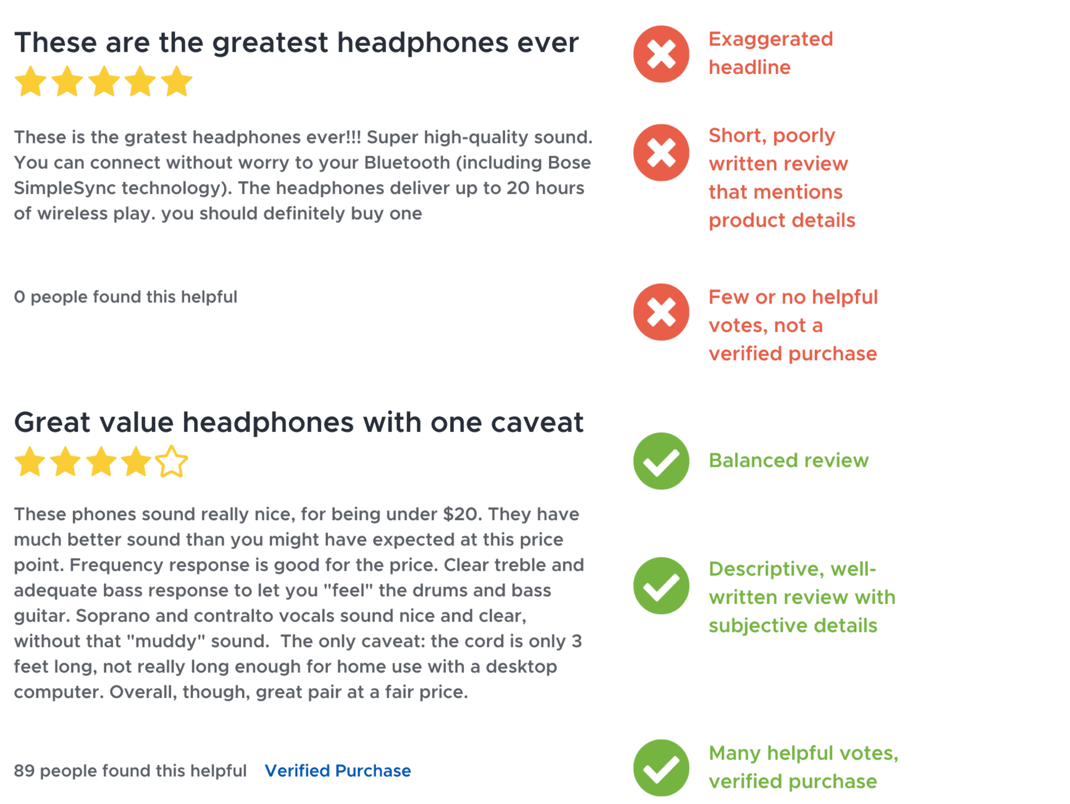
On the Role of Reviewer Expertise in Temporal Review Helpfulness Prediction

EACL 2023

Mir Tafseer Nayeem and Davood Rafiei
University of Alberta

What is a Helpful Review?

- Provides useful and informative feedback to potential customers
- Contains specific details about the product or service
- Usually includes both positive and negative aspects
- Helpful review may include suggestions for improvement



Importance of Helpful Reviews

- Help customers make quick purchase decisions
- Benefit the merchants in their sales

Challenges

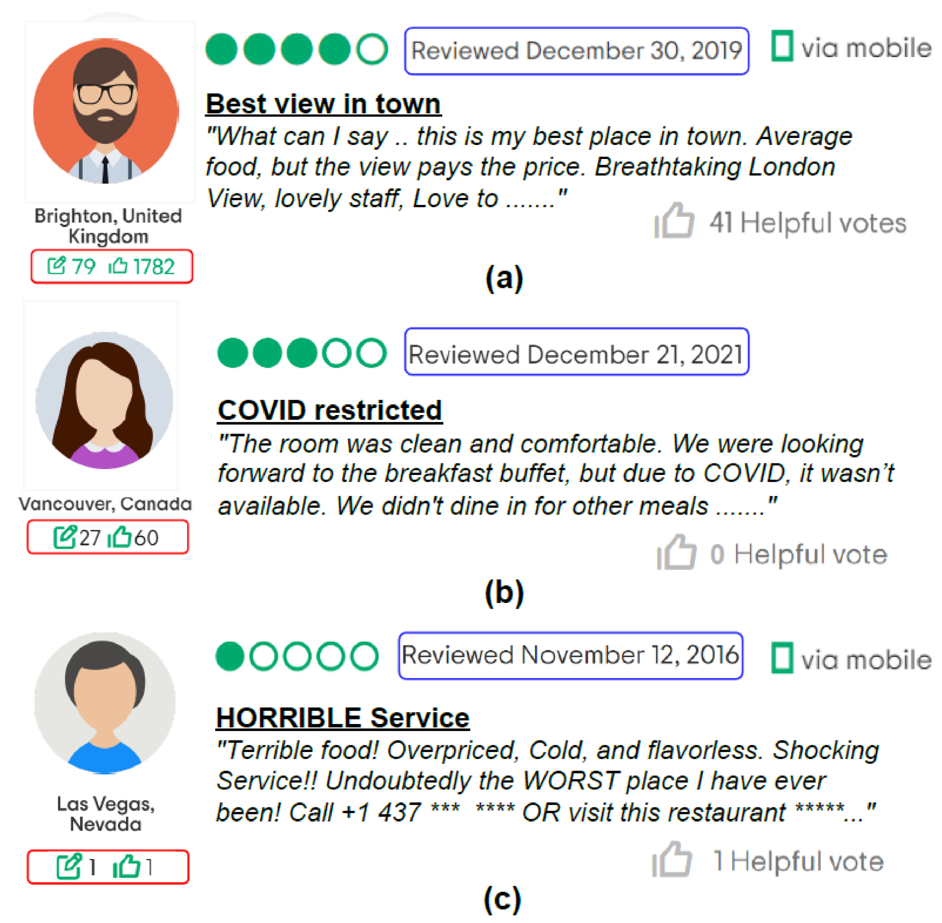
- User reviews may contain spam, excessive appraisal, or unexpected biases.
- Multiple factors that affect the quality of a review. These factors are not usually explicit in the review text **1) Reviewers' life experience 2) Educational background 3) Motive for writing the review**
- Customers usually have limited patience for reading reviews – most customers read less than 10 reviews before making a purchase decision.
- Large volume of reviews and their unpredictable quality and the limited customer patience demand better review utilization strategies.

Contributions

- We introduce a new dataset with both review text and reviewer's history, to highlight the importance of integrating the two sources for review helpfulness.
- We propose a Review Helpfulness Prediction (RHP) model incorporating the reviewer's expertise and temporal information, especially for unreliable and cold-start reviews.

Assumptions

- People who post more reviews and earn more helpful votes are more likely to be better reviewers.
- Trustworthy reviewers are less likely to be posting fake or biased reviews, and their reviews are more likely to earn more helpful votes; otherwise, they will be ruining their reputation.
- Those who have been to more hotels or restaurants across different cities have a better basis for comparison and writing critical reviews.



Review a has accumulated more helpful votes but is posted almost two years before **Review b**; on the other hand, **Review b** (*a.k.a., cold-start review*) contains time-sensitive information, describing the current conditions and **Review c** is likely a spam review.

RHP Model

A supervised machine learning task where the input contains information about the **reviews (R)** and the **reviewers (U)**.

$$\mathcal{R}_i = ([s_1, \ldots, s_N],\ t_i)$$

$$\mathcal{U}_i = (n_i,\ m_i)$$

- S = Review Sentence
- t = Review time
- n = # review posted
- m = # helpful votes received

We formulate the task where we seek to find a model f that minimizes the loss function.

$$min_{\theta} \mathcal{L}\ (f(\theta, \mathcal{R}, \mathcal{U}),\ Y)$$

We encode the review sentences using BERT.

$$[\mathrm{h}^{[\mathrm{CLS}]}, \mathrm{h}^{(1)}, \mathrm{h}^{(2)}, \ldots] = \mathrm{BERT}([\mathrm{CLS}]\ s_1, \ldots, s_N\ [\mathrm{SEP}])$$

$$x_h = \Theta\ (\mathrm{MLP}\ (\mathrm{h}_l^{[\mathrm{CLS}]})).$$

We also integrate reviewer expertise and temporal information of the reviews.

Integrating Reviewer Expertise and Time

Expertise

- Reviewers who post more reviews and earn more helpful votes are likely to be better reviewers.
- Such users may have been to more hotels and restaurants across the globe and have a better basis for comparison.

Temporal Information

- Older reviews are more likely to accumulate more helpfulness votes than newer reviews but are not necessarily the most relevant describing the current conditions (*e.g., new COVID restrictions*).
- One-time problems such as broken bathrooms and dirty pool area are likely to be addressed and to be less relevant.

Reviewer expertise is the mean number of helpful votes received per review.

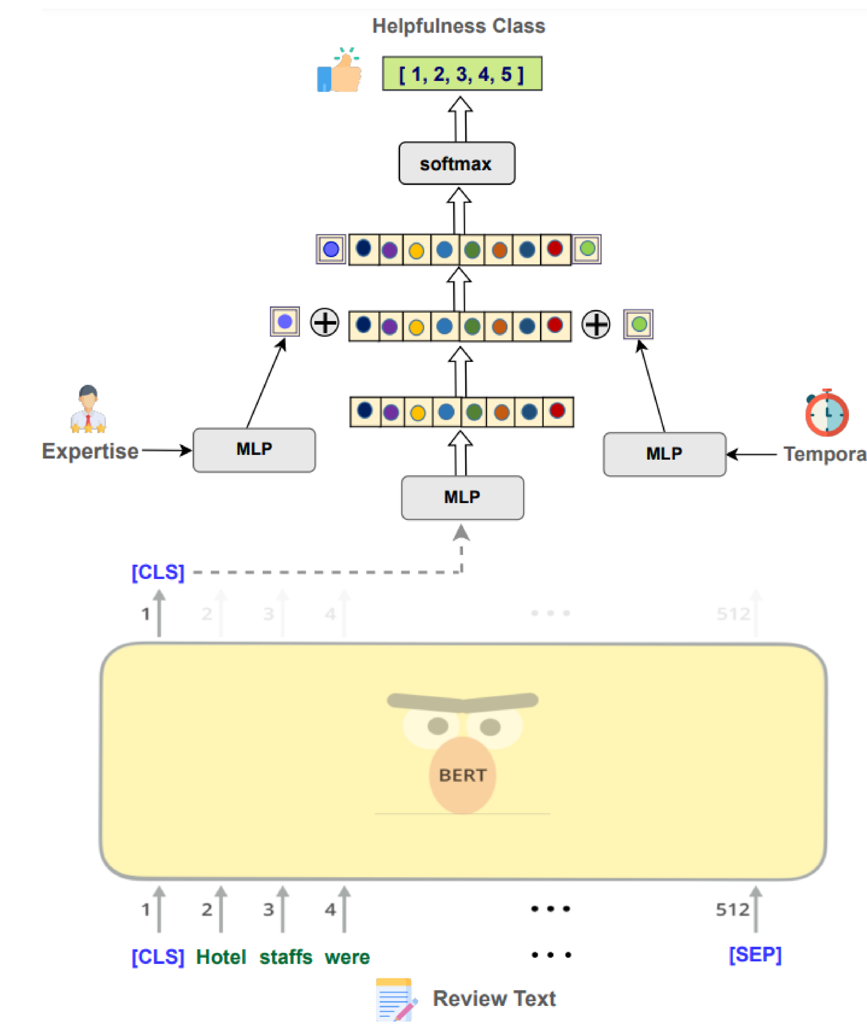
$$h_s = \mathrm{MLP}\ (e_s)$$

td be the relative age of a review in days, for example, as of the day the reviews are scraped.

$$h_t = \mathrm{MLP}\ (t_d)$$

Both the review age and the reviewer expertise are normalized to a fixed range $[a, b]$

$$z_i = (b-a)\frac{x_i - min(\mathcal{X})}{max(\mathcal{X}) - min(\mathcal{X})} + a$$



$$o_{final} = h_s \oplus x_h \oplus h_t$$

$$\hat{Y} = \mathrm{softmax}\ (W_r \cdot o_{final} + b_r),$$

$$\mathcal{L} = \mathcal{L}_{CE}\ (\hat{Y}, Y)$$

Experimental Results

Baseline Models	Acc. (↑)	MAE (↓)	MSE (↓)
ARH	58.73	0.476	0.619
UGR + BGR	62.76	0.464	0.674
TextCNN	62.82	0.444	0.608
MTNL	62.77	0.458	0.653
BERTHelp	63.03	0.432	0.591
Our Ablations	**Acc. (↑)**	**MAE (↓)**	**MSE (↓)**
RHP (*ours*)	**65.18**†	**0.393**†	**0.491**†
- *w/o Expertise*	63.87	0.421†	0.550†
- *w/o Temporal*	63.40	0.437†	0.592
- *w/o Expertise + Temporal*	62.92	0.446	0.617

Table 2: Performance compared to our baseline models and the result of our ablation study (↑ indicates higher values for a better performance and ↓ indicates lower values for a better performance). † reported results are statistically significant in paired t-test by taking BERTHelp (Xu et al., 2020) as a reference with the confidence of 95% (p-value < 0.05).

Analysis

Helpfulness Class	Unigram	Bigram
Class #1 Helpful Votes [1, 2)	'room' 'staff' 'location' 'time' 'service'	'front desk' 'coffee maker' 'breakfast buffet' 'sofa bed' 'swim pool'
Class #2 Helpful Votes [2, 4)	'room' 'staff' 'service' 'location' 'time'	'front desk' 'shampoo conditioner' 'customer service' 'resort fee' 'pool area'
Class #3 Helpful Votes [4, 8)	'room' 'staff' 'time' 'service' 'view'	'front desk' 'resort fee' 'customer service' 'coffee maker' 'city view'
Class #4 Helpful Votes [8, 16)	'room' 'staff' 'service' 'time' 'pool'	'front desk' 'resort fee' 'customer service' 'minute walk' 'life jacket'
Class #5 Helpful Votes [16, ∞)	'room' 'time' 'service' 'staff' 'pool'	'front desk' 'resort fee' 'bed bug' 'beach chair' 'cable car'

Table 3: Top 5 unigrams and bigrams extracted from five different classes of reviews divided according to helpfulness votes. For each column, green color indicates the overlap with all 5 classes, whereas blue for 4, orange for 3, and red for 2 overlaps.

We extract **Top K (where K = 5)** n-grams from each class of reviews to identify the **most relevant keywords or topics** in reviews to assess **what aspects are most talked** about the items.

Case Study

[Free WiFi, Free parking, Location, Room, Staffs, Front Desk, Food, swimming pools, foods, Bar, Air conditioning, Non-smoking rooms, Fitness center, ATM on site, Shuttle service, Room service, Spa, ...]

Aspects / Facilities

[CLS] We could not have been happier with our choice for our family's 3 night stay in Las Vegas recently. The location was perfect. We stayed in a 2 bedroom villa, which was so spacious and had a great view of the Vegas lights and airport ....The bathroom to the main bedroom had a fabulous big bath. The beds very comfortable. Dinner in the restaurant in the lobby one night, the food and service were both great. We particularly liked the restaurant and bar next to the pool on level 5, very relaxing for lunch [SEP]

Review Text

Figure 3: Top 10 ranked tokens of the RHP model shown in green colors with the color intensity indicating the importance of the tokens in the overall prediction.

- **Top-ranked words** are highly representative of the **aspects or facilities** listed on the restaurant page.
- Use of **personal pronouns (e.g., I, we, they, etc.)**, describing **personal experiences**, contributes to the helpfulness prediction.

Limitations & Future Work

- How to incorporate **personal preferences** to model the **demographics** and **cultural differences** for this task.
- We aim to extend this work to support **more languages** and **code-switched reviews** when the reviewers **alternate** between two or more languages in a single review.