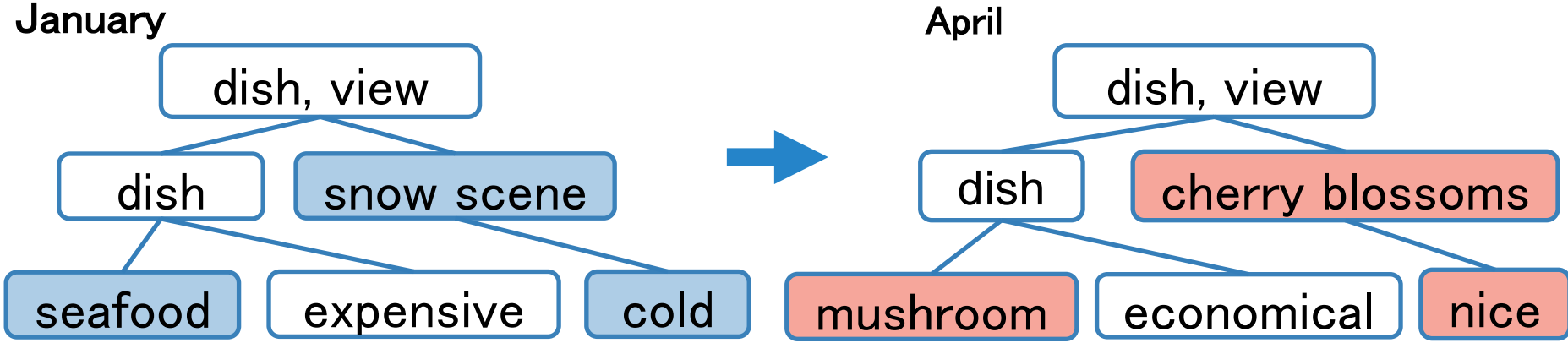
Time-Variable Analysis of Accommodation Reviews Based on Hierarchical Topic Model

The Thirteenth International Conference on Advances
in Computer-Human Interactions ACHI 2020
Graduate School of Information Science and Engineering, Ritsumeikan University

Yujiro SATO, Ryosuke YAMANISHI and Yoko NISHIHARA
Mail: ryama@kansai-u.ac.jp

1 Abstract

1. We propose a method to extract seasonal features from accommodation reviews.
2. We evaluated the change of the similarity among the features.



2 Background

Popular accommodations have many reviews.
The required information cannot be found.

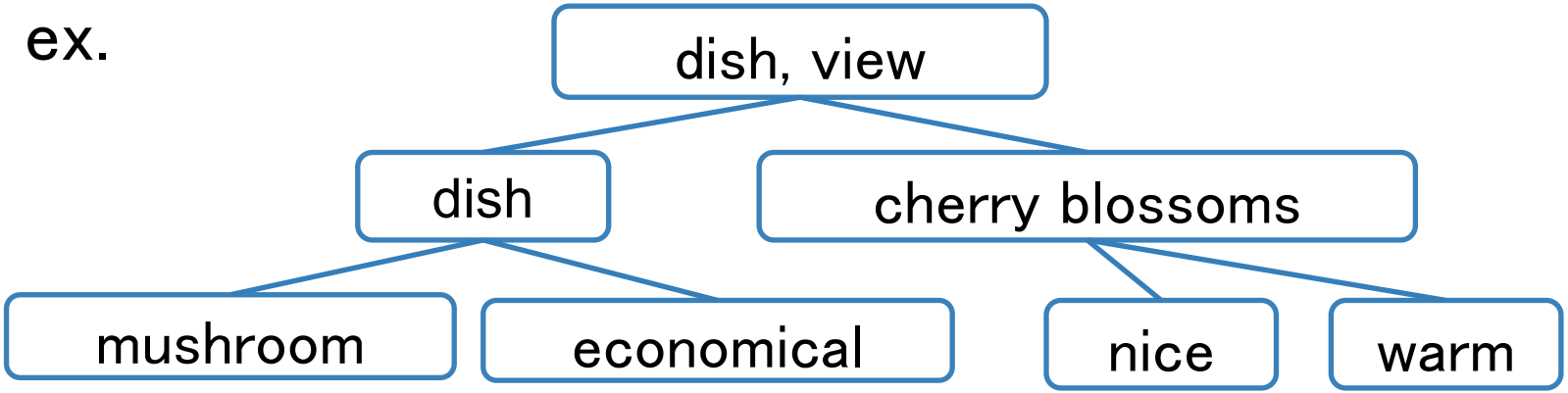
You can only see 8 or 7 reviews per page.

- It is not possible to discover seasonal changes and the characteristics of the season in which you plan to stay.

3 Purpose of study

We extract the features of accommodation facilities in a **hierarchical structure**

ex.



1. Abstract

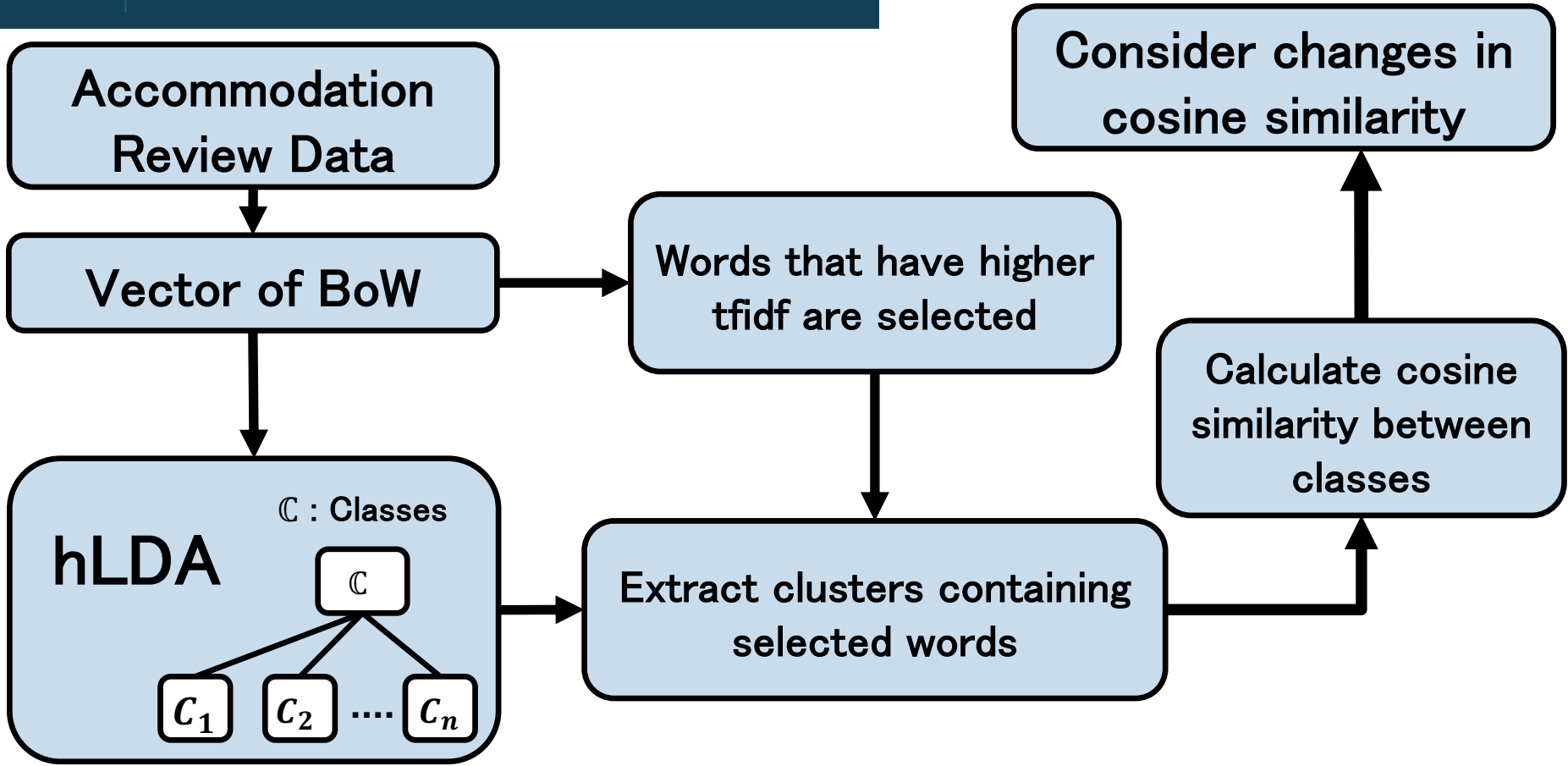
2. The proposed method

3. Result of analysis

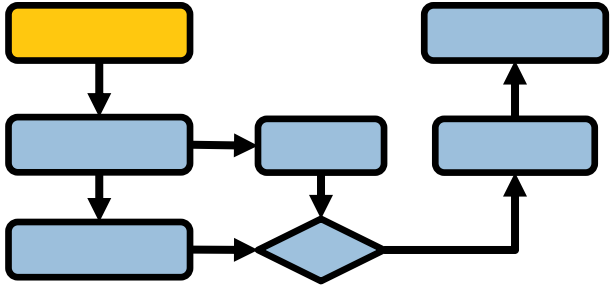
4. Discussion

5. Conclusion

4 Framework of method



5 Accommodation Review Data



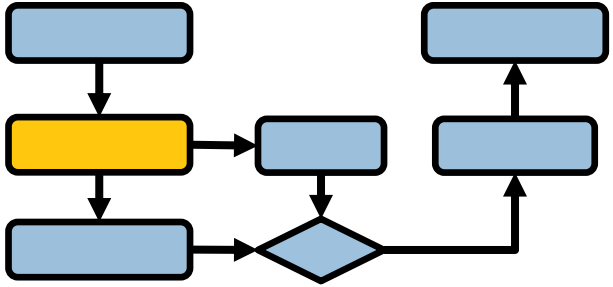
Rakuten Travel reviews

- During year of 1996 through 2016.
- The top 10% of accommodations with the highest number of reviews.

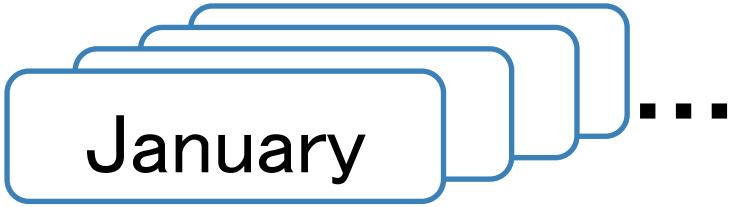
◆#1 Accommodations（3,118 reviews）

◆#2 Accommodations（2,770 reviews）

6 Text vectorization on BoW



Splitting reviews by month



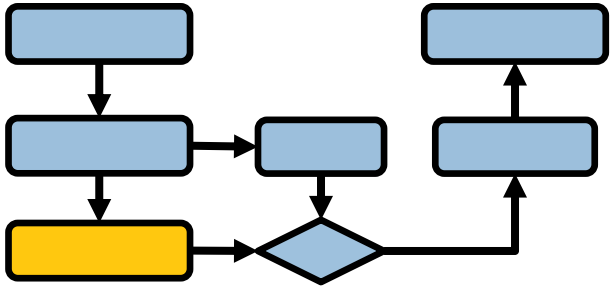
Vectorized based on the BoW

	"place"	"stay"	"room"	"host"
Review1	1	3	1	4
Review2	2	2	0	0
Review3	3	0	3	0

input → hLDA

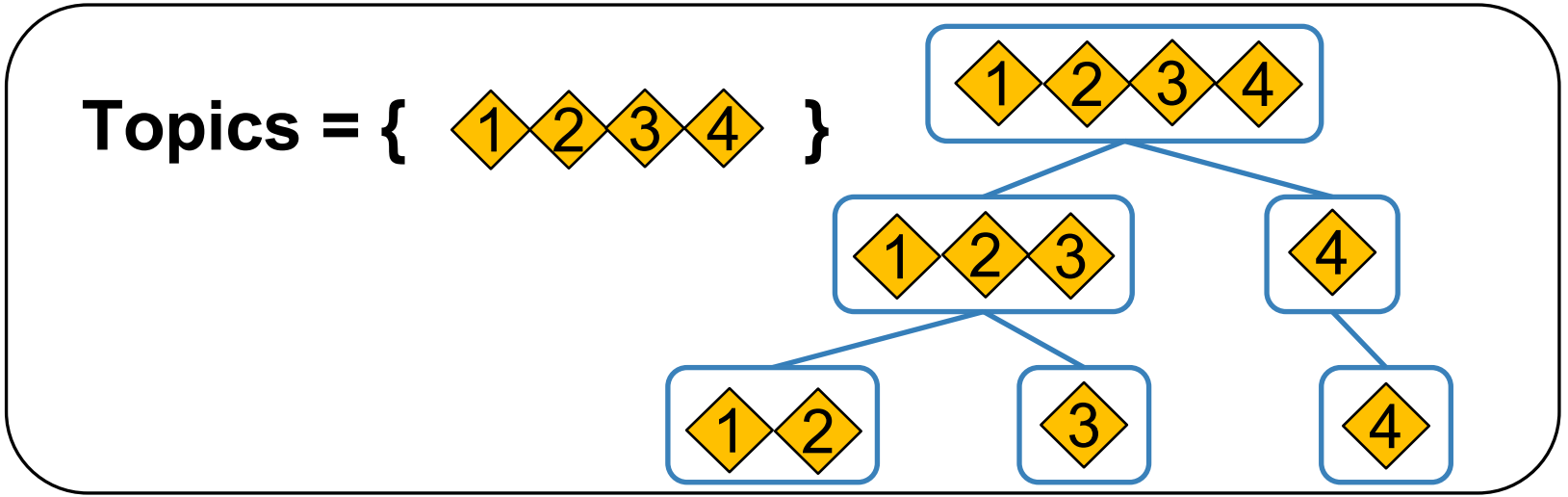
input → tfidf

7 hLDA(1/2)

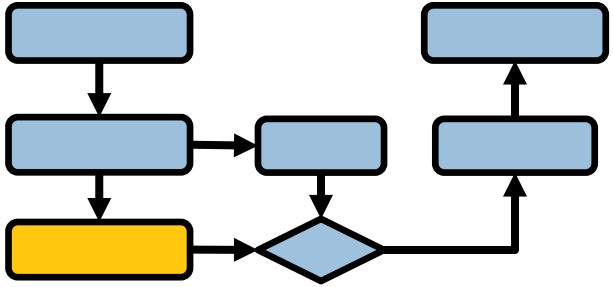


hLDA (hierarchical Latent Dirichlet Allocation)

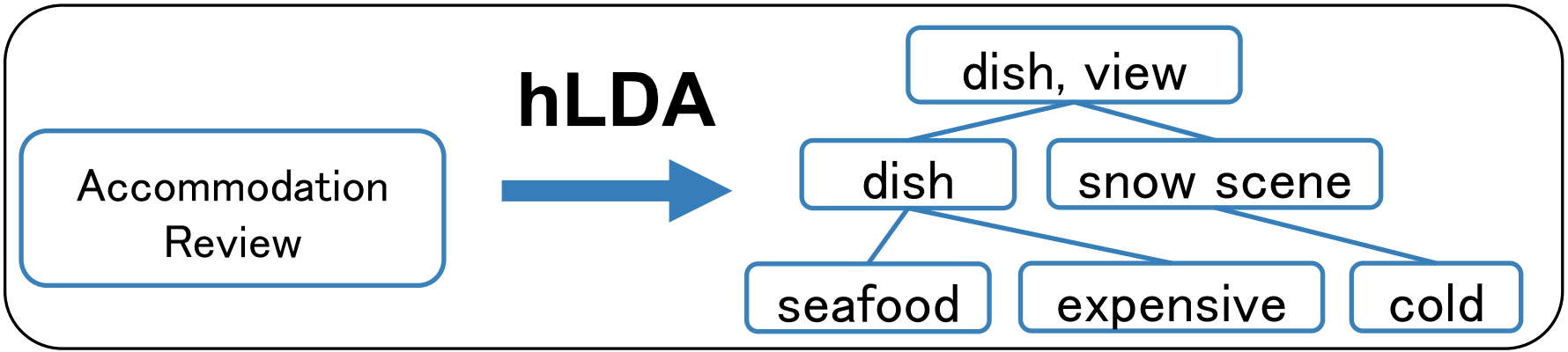
✓ The hLDA analyzes the hierarchical relationship of topics.



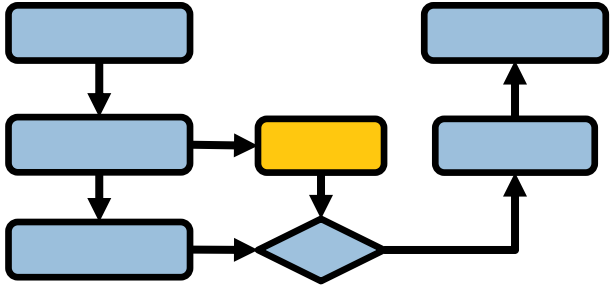
8 hLDA(2/2)



We hypothesized that hLDA could be used to extract accommodation topics from accommodation reviews.

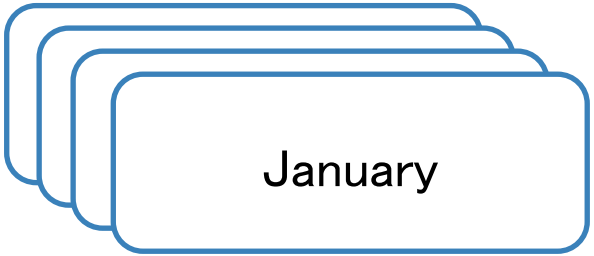


9 TF-IDF



- ✓ TF-IDF can be used to compare the relative frequency of words in a document with the inverse ratio of words throughout.
- ✓ In this study, nouns with high values can be said to be seasonal feature words, because monthly documents are used.

Monthly Documents

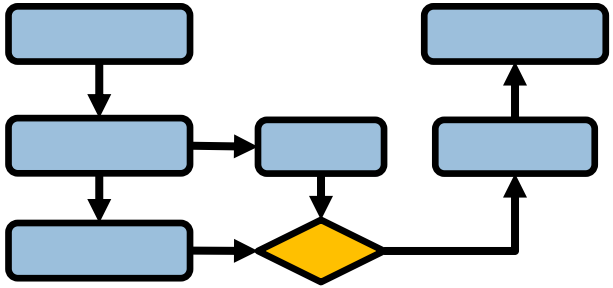


TF-IDF

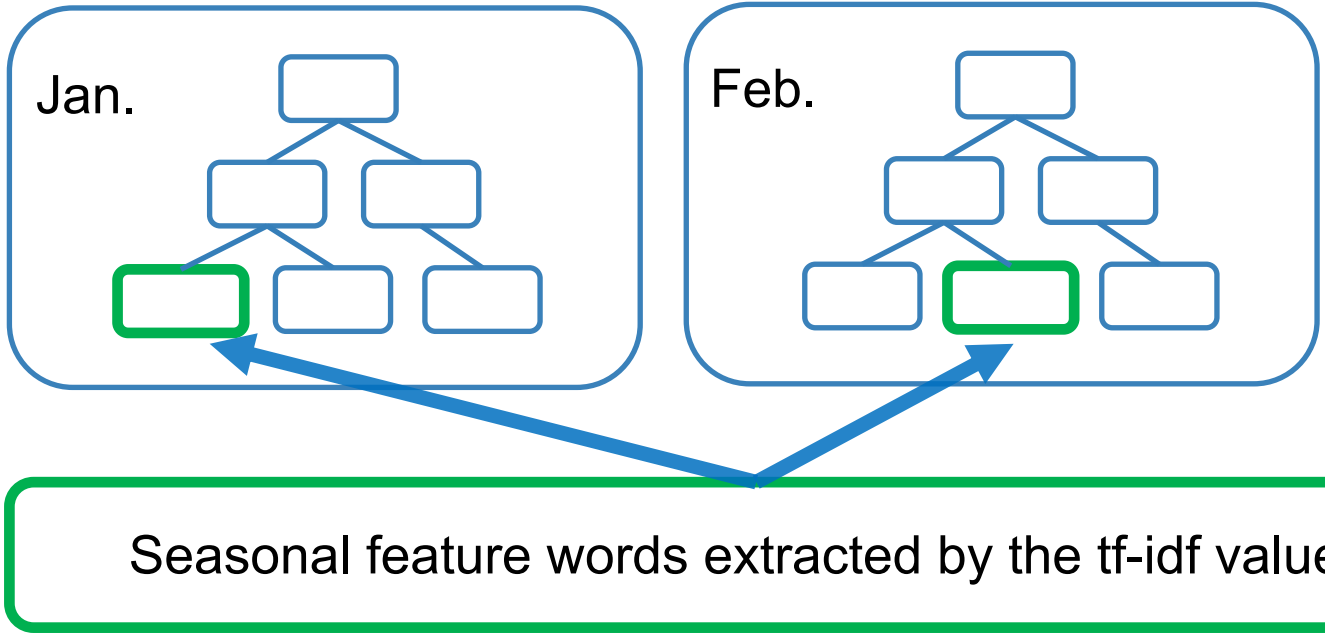
Seasonal feature words

The top 10% nouns with TF-IDF values

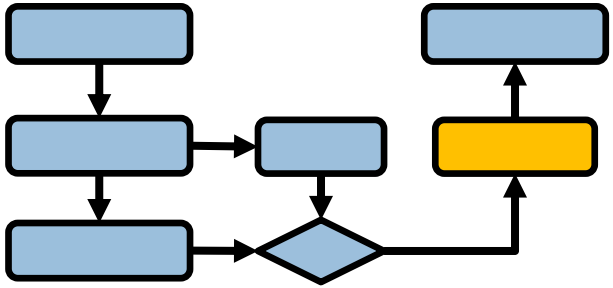
10 Focusing on the clusters



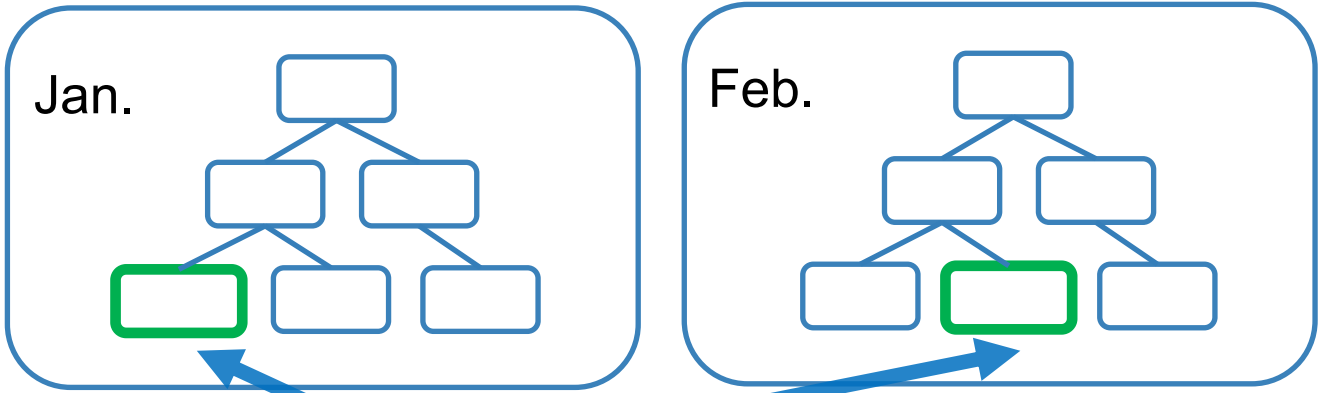
We extract clusters containing seasonal feature words



11 Cosine similarity(1/2)

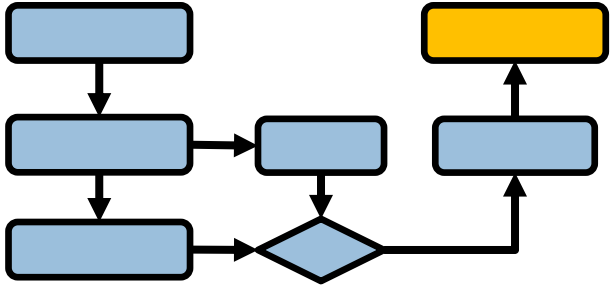


We evaluate the similarity between the clusters

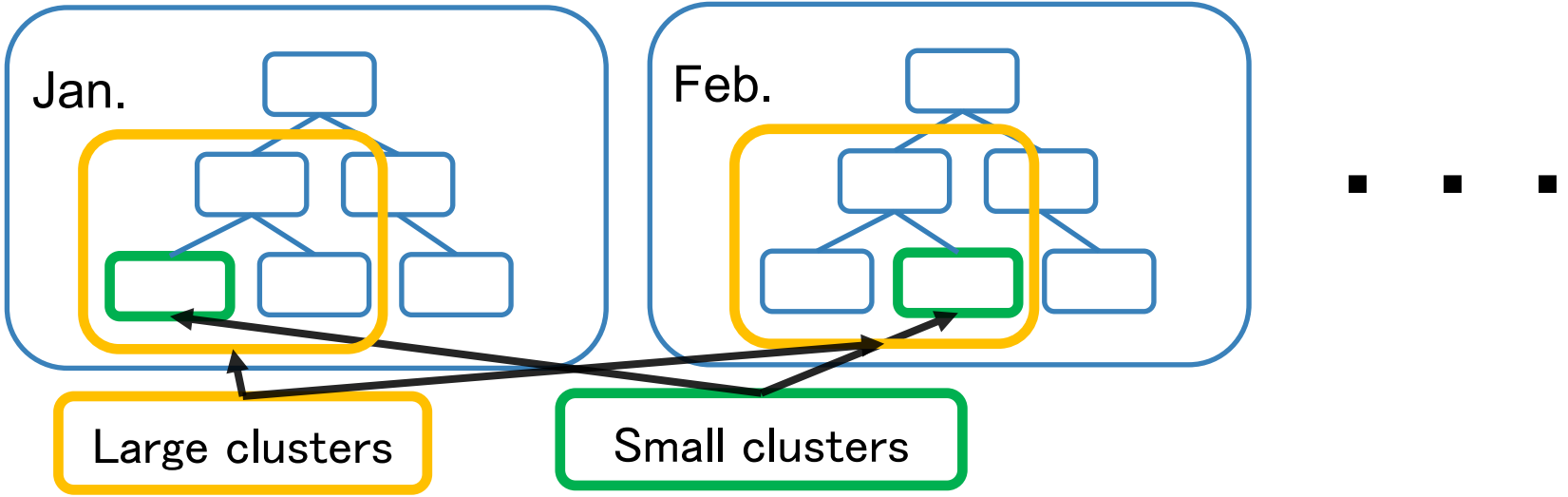


High: The clusters have similar features.
Low: Seasonal features have changed.

12 Cosine similarity(2/2)



We evaluate the amount of change in similarity



✓ We extract seasonal changes from two types of clusters.

1. Abstract

2. The proposed method

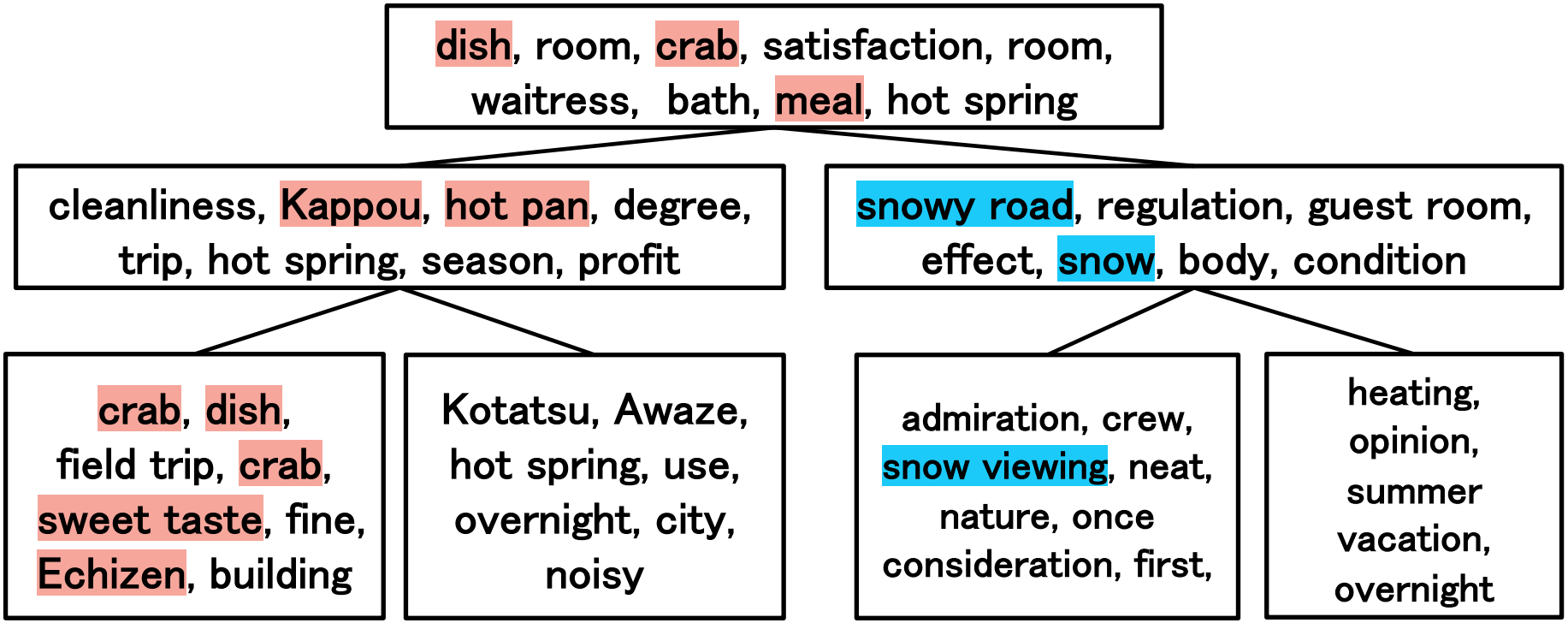
3. Result of analysis

4. Discussion

5. Conclusion

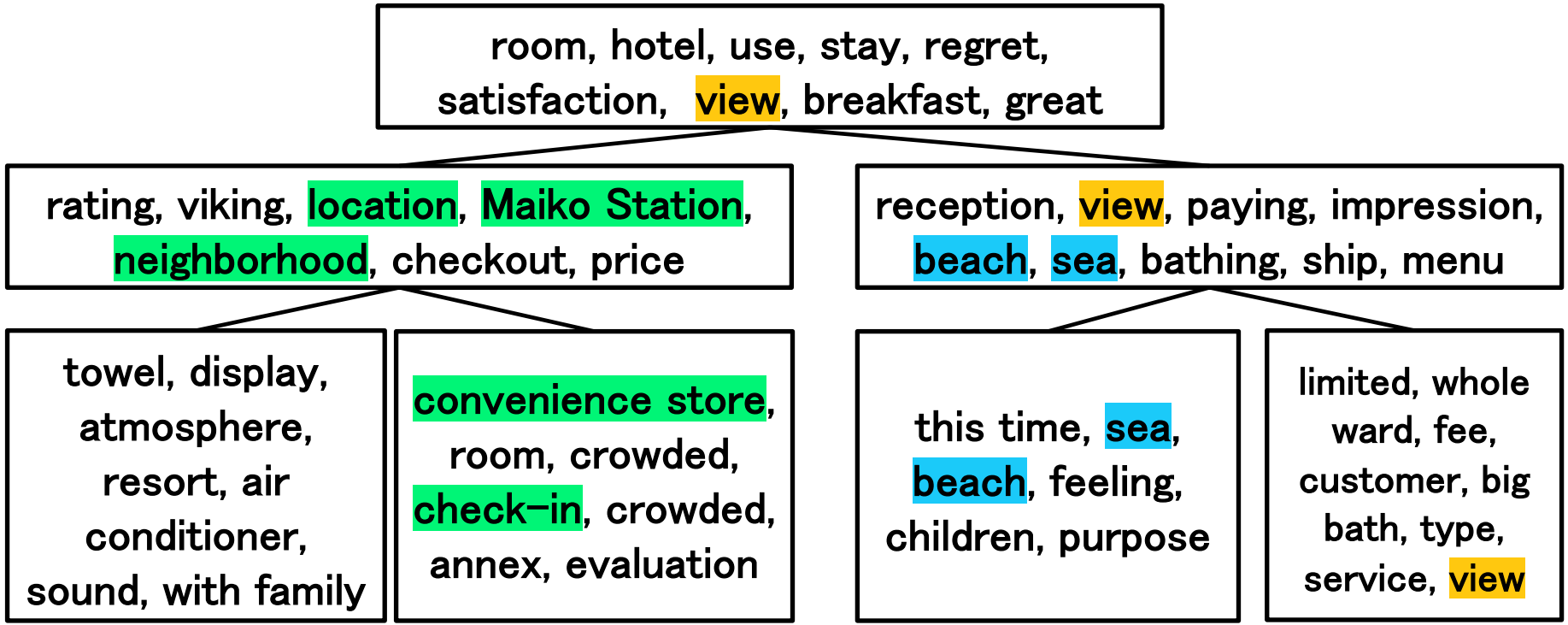
13 The result of hLDA(1/2)

The result of hLDA for accommodation #1 in January.



14 The result of hLDA(2/2)

The result of hLDA for accommodation #2 in August.



15 The result of TF-IDF

Top 5 nouns with high TF-IDF values

✓ TRANSLATED INTO ENGLISH BY THE AUTHORS

	Accommodations #1	Accommodations #2
Jan.	new year, new year's day, new year's end, **mochi pounding**, superlative degree	new year, new year's, sweets, special, anniversary
May	Golden Week, **red snapper**, spring, love, holiday	room temperature, Golden Week, weekend, grade, beauty treatment
Aug.	Obon, Gassho, Rokusaburo Michiba, **sweetfish**, Noryo	**pool**, summer vacation, **beach**, **sea bathing**, **barbecue**
Dec.	christmas, meal, breakthrough, specialty, superlative degree	**Luminarie**, christmas, win- ter, hospitality, special

1. Abstract

2. The proposed method

3. Result of analysis

4. Discussion

5. Conclusion

16 Discussion (1/3)

Accommodations #1 focus on “Red snapper”

month	Jan.~ Mar.	Apr.	May	June	July	Aug.	Aept.	Oct.	Nov.~ Dec.
Large	0.0	0.029	**1.0**	0.046	0.018	0.018	0.021	0.028	0.0
Small	0.0	0.099	**1.0**	0.199	0.099	0.099	0.099	0.105	0.0

✓ People can enjoy “Red snapper” from April to October

✓ Similar features can be seen from April to June

Nouns appeared with ``Red snapper'' were cuisines: KAISEKI (Traditional multi-course meal), nice smell, CHAZUKE (rice in Japanese broth).

17 Discussion (2/3)

Accommodations #2 focus on "pool"

month	Jan.~ May	June	July	Aug.	Sept.	Oct. ~ Dec.
Large	0.0	0.110	0.042	**1.0**	0.018	0.0
Small	0.0	0.099	0.099	**1.0**	0.099	0.0

- ✓ Reviews changed drastically due to low similarity in Jul. and Sep.
- ✓ Many families instead of single travelers stay at the accommodation in August.

Aug. cluster includes: Together, pool, Children, adults, travel …

18 Discussion (3/3)

Accommodations #1

- ✓ “Red snapper” is popular from April to June
- ✓ Guests have a strong interest in food.

Accommodations #2

- ✓ Family trips are popular in August
- ✓ Guests can use swimming pools from June to September.

1. Abstract

2. The proposed method

3. Result of analysis

4. Discussion

5. Conclusion

19 Conclusion

1. We propose a method using hLDA to extract seasonal features from accommodation reviews.
2. We verified the effectiveness of the proposed method using TF-IDF and cosine similarity.
3. We found a seasonal difference in each cluster.
4. As the future work , we will differences in features in the region.