}}, U_N(c_u), \emptyset \rangle$	0.432^* (± 0.071)
$\langle D_{\{g+c\}}, U_N(p_u), \emptyset \rangle$	0.511^* (± 0.068)
$\langle D_{\{g+c\}}, U_N(s_u), \emptyset \rangle$	0.468^* (± 0.067)
$\langle D_{\{g+c\}}, U_S(e_u), \emptyset \rangle$	0.439^* (± 0.064)
$\langle D_{\{g+c\}}, U_S(c_u), \emptyset \rangle$	0.413 (± 0.057)
$\langle D_{\{g+c\}}, U_S(p_u), \emptyset \rangle$	0.590^* (± 0.052)
$\langle D_{\{g+c\}}, U_S(s_u), \emptyset \rangle$	0.470^* (± 0.062)
$\langle D_{\{g+c\}}, U_C(e_u), \emptyset \rangle$	0.446^* (± 0.051)
$\langle D_{\{g+c\}}, U_C(c_u), \emptyset \rangle$	0.429^* (± 0.055)
$\langle D_{\{g+c\}}, U_C(p_u), \emptyset \rangle$	0.558^* (± 0.050)
$\langle D_{\{g+c\}}, U_C(s_u), \emptyset \rangle$	0.460^* (± 0.070)
$\langle D_{\{g+c\}}, U_L(e_u), \emptyset \rangle$	0.440^* (± 0.063)
$\langle D_{\{g+c\}}, U_L(c_u), \emptyset \rangle$	0.406 (± 0.056)
$\langle D_{\{g+c\}}, U_L(p_u), \emptyset \rangle$	0.586^* (± 0.055)
$\langle D_{\{g+c\}}, U_L(s_u), \emptyset \rangle$	0.469^* (± 0.062)

AUC using collaborator features

$\langle \mathsf{D}_{\{g+c\}}, \emptyset, \mathsf{P}_{\mathsf{Co}}(\mathsf{e_p}) \rangle$	$0.383 \quad (\pm 0.058)$
$\langle \mathsf{D}_{\{g+c\}}, \emptyset, \mathsf{P}_{\mathsf{Co}}(\mathsf{c_p}) \rangle$	$0.369 \quad (\pm 0.054)$
$\langle \mathsf{D}_{\{g+c\}}, \emptyset, \mathsf{P}_{\mathsf{Co}}(\mathsf{p1_p}) \rangle$	$0.556^{*} \quad (\pm 0.037)$
$\langle \mathsf{D}_{\{g+c\}}, \emptyset, \mathsf{P}_{\mathsf{Co}}(\mathsf{p2_p}) \rangle$	$0.552^{*} \quad (\pm 0.032)$
$\langle \mathsf{D}_{\{g+c\}}, \emptyset, \mathsf{P}_{\mathsf{Ag}}(\mathsf{e_p}) \rangle$	$0.397 \quad (\pm 0.043)$
$\langle \mathsf{D}_{\{g+c\}}, \emptyset, \mathsf{P}_{\mathsf{Ag}}(\mathsf{c_p}) \rangle$	$0.388 \quad (\pm 0.061)$
$\langle \mathsf{D}_{\{g+c\}}, \emptyset, \mathsf{P}_{\mathsf{Ag}}(\mathsf{p1_p}) \rangle$	$0.571^{*} \quad (\pm 0.067)$
$\langle \mathsf{D}_{\{g+c\}}, \emptyset, \mathsf{P}_{\mathsf{Ag}}(\mathsf{p2_p}) \rangle$	$0.572^{*} \quad (\pm 0.067)$
$\langle \mathsf{D}_{\{g+c\}}, \emptyset, \mathsf{P}_{\mathsf{Dg}}(\mathsf{e_p}) \rangle$	$0.375 \quad (\pm 0.053)$
$\langle \mathsf{D}_{\{g+c\}}, \emptyset, \mathsf{P}_{\mathsf{Dg}}(\mathsf{c_p}) \rangle$	$0.377 \quad (\pm 0.062)$
$\langle \mathsf{D}_{\{g+c\}}, \emptyset, \mathsf{P}_{\mathsf{Dg}}(\mathsf{p1_p}) \rangle$	$0.568^{*} \quad (\pm 0.075)$
$\langle \mathsf{D}_{\{g+c\}}, \emptyset, \mathsf{P}_{\mathsf{Dg}}(\mathsf{p2_p}) \rangle$	$0.560^{*} \quad (\pm 0.067)$

AUC using the ‘best of best’ features

$\langle \mathsf{D}_{\{g+c\}}, \mathsf{U}_{\mathsf{S}}(\mathsf{p}_u), \emptyset \rangle$	$0.590\ (\pm 0.052)$
$\langle \mathsf{D}_{\{g+c\}}, \mathsf{U}_{\mathsf{L}}(\mathsf{p}_u), \emptyset \rangle$	$0.586\ (\pm 0.055)$
$\langle \mathsf{D}_{\{g+c\}}, \emptyset, \mathsf{P}_{\mathsf{Ag}}(\mathsf{p1}_{\mathsf{p}}) \rangle$	$0.571\ (\pm 0.067)$
$\langle \mathsf{D}_{\{g+c\}}, \emptyset, \mathsf{P}_{\mathsf{Ag}}(\mathsf{p2}_{\mathsf{p}}) \rangle$	$0.572\ (\pm 0.067)$

1. using user features improves AUC performance significantly;
2. using collaborator pair features improves AUC, but not statistically significant.

Outline

- Featured articles in Wikipedia
  - Nomination of featured articles
  - Featured article candidates dataset
- Prediction on FAC outcome
  - Discussion features
  - User features
  - Collaborator features
- Conclusion

Conclusion

- We analyze user collaboration in the process of FAC nomination and review
  - users' participation, commenting, voting statistics
  - consensus is largely followed in the review process

- We address the task of predicting FAC outcome as binary classification using features derived from review data and user collaboration
  - using vote consensus gives strong performance
  - using user features improved prediction significantly

Future work

- To compare classifier performance when varying the number of active users selection, and compare with random selection.
- To look at the classifier performance for cases where consensus does not exist.
- To associate with article's editing history during the review period.
- To examine performance for controversial articles.