Developing data-driven QFD:
A systematic approach to employing text information using product manuals

박가문날비(석박사과정)[1] ,김신호(석사과정)[2], 금영정(부교수)[*]
[1 2] 서울과학기술대학교 데이터사이언스학과
* 서울과학기술대학교 산업공학과
[1 2] {nuripeace99, shk240} @seoultech.ac.kr
[*] yjgeum@seoultech.ac.kr ,

2023년 대한산업공학회 추계학술대회

Table of Contents

Introduction

Background

- QFD as a prominent method for designing new products or services, by measuring the relationships between Customer Requirements (CRs) and Functional Requirements (FRs)
- Many previous works on utilizing QFD for product design (Sireli et al, 2007; Ji et al 2014; Kirgizov, U.A & Kwak, C et al.,2021)
- A data-driven approach to the QFD as a promising research area (Chan, L. et al 2002; Xu et al, 2018; Ha & geum, 2022;)

Motivation

- Limitation of previous works on data-driven QFD

- Focusing only on CRs when employing data-driven approach (Nahm, 2013; Luo, et al, 2014)

(introducing customers purchase choice rules into the QFD based product planning process to simulate a customer's purchase behavior towards a product)

- Needs to suggest more effective ways on developing data-driven QFD, considering FR constructions and relationship measurements

Objective

- To develop a holistic data-driven QFD framework by employing a systematic approach to the data-driven FR constructions and measurements
- To utilize data mining and text mining techniques to extract proper FRs and quantitatively measure relationship between CRs and FRs

Literature review

In today, big data analytics drive businesses to integrate data analysis into their strategic management.

With the increasing accessibility of data sources, companies are finding it easier than ever to collect customer requirement from online platform.

Previous studies have analyzed mostly customer reviews and identified customer requirements from text.

1. Data-Driven Customer Requirement Analysis

Numerous studies use data-driven techniques to extract customer requirements from text data.

- Chiu et al. (2018) used text mining to predict customer preferences.
- Wang et al. (2019) combined different layers to elicit customer requirements.
- Xu & Dang. (2020) likewise developed an automated cause-and-effect diagram using a data-driven approach

Literature review

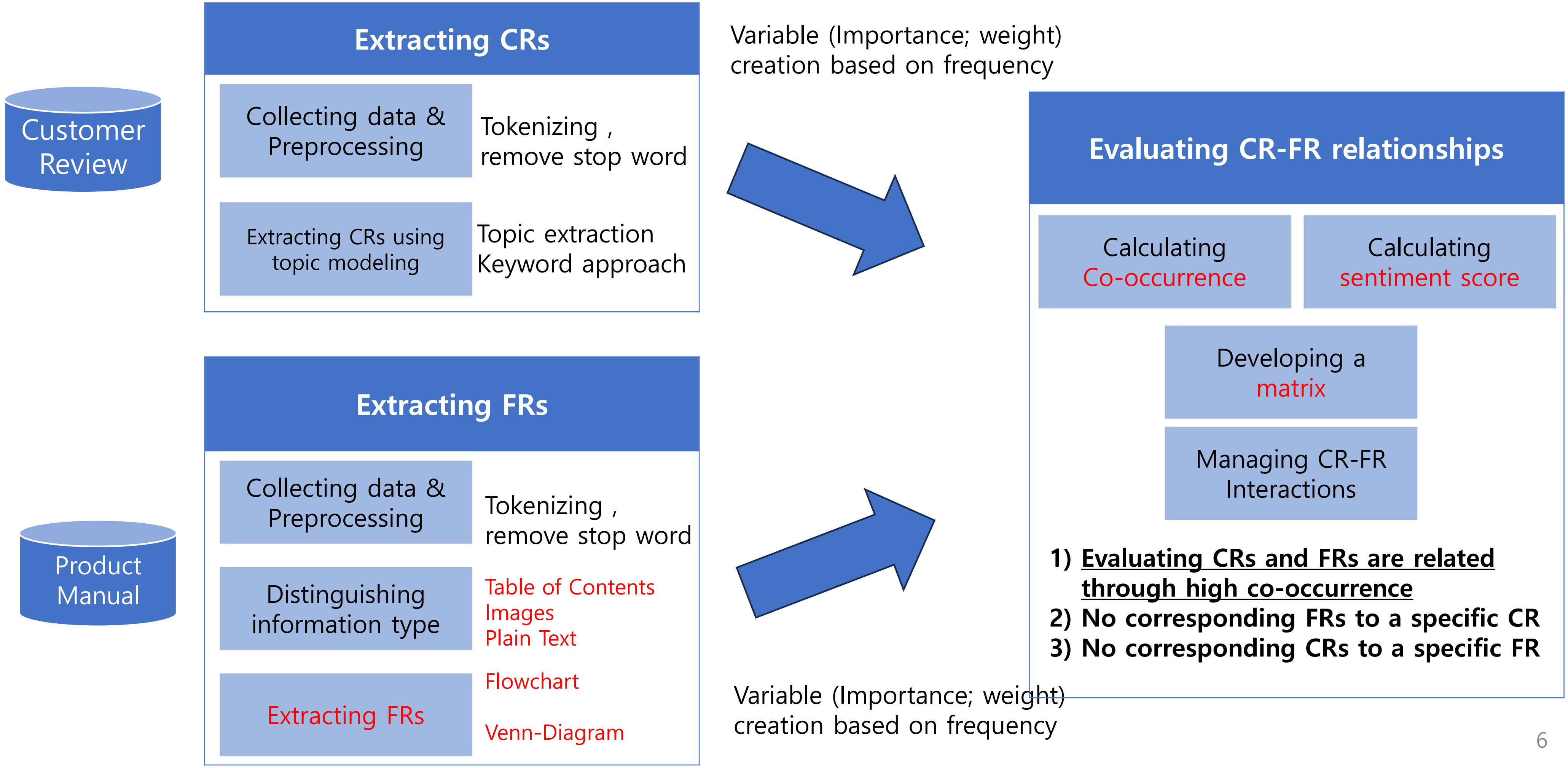
2. Quality Function Deployment (QFD) as Data-Driven

Recent advancements include the integration of data analysis with QFD.

- Jin et al. (2014) prioritized engineering characteristics based on customer reviews.
- Jin et al. (2015) used text mining for technology-driven QFD.
- Özdağoğlu et al. (2018) made an effort to integrate QFD and topic modeling.
- Huang et al. (2022) proposed an approach that combines LDA and Interval Grey Number QFD.
- Asadabadi et al. (2022) designed a method for automating the QFD process using text mining.

However, these studies mostly focused on customer requirements (CRs) and overlooked functional requirements (FRs). Given that QFD considers both CRs and FRs, a holistic framework addressing both is suggested.

Overall Framework



Framework for CRs

Collecting data

- Collecting review data from online

Preprocessing

Using collected data

- **Data Cleaning** : Removing unnecessary data, handling duplicates, and fixing format errors.
- **Outlier Treatment**: Managing unusual data.
- **Data Transformation**: Modifying data formats.
- **Scaling and Normalization**: Adjusting data scales.

Extracting CRs

1.1 Topic modeling

- Topic Selection

1.2 Keyword-based approach

- Select keyword with high frequency

2. Generating variable

- Importance : the more frequent, the more importance
- Sentiment- score : representation of text - keyword (positive, negative, neutral)

Framework for FRs

Collecting data

- Collecting manual from company website

Preprocessing

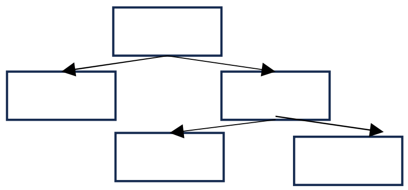
- Same CRs process

Distinguishing information type

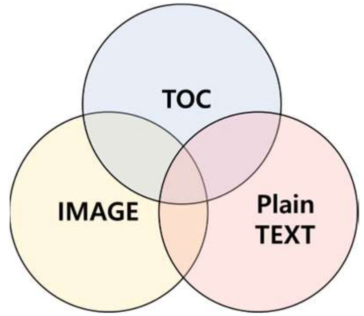
- TOC : manual structure
- Image : Product specifications
- Plain Text : product details

Extracting - FRs

1.1 Flowchart
- First, select TOC page
- Second, select specification images
- Third, select detailed product description (mostly text)
- Finally, remove unnecessary page.
- repeat Preprocessing



1.2 Venn – Diagram
- keywords occurs in each Information type



2. Generating variable
- Importance : the more frequent, the more importance
- Sentiment- score : representation of text - keyword (positive, negative, neutral)

Framework for Evaluating

Evaluating CR-FR relationships

1. Co-occurrence matrix

- **checking for missing rows or columns in each matrix**

➔ support to product design

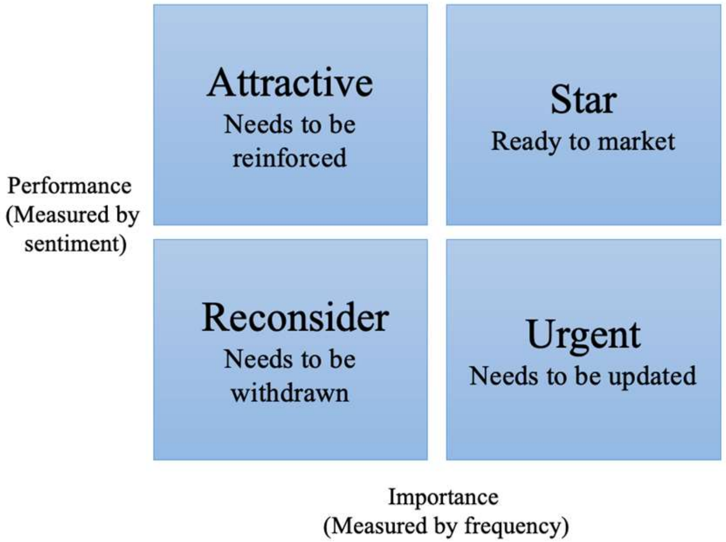
- co-occurrence matrices facilitate straightforward keyword comprehension and intensity.
- Visualization augments lucidity and insights.

2. IPA matrix

- capture sentiment alignment with CRs
- assist in decision-making by identifying which features to focus on

	assembled	guide	estimate	assurance	bougth
assembled	20.0	0.0	0.0	0.0	0.0
guide	0.0	8.0	0.0	0.0	0.0
estimate	0.0	0.0	2.0	0.0	0.0
assurance	0.0	0.0	0.0	6.0	0.0
bougth	0.0	0.0	0.0	0.0	2.0
...	...	...	...	...	...

Co-occurrence



IPA framework

Case illustration

1. Data collection

- CR : online platform (amazon.com)
- FR : company website

(bosch – washing machine)

CRs	FRs		
	TOC	Text	Image
period	detergent	detergent	panel
time	laundry	button	detergent
clothes	settings	laundry	setting
quality	panel	remove	clothes
delivery	safety	setting	control
money	blocked	hose	door
noise	removing	depending	display
value	drawer	press	
vibration	display	change	

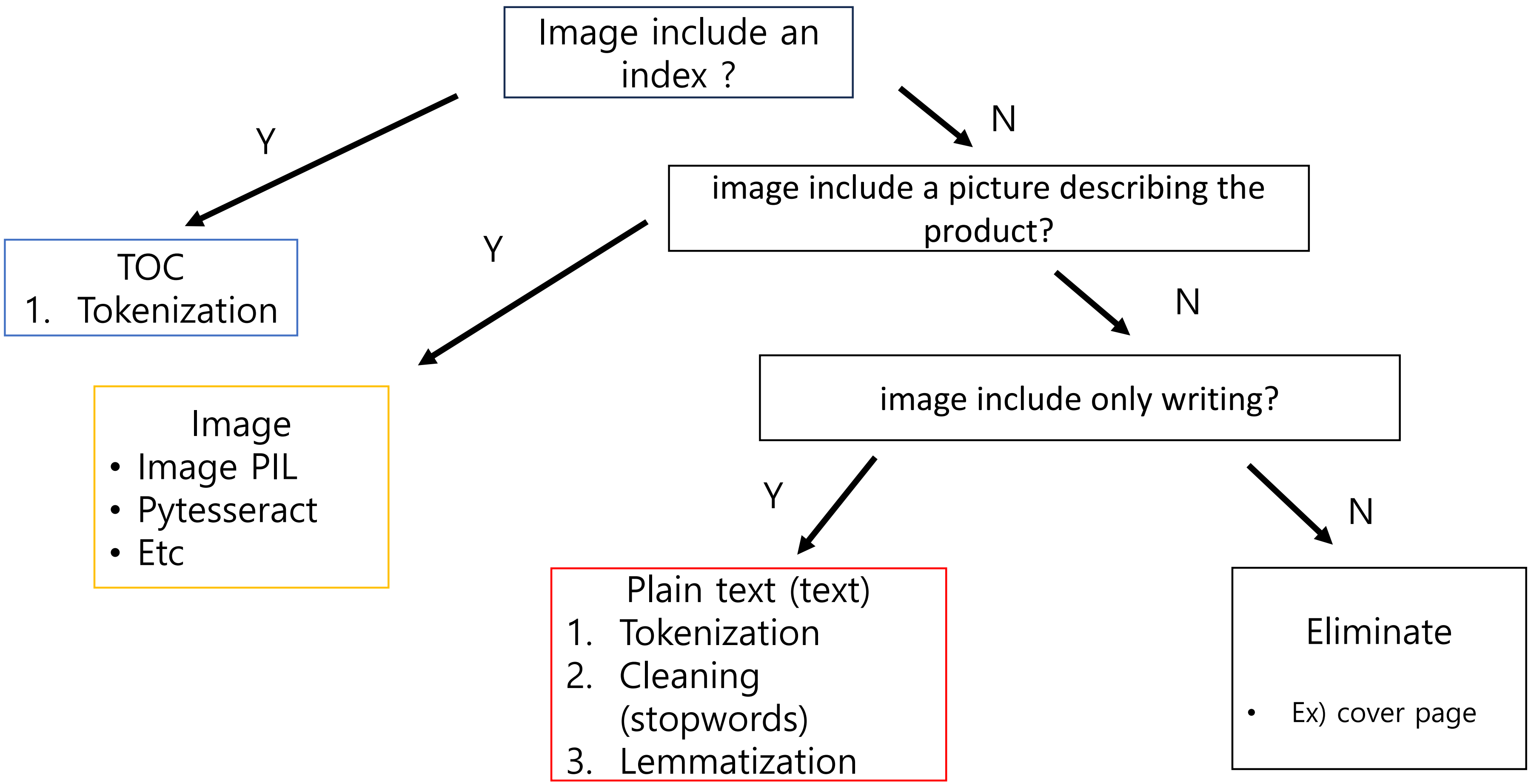
Table 1

2. Data Preprocessing (both)

- Data Cleaning : Removing unnecessary data, handling duplicates, and fixing error
- Outlier Treatment: Managing unusual data.
- Data Transformation: Modifying data formats.
- Feature Engineering: Creating new insights.
- Scaling and Normalization

Table 1 lists extracted keywords for CRs and FRs
(their high frequency occurrence)
➔ conduct steps 1,2 for CRs and steps1,2,3(distinguish) for FRs

Case illustration (FR – 3. distinguish)



Case illustration (4. generating variable)

- To assign weights, the frequencies of CRs and FRs are normalized to values between 0 and 1, and these values are utilized as weights.
- At this point, no threshold is applied to the frequencies of weighted CR and PF; in other words, weights are assigned to all keywords without imposing a frequency limit, such as freq > 2.
- The sentiment scores assigned to CRs and FRs are based on a threshold of 0.1.

CRs	Importance	sentiment_score
period	1	0
time	0.676271	0
clothes	0.383051	0
quality	0.359322	0
delivery	0.347458	0

FRs(TOC)	Importance	sentiment_score
detergent	1	0
laundry	0.71428571	0
settings	0.42857143	0
panel	0.42857143	0
safety	0.42857143	0.4215

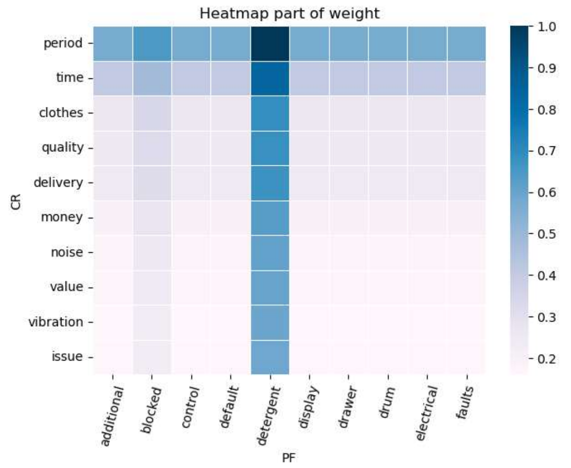
Positive

Case illustration (4. generating variable - Importance)

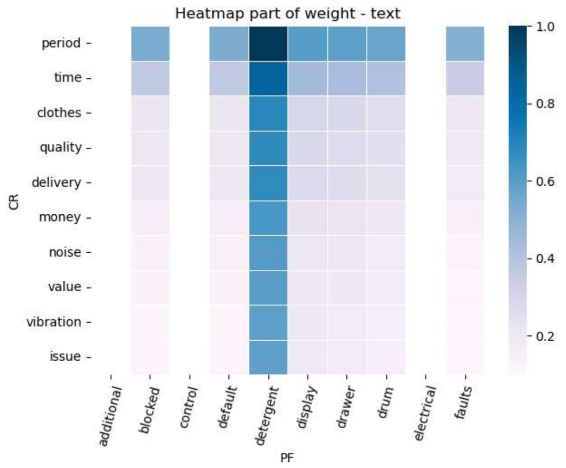
Importance

- Before construct the relationship matrix, we derived the weights (importance) for CRs and FRs
- Importance is a standardized variable generated based on frequency.
- The visualizations provided depict a subset of the importance values for the following relationships, presented in sequence: CR-TOC, CR-Text, and CR-Image.

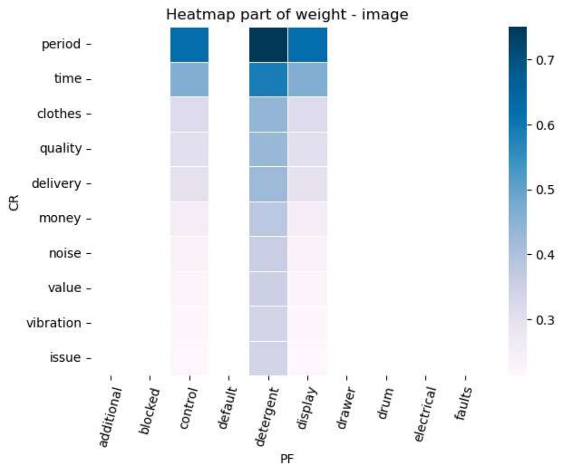
CR-TOC



CR-Text



CR-Image

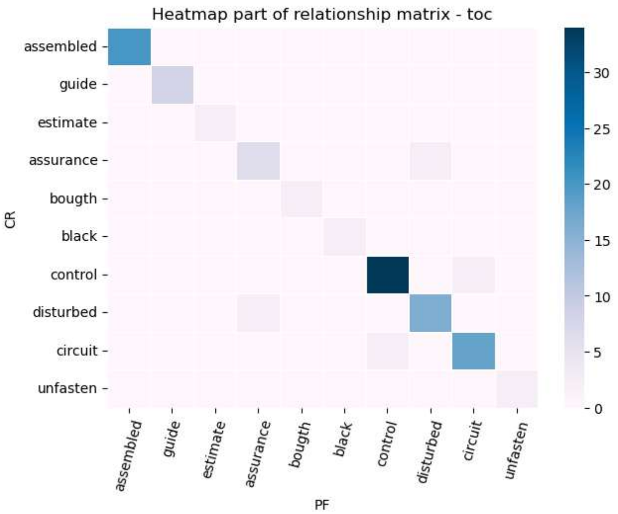


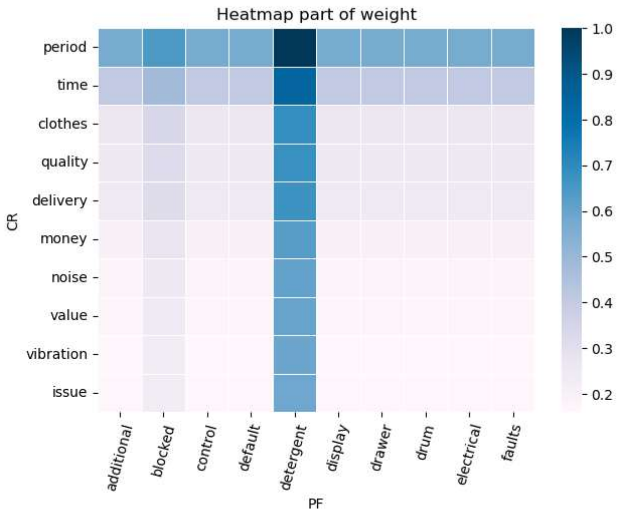
Case illustration (5. Evaluating – Relationship Matrix)

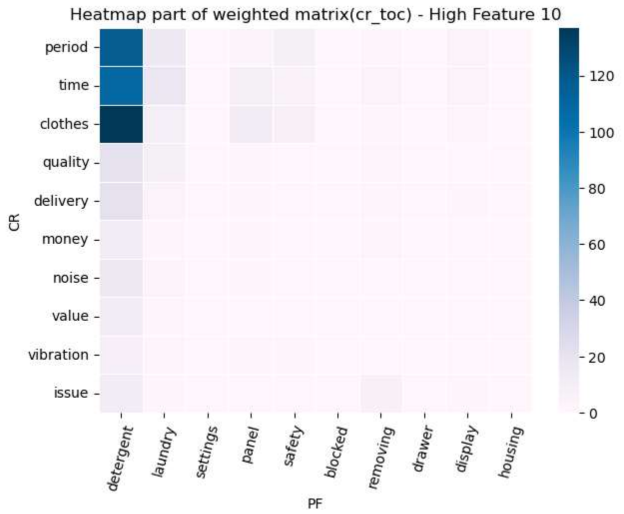
1) Co-occurrence matrix

- Defined Relationship Matrix (CR-TOC, CR-text, CR-Image)
- checking for missing rows or columns in each matrix
- support to product design

CR-TOC







Relationship matrix

Importance matrix

multiplied by the Relationship matrix to apply the importance

Case illustration (5. Evaluating – Relationship Matrix)

1) Co-occurrence matrix

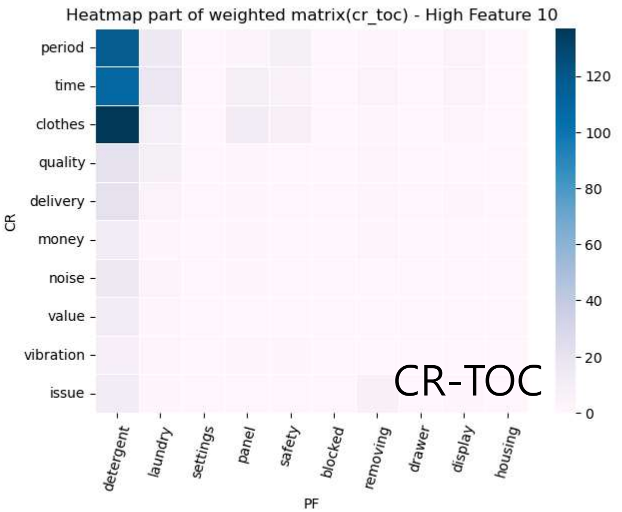
- The weighted average version of the Relationship Matrix enhances the relevance of CRs and FRs.

CR-TOC CR-Text CR-Image

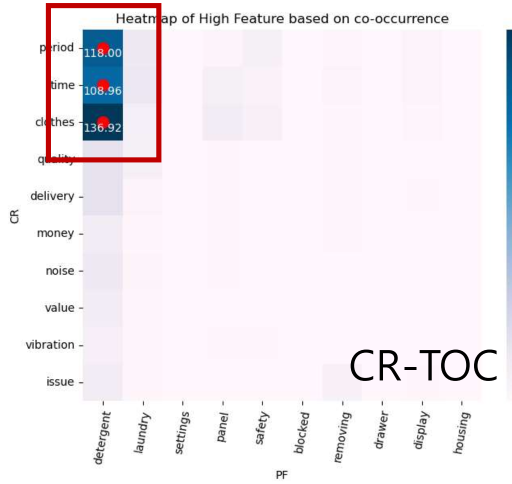
Case illustration (5. Evaluating – Relationship Matrix)

1) Co-occurrence matrix

- Additionally, detecting the substantial co-occurrence of CRs and FRs. (High value)
- The Relationship Matrix offers a more comprehensive analysis than a simple examination of keyword occurrences.
- Its applicability to the entire plot, rather than just subplots, contributes to a better comprehension of data distribution.

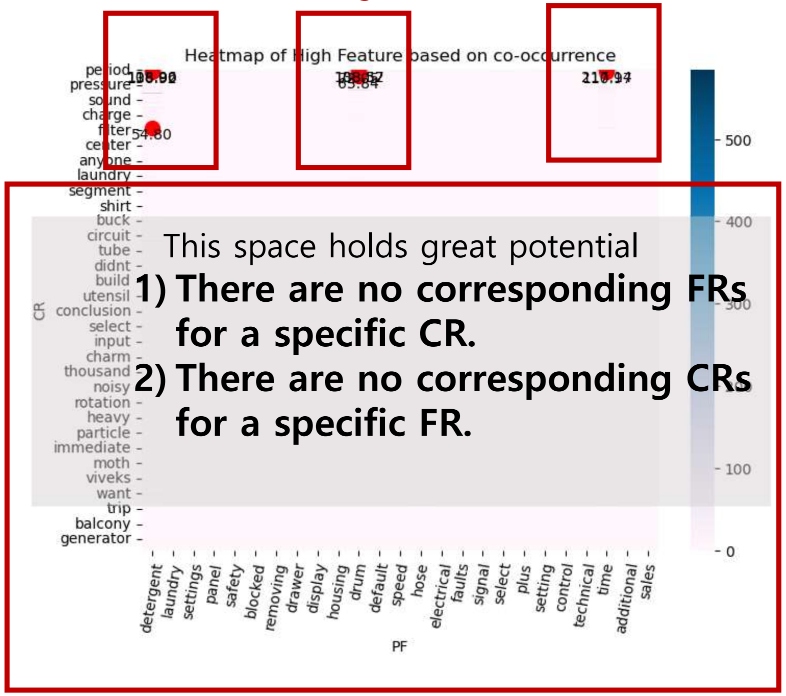


Relationship matrix Average ver.



Subset of high co-occurrence

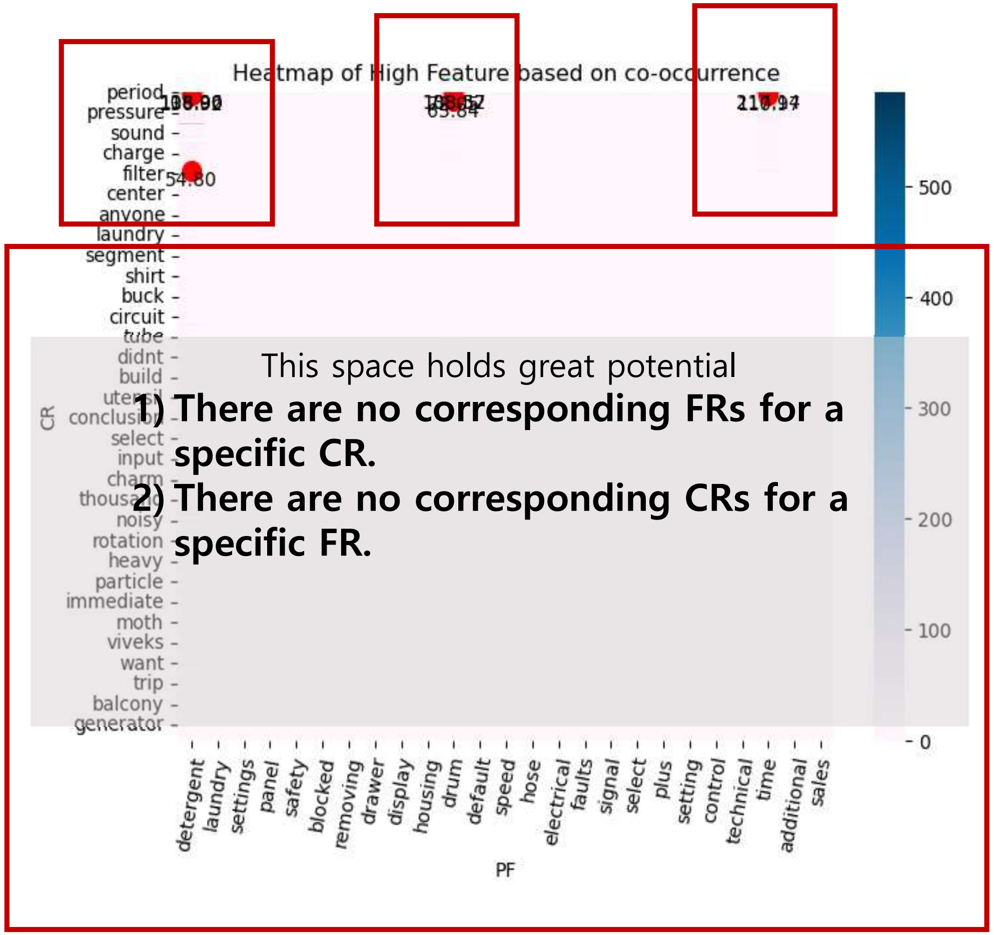
Subset of high co-occurrence



CR-TOC

Case illustration (5. Evaluating – Relationship Matrix)

1) Co-occurrence matrix



CR-TOC

1. In the figure, CRs (word: period, filter,..) exhibit a high co-occurrence of many points

2. Period-detergent, Period-drum, etc. can be considered as important.

➔ Components and consumables that are periodically used, such as, 'detergent','drum'

➔ Customers are concerned about consumables

3. The boxes correspond to the risks associated with 2) and 3).

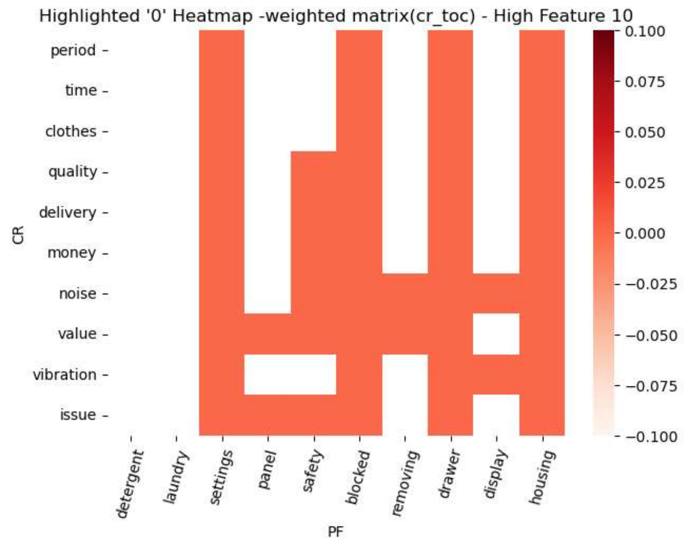
-> **Thorough examination of these boxes is warranted**

Case illustration (5. Evaluating – Relationship Matrix)

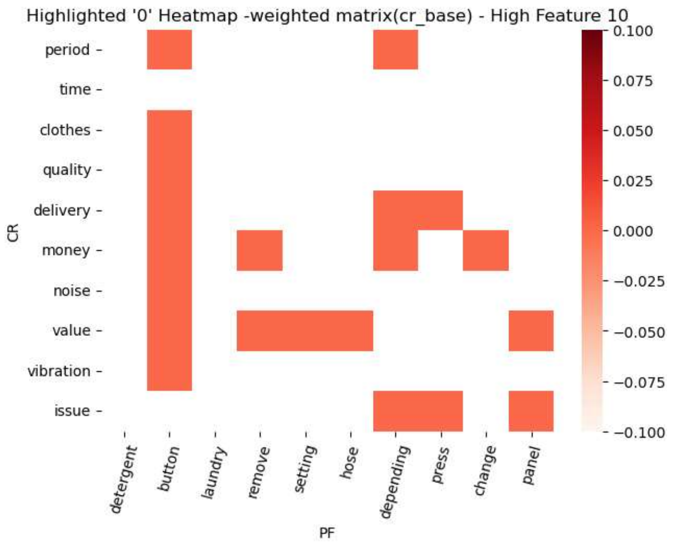
2) No corresponding FRs to a specific CR

- All matrices do not correspond to 2).

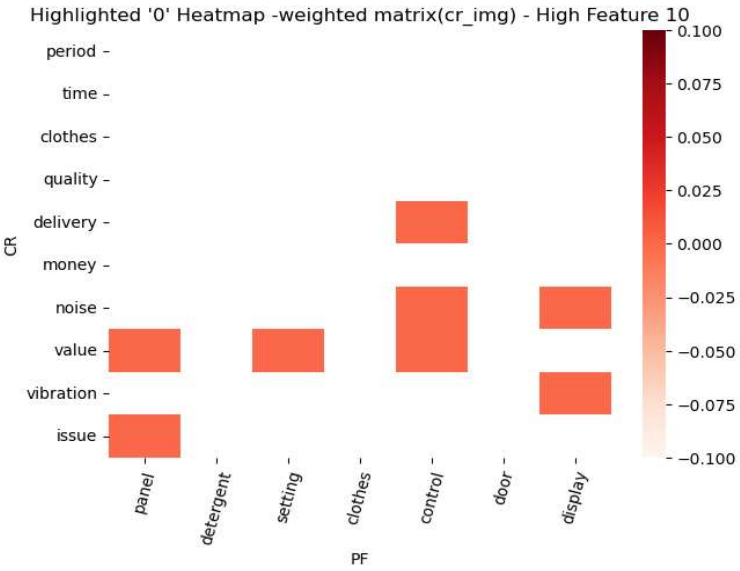
CR-TOC



CR-Text



CR-Image

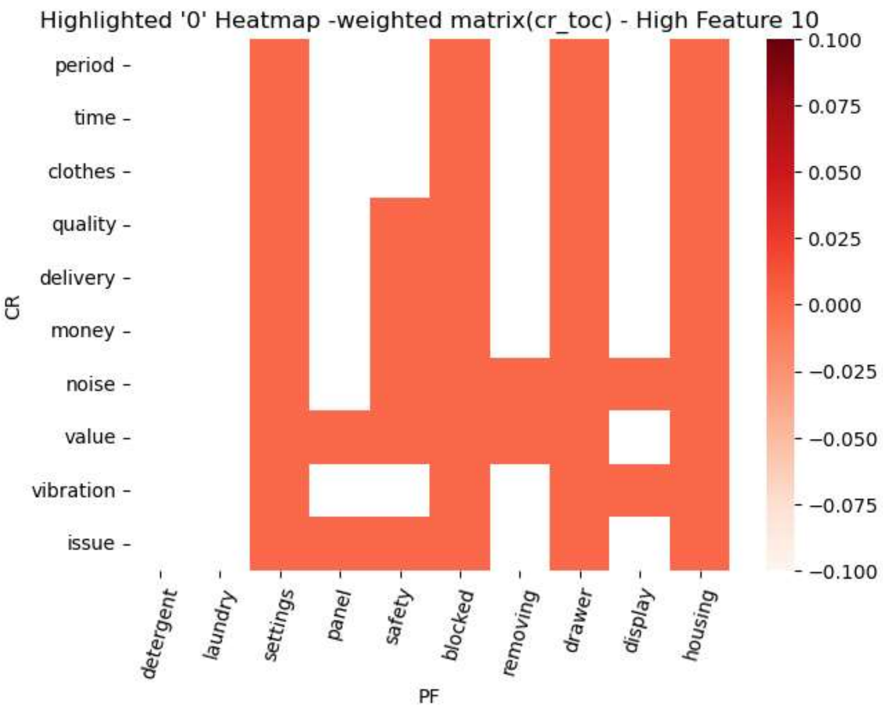


- In this case, there is unlikely to be a critical risk

Case illustration (5. Evaluating – Relationship Matrix)

3) No corresponding CRs to a specific FR

- CR-TOC matrices correspond to 3)



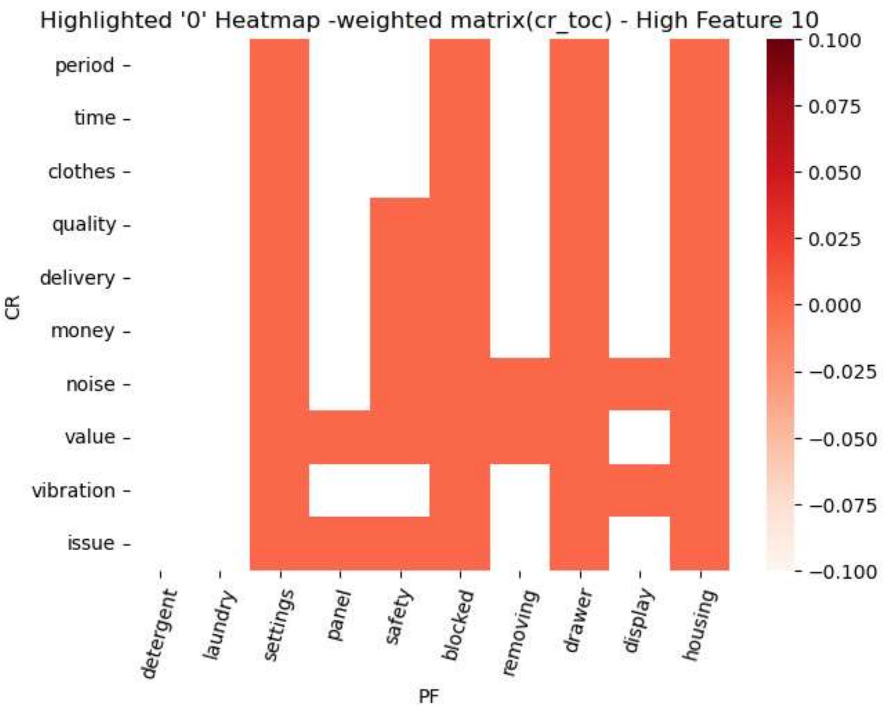
- **In the CR-TOC Relationship Matrix, FR : ‘setting’, ‘blocked’,’ drawer’ etc. have no corresponding FRs**

-> <u>Although it is a risk, there is no need to delete it</u>.
However, it is essential to consider whether it is truly necessary.

Case illustration (5. Evaluating – Relationship Matrix)

3) No corresponding CRs to a specific FR

- CR-TOC matrices correspond to 3)



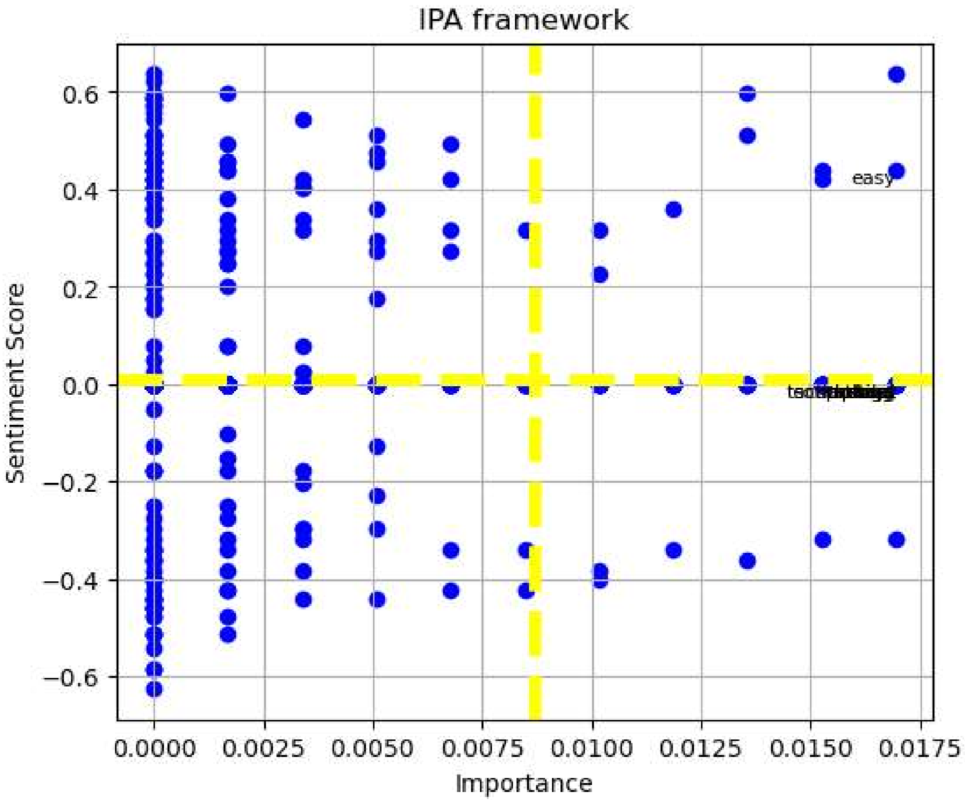
- **In the CR-TOC Relationship Matrix,**

FR : ‘setting’, ‘blocked’,’ drawer’ etc. have no corresponding FRs

-> <u>Although it is a risk, there is no need to delete it</u>.
However, it is essential to consider whether it is truly necessary.

Case illustration (5. Evaluating – IPA Matrix)

CRs - this value is greater than or equal to the lower 90% of the data and less than or equal to the upper 10% of the data

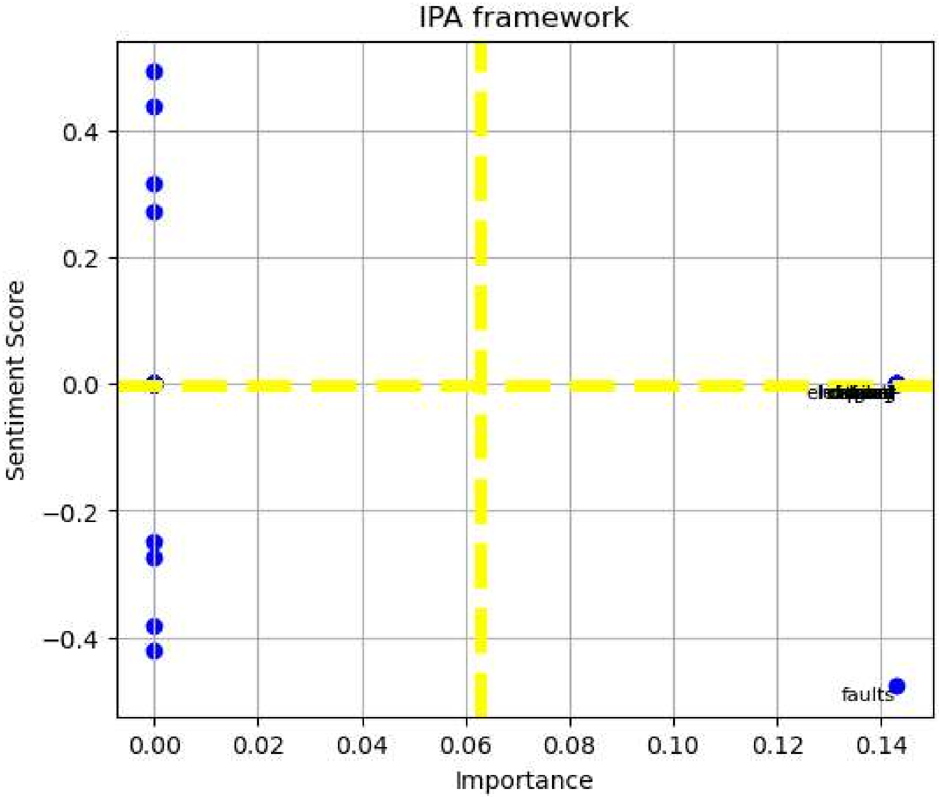


word	Importance	sentiment	score
something	0.016949	neutral	0.0000
show	0.016949	neutral	0.0000
prompt	0.016949	neutral	0.0000
duration	0.016949	neutral	0.0000
laundry	0.016949	neutral	0.0000
technology	0.016949	neutral	0.0000
panel	0.016949	neutral	0.0000
easy	0.016949	positive	0.4404
steel	0.016949	neutral	0.0000
change	0.016949	neutral	0.0000

'Easy' is Star FRs.
- performance and importance are at high levels, indicating well-defined FRs and positive customer perception

Case illustration (5. Evaluating –IPA Matrix)

TOC



word	Importance	sentiment	score
drawer	0.142857	neutral	0.0000
display	0.142857	neutral	0.0000
housing	0.142857	neutral	0.0000
drum	0.142857	neutral	0.0000
default	0.142857	neutral	0.0000
speed	0.142857	neutral	0.0000
hose	0.142857	neutral	0.0000
electrical	0.142857	neutral	0.0000
faults	0.142857	negative	-0.4767
signal	0.142857	neutral	0.0000

‘Faults’ is Urgent
- high importance but low levels of performance, as measured by customer feedback

Conclusion

Contribution

- The primary contribution of this study is the introduction and application of data-driven QFD.
- This approach is expected to enhance product design and quality management processes, aligning them more effectively with customer requirements.
- Additionally, it highlights the practicality of data integration throughout the entire QFD process.

Limitation

- Complexity of Analysis: The study suggests advanced methods for data-driven PF extraction, but lacks in-depth explanation for methodology
- Data Integration Hurdles: Integrating data from various sources is not thoroughly addressed, potentially limiting practical application.

Reference

- Kirgizov, U.A., & Kwak, C. (2021). Quantification and integration of Kano’s model into QFD for customer-focused product design. Quality Technology & Quantitative Management, 19, 95 - 112.
- Sireli, Y., Kauffmann, P.J., & Ozan, E. (2007). Integration of Kano's Model Into QFD for Multiple Product Design. *IEEE Transactions on Engineering Management, 54*, 380-390.
- Ji, P., Jin, J., Wang, T., & Chen, Y. (2014). Quantification and integration of Kano’s model into QFD for optimising product design. *International Journal of Production Research, 52*, 6335 - 6348.
- Xu, Y., & Yan, H. (2018). Identifying Core Technongies Based on a Two-Stage QFD Model. *2018 10th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), 01*, 356-360.
- Ha, S., & Geum, Y. (2022). Identifying new innovative services using M&A data: An integrated approach of data-driven morphological analysis. *Technological Forecasting and Social Change*.
- Chan, L. K., & Wu, M. L. (2002). Quality function deployment: A literature review. European journal of operational research, 143(3), 463-497.
- Wei, W., & Zixin, W. (2023). An Improved QFD Method for Rapid Response to Customer Requirements in Product Optimization Design. Procedia CIRP.
- Luo, X.G. & Kwong, C.K. & Tang, Jiafu & Sun, F.. (2014). QFD-Based Product Planning With Consumer Choice Analysis. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 45. 1-1.
- Nahm, YE. (2013). A novel approach to prioritize customer requirements in QFD based on customer satisfaction function for customer-oriented product design. J Mech Sci Technol 27, 3765–3777
- Chiu, M. C., & Lin, K. Z. (2018). Utilizing text mining and Kansei Engineering to support data-driven design automation at conceptual design stage. Advanced Engineering Informatics, 38, 826-839.
- Jin, J., Ji, P., & Liu, Y. (2014). Prioritising engineering characteristics based on customer online reviews for quality function deployment. Journal of Engineering Design, 25(7-9), 303-324.

Reference

- Chiu, M. C., & Lin, K. Z. (2018). Utilizing text mining and Kansei Engineering to support data-driven design automation at conceptual design stage. Advanced Engineering Informatics, 38, 826-839
- Wang, Z., Chen, C. H., Zheng, P., Li, X., & Khoo, L. P. (2019). A novel data-driven graph-based requirement elicitation framework in the smart product-service system context. Advanced engineering informatics, 42, 100983.
- Xu, Z., & Dang, Y. (2020). Automated digital cause-and-effect diagrams to assist causal analysis in problem-solving: a data-driven approach. International Journal of Production Research, 58(17), 5359-5379.
- Özdağoğlu, G., Kapucugil-Ikiz, A., & Celik, A. F. (2018). Topic modelling-based decision framework for analysing digital voice of the customer. Total Quality Management & Business Excellence, 29(13-14), 1545-1562
- Huang, S., Zhang, J., Yang, C., Gu, Q., Li, M., & Wang, W. (2022). The interval grey QFD method for new product development: Integrate with LDA topic model to analyze online reviews. Engineering Applications of Artificial Intelligence, 114, 105213.
- Asadabadi, M. R., Saberi, M., Sadghiani, N. S., Zwikael, O., & Chang, E. (2022). Enhancing the analysis of online product reviews to support product improvement: integrating text mining with quality function deployment. Journal of Enterprise Information Management, (ahead-of-print)