

Addressing Overchoice: Automatically Generating Meaningful Filters from Hotel Reviews

István Varga, Yuta Hayashibe
Megagon Labs, Recruit Co., Ltd., Japan
September 26 2021

Motivation

- Online booking systems brought forward the benefit of **an increased selection** of choices
- Having too many options to choose from leads to **overchoice**
 - **Too many similar choices** lead to extra cognitive load
 - Consumers **can't grasp all information** about a choice, making them less convinced that they made the best decision

Choosing the right hotel can be even more difficult

- Hotels are extremely complex items
 - Large number of features to consider
 - The weight of the decision is big
- Faceted search offers **objective filters** (e.g., *accommodation type, room facilities*, etc.)

The image shows a screenshot of a hotel search interface. On the left, a sidebar contains various filters, all of which are currently unchecked. These filters are: Price (range \$0 to \$4000), Accommodation Type (Hotel, Ryokan, Villa, Public hotel, Pension etc), Room Type (Single room, Twin room, Double room, Triple room, Quadruple room, Japanese-style room, Japanese-Western style room, Other), Meals (No meal, With breakfast, With dinner, With breakfast/dinner), Room facilities (Non-smoking, Smoking), and Internet (Wi-Fi (free), Wi-Fi (paid), LAN cable (free), LAN cable (paid)). A red border highlights this sidebar. To the right, the main content area displays search results. The top result is for 'HOTEL UNIKAWA TATE' with a 4.3 rating and a reservation note. Below it are three room options: 'Standard Double room, Non-smoking', 'Standard Twin room, Non-smoking', and 'Twin room, Non-smoking', each with a 'With Breakfast' icon and a reservation note. The second result is for 'Rihga Royal Gran Okina' with a 4.7 rating and a reservation note. Below it are two room options: 'Superior Twin room, Non-smoking, Harbor view' and 'Superior Twin room, Non-smoking, City view', each with a 'With Breakfast' icon and a reservation note.

Choosing the right hotel can be even more difficult

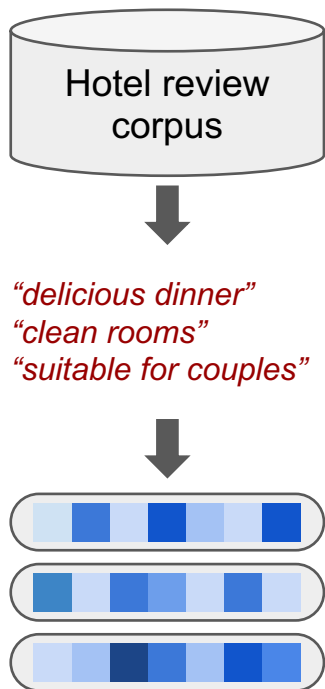
- Hotels are extremely complex items
 - Large number of features to consider
 - The weight of the decision is big
- Faceted search offers **objective filters** (e.g., *accommodation type*, *room facilities*, etc.)
- **Subjective experiences or fine-grained details** are the decisive factors, sometimes only found in customer reviews

The screenshot displays a hotel search interface with various filters and search results. On the left, there are filter sections for 'Price' (ranging from \$0 to \$4000), 'Accommodation Type' (Hotel, Ryokan, Villa, Public hotel, Pension etc.), 'Room Type' (Single room, Twin room, Double room, Triple room, Quadruple room, Japanese-style room, Japanese-Western style room, Other), 'Meals' (No breakfast, With breakfast), 'Room facilities' (Non-smoking, Smoking), and 'Internet' (Wi-Fi, LAN cable, LAN cable (free), LAN cable (paid)). On the right, search results are shown for 'Rihga Royal Gran Okina' in Okinawa > Naha > Naha. The results include a 4.3-star rating, a 'Reservation possible without a credit card' badge, and several room options: 'Standard Double room, Non-smoking', 'Standard Twin room, Non-smoking', and 'Twin room, Non-smoking', each with 'With Breakfast' and 'Reservation possible without a credit card' options. A red text overlay on the right side of the screenshot lists: 'Suitable for families', 'Great view of Mt. Fuji', 'Clean rooms', 'Delicious breakfast', and '...'. The background of the search results shows a scenic view of a beach and a large umbrella.

Proposal overview

- We propose a simple **clustering** based framework to automatically identify filters related to hotels that already match the user's initial query
- We focus on **customer experiences** and quality judgements extracted from **customer reviews**
- We **define and implement key concepts** to identify filters that reduce the choice set in an intuitive way
 - **Size control rules** as hard constraints in acquiring filters
 - **Relevance** and **uniqueness** to rank filters

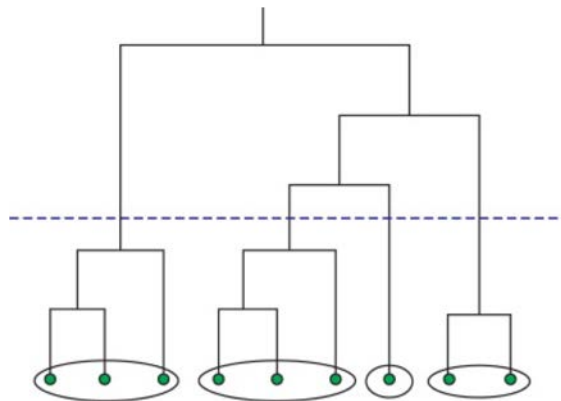
Data source and filter representation



- Corpus: 20 million sentences extracted from hotel reviews, collected from jalan.net (in Japanese)
- Filters units: predicate-argument structures extracted from customer reviews
- 167,886 unique filter candidates with non-negative polarity that are related to the hotels or hotel services
- Filter representation: sentenceBert embeddings of filters units
 - Pre-trained BERT on the review corpus
 - Fine-tuned with sentenceBert with triplet-loss function

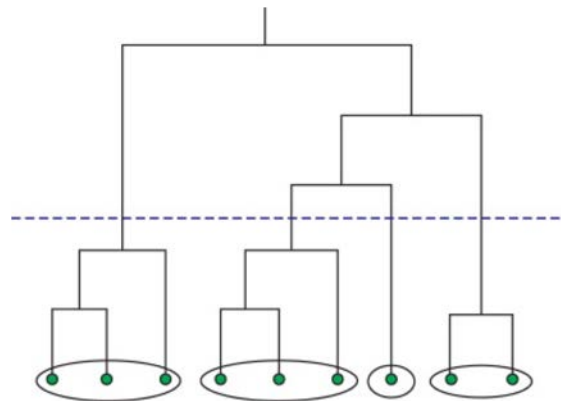
2 stage filter acquisition

- **Stage 1:** cluster filter candidates from reviews that match the user query, identifying latent topics (e.g., *food*, *location*, *hot-spring*, etc.)
 - Ward's agglomerative clustering with cosine similarity as metric



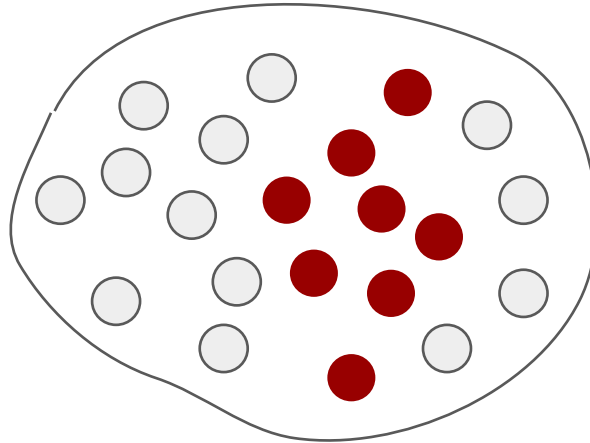
2 stage filter acquisition

- **Stage 1:** cluster filter candidates from reviews that match the user query, identifying latent topics (e.g., *food*, *location*, *hot-spring*, etc.)
 - Ward's agglomerative clustering with cosine similarity as metric
- **Stage 2:** for each topic, we adjust the link thresholds of the hierarchical subtrees to identify filter clusters that obey *size control rules*
 - Score and rank each resulting cluster C_x by combining *relevance* and *uniqueness*
$$\text{rank}(C_x) = (\alpha + \text{relevance}(C_x)) \times (\beta + \text{uniqueness}(C_x))$$



Size control rules

- Filters are designed to address overchoice and reduce the choice set
- The **degree** of the size reduction is also crucial (not too drastic or too shallow)
- A set of *lower_bound* and *upper_bound* thresholds (%) which guarantee that a filter reduces the choice set in a reasonable way



Relevance acquisition

- **Relevance:** the usefulness or popularity of a filter (e.g., *close to the city center* >> *bright pink curtains*)
- K-nearest neighbour classifier with similarity as weight

$$\text{relevance}(x) = \frac{\sum_{i=1}^k \text{cossim}(x, x_i) \times \text{relevance}_{\text{gold}}(x_i)}{\sum_{i=1}^k \text{cossim}(x, x_i)}$$

- Training data: 8000 filters, scored from 5 (most relevant) to 1 (least relevant) by 5 crowd workers
 - Pairwise Cohen's Kappa showed fair to moderate agreement (between 0.24 and 0.56), thus we used the truncated mean of the workers' scores

Uniqueness acquisition

- **Uniqueness:** the representativeness of a filter within the choice set (with the potential to be unknown to the customer) (e.g., *close to the city aquarium, private hot-spring*)
- Tf-idf as uniqueness score
 - To prevent sparseness, we pre-clustered semantically similar filters
 - Individual filters inherit the uniqueness scores of their parent cluster

$$\text{uniqueness}(x, d) = \text{tf}(x, d) \times \text{idf}(x)$$

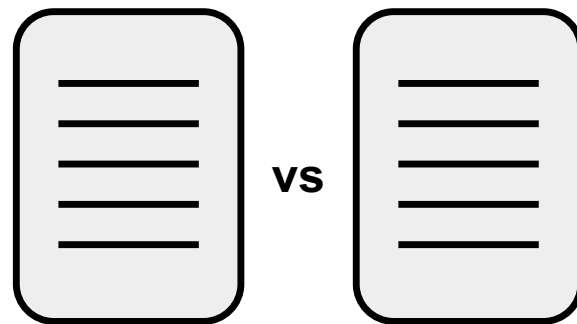
x : group of semantically similar filters
 d : all filters of a given hotel

Experiment settings

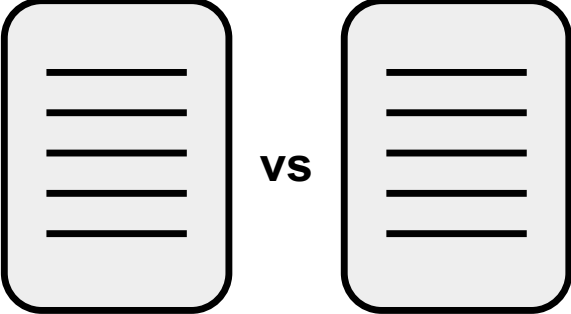
- Comparative evaluation of our **proposal** against a set of baselines
- **Human**: manually compiled set of filters
 - 10 (query, location) tuples (e.g., *Relaxing atmosphere*, @Nagano)
 - 3 annotators were asked to manually extract the most useful filters from all hotel reviews that match the input query
- **Relevant**: our proposal without uniqueness
- **Unique**: our proposal without relevance

Filter list evaluation

- Compared the top 5 filters of the competing models
- 10 (query, location) tuples
- 300 crowd workers chose the more useful of the competing lists

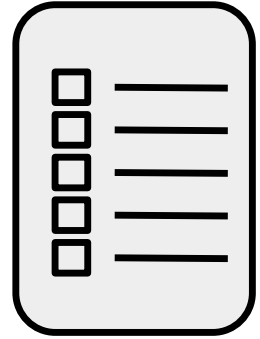


Filter list evaluation

- Compared the top 5 filters of the competing models
 - 10 (query, location) tuples
 - 300 crowd workers chose the more useful of the competing lists
- 
- **Proposed** significantly outperformed **human** (5/10), with highly specific (e.g., *delicious food with local ingredients*), and very localized or fine-grained filters (e.g., *the splendid alfonsino was very delicious*)
 - **Human** significantly outperformed **proposed** (2/10) with filters that are relevant, but not unique enough (e.g., *large room, clean hotel*)
 - **Proposed** also outscored **relevant** and **unique**, with **relevant** having similar behaviour to **human**

Individual filter evaluation

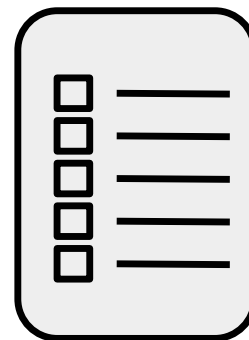
- Compared the mixed outputs of proposed and the baselines
- 10 (query, location) tuples
- 300 crowd workers chose the most useful individual filters from the mixed list



Individual filter evaluation

- Compared the mixed outputs of proposed and the baselines
- 10 (query, location) tuples
- 300 crowd workers chose the most useful individual filters from the mixed list

- **Proposed** significantly outperformed **human** (5/10) with filters expressing experiences or quality judgements (e.g., *the free breakfast was delicious*) as opposed to factual filters (e.g., *free breakfast is available*)
- Highly relevant but less unique (e.g., *excellent service, free wifi*) or highly unique but less relevant filters (e.g., *karaoke machine is available*) also received numerous votes



Quality judgements always more valuable than factual filters?

- Evaluation results hinted that **quality judgements** are more appealing than **factual filters**
- We asked 20 crowd workers to decide between 30 (*quality judgement, factual*) filter tuples (e.g., *food was delicious* versus *food available*) across multiple topics
- Quality judgements were preferred with the majority of the topics (e.g., *location, food, hot-spring*, etc.)
- Factual filters are still preferred with topics where experience or quality is not very important (e.g., *parking: free parking available* >> *the parking space was very accessible*)

Key takeaways and future work

- We proposed a simple, **hierarchical clustering based approach** to identify customer experiences as potentially interesting filters in the hotel industry domain, using customer reviews
- Customers have a strong preference towards experience based or **quality judgments over factual filters**
- We still need to investigate how to incorporate **subjective filters** into real-life hotel booking systems
- Look into **other factors** besides relevance and uniqueness (and their importance)