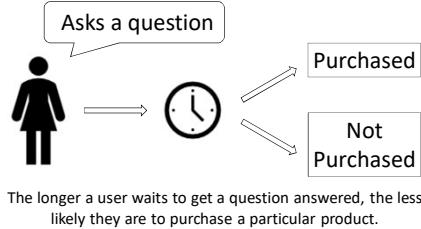


Answering Amazon product queries using relevant reviews

Mikhail Boulgakov, Silvia Jinrong Gong, Lavanya Satyan
Advised by Prof. Julian McAuley

Motivation & Background

It is important for sellers on Amazon to quickly answer users' questions. Our goal is to automatically answer a user's yes/no question about an Amazon product using the information contained in the reviews for that product.



We aim to build upon Moqa¹, a framework that predicts the answer to a question using a bilinear regression between a bag of words representation of a review and a question.

Problem Statement

We aim to improve the accuracy of the predicted answers by improving the measurement of the relevance of a review to the question. Improving the relevance helps us to determine which reviews contain the most useful and correct information, which allows us to answer the question as accurately as possible.

Dataset/Input

- Our data came from the Tools and Home Improvement category of Amazon².
- To train relevance, we created a dataset of 760 hand labelled question-review pairs labelled as relevant or irrelevant.
- To train voting, we used a dataset with 8280 questions labeled with yes/no answers and 148540 reviews.

1. Julian McAuley and Alex Yang. 2016. Addressing Complex and Subjective Product-Related Queries with Customer Reviews.
2. <http://jmcauley.ucsd.edu/data/amazon/qa/>

Model



Example Question: Does the router attachment work with the DeWalt Dw625 router?

Relevance

Break question and review into words and discard the stop words

Example:

Example relevant review: I used this attachment with the DeWalt Dw625 and it was great.

$R = \{ \text{used, attachment, dewalt, dw625, great} \}$

$Q = \{ \text{router, attachment, work, dewalt, dw625} \}$

Featurization:

$\frac{R \cdot Q}{ R \cdot Q }$	0.403	Cosine similarity finds the distance between the question and review
$ R \cap Q $	3	Number of common words between the question and review
OkapiBM25(R,Q)	2.75	OkapiBM25 finds the relevance of a document to a search query
	0	The 3 rd most uncommon word in the question is not in the review
$Q[-3:] \in R$	1	The 2 nd most uncommon word in the question is in the review
	1	The 1 st most uncommon word in the question is in the review

Output:

I used this router attachment with the DeWalt Dw625 and it was great	0.93
I like how portable it is	0.1
The router does indeed work	0.46

Voting

Take a dictionary of the 200 most common words, ignore stop words, and check how many times each word is contained in the question and in the reviews

Example:

Bag of words representation

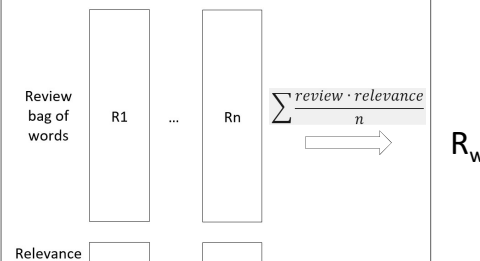
...	
does	— stopword
the	— stopword
router	— word
attachment	— word
work	— word
with	— stopword
the	— stopword
DeWalt	— not common
Dw625	— not common
router	— second time

...	
item	0
work	1
broom	0
...	
clean	0
attachment	1
play	0
...	
red	0
router	2
never	0
...	

Q

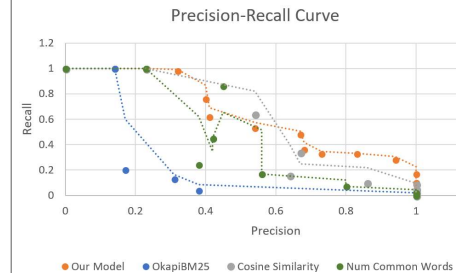
Weighted average of reviews

We use the relevance scores as the weights



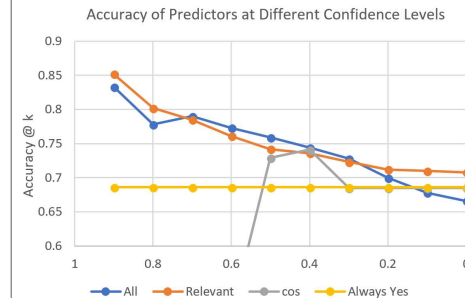
Cosine similarity between question (Q) and weighted reviews (R_w)

Results - Relevance



We tested 4 different relevance functions: our model, OkapiBM25 (used by search engines), cosine similarity between question and review, and number of common words between question and review. The respective areas under the curve (a measure of a relevance function's performance) are .64, .20, .59, and .45. Our model outperformed all the other functions for this task.

Results - Voting



We tested 4 different models: our model trained on all reviews, our model trained only on relevant reviews, cosine similarity between review and question, and always predicting yes. For the 50% most confident predictions, we got 74.5% accuracy, which is better than Moqa performed for this category (73%).

Future Work

Our model is currently only able to answer binary questions. In the future, we also want to be able to answer open ended questions.