TO EACH HIS OWN: PERSONALIZED CONTENT SELECTION BASED ON TEXT COMPREHENSIBILITY

CHENHAO TAN, EVGENIY GABRILOVICH, BO PANG
CORNELL UNIVERSITY, YAHOO! RESEARCH
MOTIVATION

Various factors explain users’ choices in content consumption

- Topic (personalized search, user modeling etc)
- Beyond Topical Relevance
MOTIVATION

Various factors explain users’ choices in content consumption

• Topic (personalized search, user modeling etc)
• Beyond Topical Relevance

Text comprehensibility

• The degree of difficulty of text, e.g. as judged by average sentence length and vocabulary size
• Motivating example
  A search on antibiotic resistance:
  a physician vs. a patient
An antibacterial is a compound or substance that kills or slows down the growth of bacteria. The term is often used synonymously with the term antibiotic(s); today, however, with increased knowledge of the causative agents of various infectious diseases, antibiotic(s) has come to denote a broader range of antimicrobial compounds, including antifungal and other compounds.

Antibiotics are medicine that kills bacteria or slows the growth of bacteria. They are used to cure diseases. Antibiotics do not harm people. Penicillin is a popular antibiotic. Antibiotics started to be produced in 1939. Antibiotics can not stop a virus. Antibiotics are not the same thing as antibodies.

Intuitively, we see these texts differ in:
- Complexity of syntax
- Technical terms
- Topic independent vocabulary
- …
CHALLENGES

Estimate the comprehensibility of text

Model and predict users’ comprehensibility preferences without explicit preference information
  • Topic independent
  • Topic dependent

Improve the ranking in more than one setting
  • Web search
  • Community question answering
CHALLENGES

Estimate the comprehensibility of text

Model and predict users’ comprehensibility preferences without explicit preference information

- Topic independent
- Topic dependent

Improve the ranking in more than one setting

- Web search
- Community question answering

Related work

  Characterizing Web Content, User Interests, and Search Behavior by Reading Level and Topic [Kim et al. 2012]
  Personalizing web search results by reading level [Collins-Thompson et al. 2011]
ESTIMATE TEXT COMPREHENSIBILITY


(40,032 aligned article pairs with the same title)

Features

- 6 linguistic readability indexes based on the length of sentences, the syllables of words, etc [CL, G, KFRC, M, M]
- A basic English word list: just 850 unigrams

Hard vs. Easy classification with logistic regression

Global threshold: 88.3%

Per-title comparison: 97.4%
An antibacterial is a compound or substance that kills or slows down the growth of bacteria. The term is often used synonymously with the term antibiotic(s); today, however, with increased knowledge of the causative agents of various infectious diseases, antibiotic(s) has come to denote a broader range of antimicrobial compounds, including antifungal and other compounds.

Antibiotics are medicine that kills bacteria or slows the growth of bacteria. They are used to cure diseases. Antibiotics do not harm people. Penicillin is a popular antibiotic. Antibiotics started to be produced in 1939. Antibiotics can not stop a virus. Antibiotics are not the same thing as antibodies.

Regular English Wikipedia

Simple English Wikipedia
ESTIMATE TEXT COMPREHENSIBILITY

*English Wikipedia VS. Simple English Wikipedia*

(40,032 aligned article pairs with the same title)

Features

- 6 linguistic readability indexes based on the length of sentences, the syllables of words, etc [CL, G, KFRC, M, M]
- A basic English word list: just 850 unigrams

*Hard vs. Easy classification with logistic regression*

  Global threshold: 88.3%

  **Per-title comparison: 97.4%**
MODEL USER PREFERENCES: TOPIC INDEPENDENT (BASIC)

We get preference pairs for each user

- From click log in web search
- From choosing the best answer in community question answering
GENERATE PREFERENCE PAIRS

Click log

- Three different ways, e.g. click > skip above
- Weight
  the closer two search results are, the larger the weight is

Best answer

- Best > Any other
- Weight
  \[ \frac{1}{\text{#answers}} \]
MODEL USER PREFERENCES: TOPIC INDEPENDENT (BASIC)

We get preference pairs for each user

\[ \Omega_u^{pref} = \{ (< a_i, b_i >, w_i) \mid a_i >_u b_i, \text{ with weight } w_i \} \]

Treat each tuple as a sample

\( P_u: \) the probability that user \( u \) prefers harder text

MLE estimation with smoothing

\[ P_u = \frac{\# \text{Samples where } u \text{ prefers harder text} + 1}{\# \text{Samples} + 2} \]

Weighted version

\[ P_u = \frac{\# \text{Weighted samples where } u \text{ prefers harder text} + 1}{\# \text{Weighted Samples} + 2} \]
MODEL USER PREFERENCES:
TOPIC DEPENDENT (TOPICAL)

Topic dependent

• Topic hierarchy (e.g. Yahoo!’s classifier for queries, or categories in Yahoo! answers)

\[ t_2 <_h t_1 \iff t_2 \text{ is a descendant of } t_1 \]

• Pairwise preferences for a topic \( t \) and a user \( u \)
  All the preference pairs in the descendants of \( t \) and \( t \)

\[ \Omega_{u,t}^{pref} = \{ pp_i \in \Omega_u^{pref} \mid t_i \leq_h t \} \]
MODEL USER PREFERENCES: TOPIC DEPENDENT (COLLABORATIVE)

Data sparseness

- Predict comprehensibility preferences for unseen topics
- Collaborative filtering
  Maximum margin matrix factorization [Weimer et al. 2007]

\[
\sum_{i,j,G_{ij} \neq 0} \|U^T V_{ij} - G_{ij}\|^2 + \|U\|_F + \|V\|_F
\]
COMBINE THE RANKINGS

*R(d)*: the original topic-relevance-based ranking

*R_u(d)*: the ranking in the descending order of the difficulty of the text

*P_u*: the probability that user *u* prefers harder text

*β*: a parameter tuned on a development set

Combined Score: \( V = R(d) + \beta \times (2 \times P_u - 1) \times R_u(d) \)
COMBINE THE RANKINGS

\( R(d) \): the original topic-relevance-based ranking

\( R_u(d) \): the ranking in the descending order of the difficulty of the text

\( P_u \): the probability that user \( u \) prefers harder text

\( \beta \): a parameter tuned on a development set

Combined Score: \( V = R(d) + \beta \times (2 \times P_u - 1) \times R_u(d) \)

\( P_u > 0.5 \iff (2 \times P_u - 1) > 0 \),

- text harder
- \( R_u(d) \) smaller
- \( V \) smaller
- final rank higher
COMBINE THE RANKINGS

$R(d)$: the original topic-relevance-based ranking

$R_u(d)$: the ranking in the descending order of the difficulty of the text

$P_u$: the probability that user $u$ prefers harder text

$\beta$: a parameter tuned on a development set

Combined Score: $V = R(d) + \beta \times (2 \times P_u - 1) \times R_u(d)$

$P_u > 0.5 \Rightarrow (2 \times P_u - 1) > 0$,
- text harder
- $R_u(d)$ smaller
- $V$ smaller
- final rank higher

$P_u < 0.5 \Rightarrow (2 \times P_u - 1) < 0$,
- text easier
- $R_u(d)$ larger
- $V$ smaller
- final rank higher
**EXPERIMENT ON SEARCH DATASET**

**Task:** Use our approach to improve the original web rank by personalization based on text comprehensibility

**Evaluation Measures [Dou et al. 2007]**

- Average Clicked Rank
- Rank Scoring

**Our Approach**

\[ 3 \times 2 \times 3 = 18 \]

- Click1
- Click2
- Click3
- Weighted
- Unweighted
- Basic
- Topical
- Collaborative

**Strength of preference**

\[ Q_u = |P_u - 0.5| \]

The larger \( Q_u \) is, the stronger preference towards harder or easier text \( u \) has
OVERALL PERFORMANCE: AVERAGE CLICKED RANK

The graph shows the improvement in web rank for different percentages of users, sorted by the strength of preference. The lines represent:
- web rank
- our: all queries
- our: non-repeated

The x-axis represents the percentage of users (log-scale), while the y-axis shows the improvement.

Sorted by the strength of preference:
- Strongest preference
- All users
OVERALL PERFORMANCE:
RANK SCORING

![Graph showing improvement in web rank and user queries over the percentage of users.](image)

- web rank
- our: all queries
- our: non-repeated

**OVERALL PERFORMANCE:**

**RANK SCORING**

- Sorted by the strength of preference

**Strongest preference**

**Percentage of Users (Log-scale)**

**Improvement**

All users
## Paired T-Test Against Web Rank on Average Clicked Rank

\[ *(p < 0.05), **(p < 0.01), ***(p < 0.001) \]

<table>
<thead>
<tr>
<th>Method</th>
<th>Weighted</th>
<th>Unweighted</th>
<th>strong 10%</th>
<th>50%</th>
<th>all 100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Click1</td>
<td><strong>Basic Topical Collaborative</strong></td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Click2</td>
<td><strong>Basic Topical Collaborative</strong></td>
<td>**</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Click3</td>
<td><strong>Basic Topical Collaborative</strong></td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Click1</td>
<td><strong>Basic Topical Collaborative</strong></td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Click2</td>
<td><strong>Basic Topical Collaborative</strong></td>
<td>**</td>
<td>**</td>
<td>***</td>
<td>**</td>
</tr>
<tr>
<td>Click3</td>
<td><strong>Basic Topical Collaborative</strong></td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>**</td>
</tr>
</tbody>
</table>
**PAIRED T-TEST AGAINST WEB RANK ON AVERAGE CLICKED RANK**

\[ *(p < 0.05), **(p < 0.01), *** (p < 0.001) \]

<table>
<thead>
<tr>
<th>Method</th>
<th>strong 10%</th>
<th>50%</th>
<th>all 100%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Weighted</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Click1</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Click2</td>
<td>**</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>***</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Click3</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td><strong>Unweighted</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Click1</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Click2</td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Click3</td>
<td>***</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
</tbody>
</table>
PAIRED T-TEST AGAINST WEB RANK ON AVERAGE CLICKED RANK

*(p < 0.05), **(p < 0.01), ***(p < 0.001)

<table>
<thead>
<tr>
<th>Method</th>
<th>strong 10%</th>
<th>50%</th>
<th>all 100%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WEIGHTED</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Click1</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Click2</td>
<td>**</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>***</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Click3</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td><strong>UNWEIGHTED</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Click1</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Click2</td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Click3</td>
<td>***</td>
<td>***</td>
<td>**</td>
</tr>
</tbody>
</table>
DIFFERENT USER PROFILE MODELS

Strongest preference

Improvement

Percentage of Users

web rank

collaborative

topical

basic
DIFFERENT USER PROFILE MODELS

See paper for more experiments and analysis!
EXPERIMENT ON YAHOO! ANSWERS

Task: Users can choose the best answer for questions they posted, we rank all the answers and try to make the rank of the best answer as small as possible

Method

- Random
- Majority (a preference for harder text)
- Our model

Performance

<table>
<thead>
<tr>
<th>Fraction of users</th>
<th>Random</th>
<th>Majority</th>
<th>Our model</th>
</tr>
</thead>
<tbody>
<tr>
<td>5% (strongest)</td>
<td>3.375</td>
<td>2.947</td>
<td>2.895 (***, ***)</td>
</tr>
<tr>
<td>10%</td>
<td>3.596</td>
<td>3.096</td>
<td>3.079 (***, ***)</td>
</tr>
<tr>
<td>100% (all)</td>
<td>4.525</td>
<td>4.093</td>
<td>4.149 (***, ***)</td>
</tr>
</tbody>
</table>

*(p < 0.05), **(p < 0.01), ***(p < 0.001)
## Differences from Collins-Thompson et al. 2011

<table>
<thead>
<tr>
<th></th>
<th>Collins-Thompson et al.</th>
<th>Our work</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Readability/comprehensibility classifier</strong></td>
<td>Explicitly models school reading levels</td>
<td>Trained on English Wikipedia vs. simple English Wikipedia. More general, e.g., improvement in topic health</td>
</tr>
<tr>
<td><strong>Approach</strong></td>
<td>A generative model</td>
<td>Extract preference pairs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Collaborative filtering</td>
</tr>
<tr>
<td><strong>Application</strong></td>
<td>Web search</td>
<td>Web search</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Community question answering</td>
</tr>
</tbody>
</table>
CONCLUSION

Develop a unified framework for personalized content selection using text comprehensibility

Model users’ comprehensibility preferences by extracting preference pairs and apply collaborative filtering to alleviate the problem of data sparseness

Modeling text comprehensibility can significantly improve content ranking in both web search and community question answering
Thank you!

Q & A