Modeling Ambiguity, Subjectivity, and Diverging Viewpoints in Opinion Question Answering Systems

Mengting Wan (m5wan@eng.ucsd.edu)    Julian McAuley (jmcauley@eng.ucsd.edu)
Opinion Question Answering Systems

- Many review websites (Amazon, Ebay, Yelp, Tripadvisor, etc.) do provide community question answering systems, where users can ask and answer item-related, complex and subjective opinion-oriented questions.
Q: “How will this camera lens work with my Nikon D3300 camera?”

Q: “Can I get to the San Diego safari zoo from my hotel by public transportation?”

- Ask the community? must wait for a response 😞
- Search for information among reviews? time-consuming to digest 😞
- Can we help users to automatically navigate reviews, in response to a particular question?
Opinion-oriented question is complicated

- **Opinion QA vs. traditional community QA**
  - Traditional cQA – **objective** information
    - Can be answered by constructing and exploring a factual knowledge-base
  - Opinion QA – **complicated 😊**
    - Objective, Subjective, Personal
    - Solution: retrieve relevant reviews rather than address the answer directly (WWW’2016, J. McAuley & A. Yang)

“Is this a good lens for my Nikon D3300 camera?”

http://www.amazon.com/Nikon-70-300mm-4-5-6D-Nikkor-Camera/dp/B00005LENR
Even more complicated!

Multiple Answers

Q: Will this work with the D3300

- **Probably not**, it did not work for AF on my D5000, I got this lens with my N70 years ago, still a good lens though. *(No)*
- **Yes it will** but the autofocus will not. There is no drive motor in the 3000 series cameras. Manual focus works well! *(Yes)*
- Hi, this lens **can not work** autofocus for D3300. Thanks in advance. *(No)*
- The lens **will work** but it will not have autofocus. You would have to focus manually. Rich *(Yes)*

Subjective Reviews

- *(1 of 1 people found the following review helpful)*
  - **★★★★☆** Great price on a 70-300mm lens… It **will not auto focus** with D3000 series and I knew that. I personally prefer manual focus in larger lenses … *(No)*
- *(0 of 0 people found the following review helpful)*
  - **★★★★☆** Nikon Nikkor 70-300mm f4-5.6 ED AF lens… It **works perfectly** on my Nikon D80 … *(Yes)*
- *(0 of 0 people found the following review helpful)*
  - **★★★★★** the best for my budget… Autofocus **works great with** my D70 camera … *(Yes)*
- *(0 of 0 people found the following review helpful)*
  - **★★★★★** Solid product… This lens **auto focus greatly with** the D7000 … *(Yes)*

http://www.amazon.com/Nikon-70-300mm-4-5-6D-Nikkor-Camera/dp/B00005LENR
Exploratory Analysis on Amazon QA Data

- QA data from Amazon
  - 808K questions, 3M answers in 8 large categories
- Questions are categorized into two types
  - Binary, Open-ended
- Build a conservative auto-labeler to label binary answers
  - logistic regression
  - keep the top 50% of the most confident predictions so that ambiguity arises due to the questions rather than any errors
- Answer distribution

<table>
<thead>
<tr>
<th>Binary Question</th>
<th>Open-ended Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>~14% binary questions are ambiguous (labels are inconsistent)</td>
<td>~4 answers per open-ended question in average</td>
</tr>
</tbody>
</table>
Goal:

- Given an item-related question, we’d like to determine
  - how relevant each review is to the question
  - measured by how helpful it is to identify proper response to the question
- with emphasis on modeling **ambiguity** and **subjectivity**.
A supervised approach for binary questions

Q: Will this work with the D3300

- Probably not, it did not work for AF on my D5000, I got this lens with my N70 years ago, still a good lens though. (No)
- Yes it will but the autofocus will not. There is no drive motor in the 3000 series cameras. Manual focus works well! (Yes)
- Hi, this lens can not work autofocus for D3300. Thanks in advance. (No)
- The lens would not have autofocus anyway, would have to focus manually. (Yes)

Nikon 70-300mm f/4-5.6D ED Auto Focus Nikkor SLR Camera Lens

A relevant review will help us to identify the proper answer(s) to the question.

Relevance finding → Binary prediction

- (1 of 1 people found the following review helpful)
  ★★★★★ Great price on a 70-300mm lens
  … It will not auto focus with D3000 series and I knew that. I personally prefer manual focus in larger lenses …
- (0 of 0 people found the following review helpful)
  ★★★★★ Nikon Nikkor 70-300mm f4-5.6 ED AF lens
  … It works perfectly on my Nikon D80 …
- (0 of 0 people found the following review helpful)
  ★★★★★ the best for my budget
  … Autofocus works great with my D70 camera
- (No)
- Hi, this lens can not work autofocus for D3300. Thanks in advance.
- The lens will work but it will not have autofocus. You would have to focus manually.

A pool of observed answers
Task 1: Resolve Labels

A set of votes from reviews
Task 2: Aggregate Predictions
Resolve Labels

Q: Will this work with the D3300?

... (No)
... (Yes)
... (No)
... (Yes)
...

- Use the fraction of positive answers \( r_q = \frac{n_q^+}{n_q^+ + n_q^-} \) as the label?

\[
\loglik = \sum_q \{ r_q \log p_q + (1 - r_q) \log (1 - p_q) \}
\]

If we assume the binary answer follows \( \text{Bernoulli}(r_q) \), this is the summation of the KL-divergences between answers and predictions.

- Limitation: real counts of positive and negative labels are discarded!
  - a question with 10+/10- labels (seems more controversial)
  - a question with 1+/1- labels

- Consider a different assumption for the label generation
Q: Will this work with the D3300?

... (No) – \( y_{q,1} \)  
... (Yes) – \( y_{q,2} \)  
... (No) – \( y_{q,3} \)  
... (Yes) – \( y_{q,4} \)  

\( y_q \)  

‘true’ label (hidden)

Ambiguity

‘sensitivity’ \( \alpha_q \): \( P(y_{q,j} = yes \mid y_q = yes) = \sigma(y_1, f_q) \)

‘specificity’ \( \beta_q \): \( P(y_{q,j} = no \mid y_q = no) = \sigma(y_2, f_q) \)

Joint distribution:

\( a_q = P(y_{q,1}, ..., y_{q,n_q} \mid y_q = yes, data) = \alpha_q^{n_q^*}(1 - \alpha_q)^{n_q^*} \)

\( b_q = P(y_{q,1}, ..., y_{q,n_q} \mid y_q = no, data) = (1 - \beta_q)^{n_q^*}\beta_q^{n_q^*} \)

Resolve Labels (EM)

Assume the ‘true’ label \( y_q \) of a question is hidden, but we have a set of observed labels \( \{y_{q,1}, y_{q,2}, ..., y_{q,n_q}\} \)

Training: Now embed them into the loglikelihood function (with the hidden ‘true’ label \( y_q \) )

\[
\loglik = \sum_q \{y_q \log a_q p_q + (1 - y_q) \log b_q (1 - p_q)\}
\]

Inference (EM-algorithm):

E-step: take the expectation of \( y_q \)

M-step: optimize \( \loglik \)
Aggregate Predictions (MoE)

- **Idea:**
  - each review (sentence) can be regarded as a weak predictor – give us a vote “yes/no”
  - a relevant review can help to predict a proper label

- **Mixture of Experts (MoE):**

  \[
  p_q = P(y_q | data) = \sum_r P(r | data) \times P(y_q | r, data)
  \]

  - How relevant is \( r \)
  - Prediction from \( r \)
  - \( p_{q,r} = \sigma(w_{q,r}) \)
  - pred. from text * \( (1 + \text{rating} + \text{bias}) \)
  - \( \alpha \exp(v_{q,r}) \)

  text relevance + **helpfulness** + expertise

  \[ \text{Subjectivity} \]

---

**Great price on a 70-300mm lens**
... It will not auto focus with D3000 series and I knew that. I personally prefer manual focus in larger lenses...
(from user 1)

**Nikon Nikkor 70-300mm f4-5.6 ED AF lens**
... It works perfectly on my Nikon D80...
(from user 2)

**the best for my budget**
... Autofocus works great with my D70 camera...
(from user 3)

**Solid product**
... This lens auto focus greatly with the D7000...
(from user 4)
Model Wrap-Up: EM-MoE

Ambiguity!

\[ y_{q,1} \rightarrow \quad y_q \rightarrow \quad p_q \rightarrow \quad p_{q,r_1}, p_{q,r_2}, p_{q,r_3}, p_{q,r_4} \]

Update the hidden \( y_q \) and other parameters by the EM-algorithm.

Subjectivity!

Resolve Labels: Learning from Crowds

Aggregate Predictions: Mixture of Opinions

observed labels

weighted based on relevance

predictions from reviews
Open-ended Questions

- Do not predict the answer directly, convert it to be a binary task.
- Preference prediction: Rank ‘true’ answers higher than non-answers

Resolve Labels

Resolve Labels

Include all answers in training

Loglik = \sum_{q} \sum_{a_q} \log p_{a_q > \bar{a}_q}

For each ‘true’ answer, randomly sample an non-answer from reviews

Aggregate Predictions

p_{a_q > \bar{a}_q} = P(a_q > \bar{a}_q | data) = \sum P(r | data) \times P(a_q > \bar{a}_q | r, data)

prediction from r

\alpha \exp(v_{q,r})
text relevance + helpfulness + expertise

\sigma(w_q, a_q > \bar{a}_q, r)

How relevant is r

For each 'true' answer, randomly sample an non-answer from reviews
Experiments (Binary Questions)

- Amazon QA Data
- Evaluation measure: Area Under Curve (AUC) for binary prediction
- EM-like methods (with real counts of labels) generally outperform others
- Subjective information only works on large categories
  - We have enough observations to model reviewer bias, expertise, etc.

Gold Standard: labels for non-ambiguous questions
Result (Open-ended Questions)

- **AUC for open-ended question**
  - measure if the model can successfully distinguish the ‘true’ answer from non-answers
  
  \[
  AUC = \frac{1}{|Q|} \sum_{q} \frac{1}{|A_q|} \sum_{a_q} 1(p_{a_q > \bar{a}_q} > 0.5)
  \]

- Including multiple answers consistently helps

- Incorporating subjective information was not effective

- Open-ended questions may not be as polarized as binary questions

- **AUC**
  - Single answer (the 1st answer) + text feature only
  - Multiple answers + text feature
  - Multiple answers + text feature + subjective feature
Conclusion and Future Direction

- Modeling ambiguity and subjectivity in opinion QA systems
- Resolve labels + Aggregate Subjective Classifiers
- Improvement from subjective information is limited
- A small gap between relevant reviews and direct answers
Thanks!

Mengting Wan (m5wan@eng.ucsd.edu)
Julian McAuley (jmcauley@eng.ucsd.edu)