Service Intelligence Week 5.

[Service Customization and Customer Segmentation]

Chiehyeon Lim

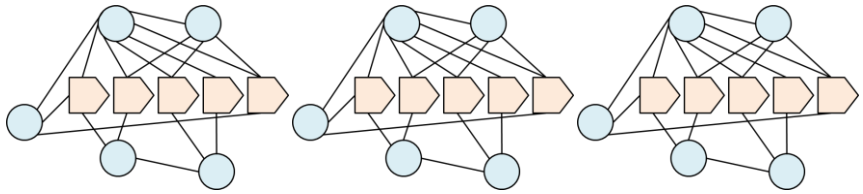
2022. 9. 26

Further Discussion on the Topics of Previous Weeks

Shouldn't We Consider the Service Quality Dynamics in Time?

consumer behaviors (user behaviors)

service delivery and experience



quality

relationship

productivity

technology and self-service

value cocreation

innovation

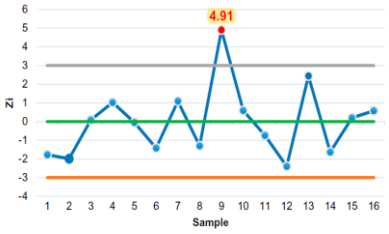
design and development

knowledge and employee management

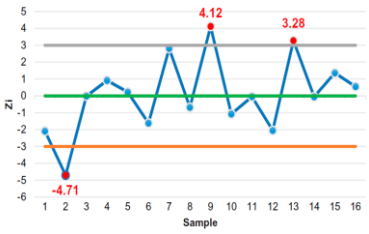
cost and pricing

service operations and supply chain management

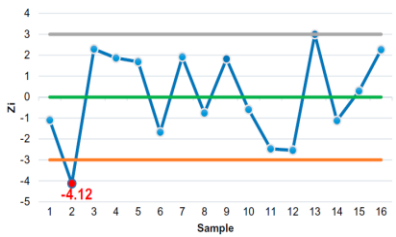
Monitoring the Quality Evaluation Results



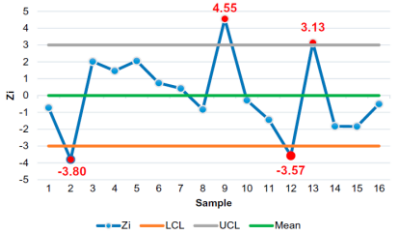
Tangibles



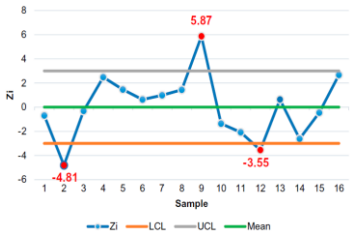
Reliability



Responsiveness



Assurance



Empathy

Reference: Altuntas, S., Dereli, T., & Kaya, İ. (2020). Monitoring patient dissatisfaction: a methodology based on SERVQUAL scale and statistical process control charts. Total Quality Management & Business Excellence, 31(9-10), 978-1008.

Customer Complaints Monitoring with Review Mining

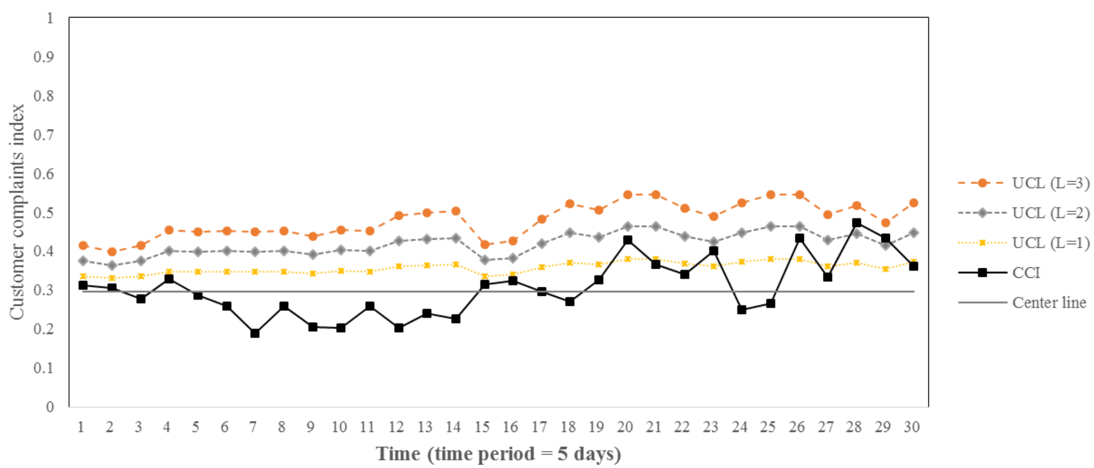
On the Time Dynamics of CustomerReviews–ServiceFeatures

	Feature 1	Feature 2	Feature 3	…	Feature m-1	Feature m	Rating
Review 1	…	…	…	…	…	…	5
Review 2	…	…	…	…	…	…	5
Review 3	…	…	…	…	…	…	4
…	…	…	…	…	…	…	3
…	…	…	…	…	…	…	4
…	…	…	…	…	…	…	1
…	…	…	…	…	…	…	4
…	…	…	…	…	…	…	5
…	…	…	…	…	…	…	4
…	…	…	…	…	…	…	3
Review n-1	…	…	…	…	…	…	2
Review n	…	…	…	…	…	…	4

	Feature 1	Feature 2	Feature 3	…	Feature m-1	Feature m	Rating
Review 1	…	…	…	…	…	…	5
Review 2	…	…	…	…	…	…	5
Review 3	…	…	…	…	…	…	4
…	…	…	…	…	…	…	3
…	…	…	…	…	…	…	4
…	…	…	…	…	…	…	1
…	…	…	…	…	…	…	4
…	…	…	…	…	…	…	5
…	…	…	…	…	…	…	4
…	…	…	…	…	…	…	3
Review n-1	…	…	…	…	…	…	2
Review n	…	…	…	…	…	…	4

	Feature 1	Feature 2	Feature 3	…	Feature m-1	Feature m	Rating
Review 1	…	…	…	…	…	…	5
Review 2	…	…	…	…	…	…	5
Review 3	…	…	…	…	…	…	4
…	…	…	…	…	…	…	3
…	…	…	…	…	…	…	4
…	…	…	…	…	…	…	1
…	…	…	…	…	…	…	4
…	…	…	…	…	…	…	5
…	…	…	…	…	…	…	4
…	…	…	…	…	…	…	3
Review n-1	…	…	…	…	…	…	2
Review n	…	…	…	…	…	…	4

Time



Case Study on the Mobile Game Service Quality Monitoring

- Case study of mobile game service
  - Representing 58% of all downloads in the mobile markets (as of 2020)
  - One of the fastest-changing areas with fierce competition
  - Need to manage service quality and update their game services by adding new attributes or fixing bugs

- Data summary
  - Database: Apple app store
  - Application: Angry bird 2
  - Data variables: Date, reviewer ID, rating, title, review content, version
    - Pre-processing the review contents via NLTK in python
  - Period: 2017-07-31 ~ 2017-12-25
  - Country: United States
  - Total number of reviews: 2,010

Case Study on the Mobile Game Service Quality Monitoring

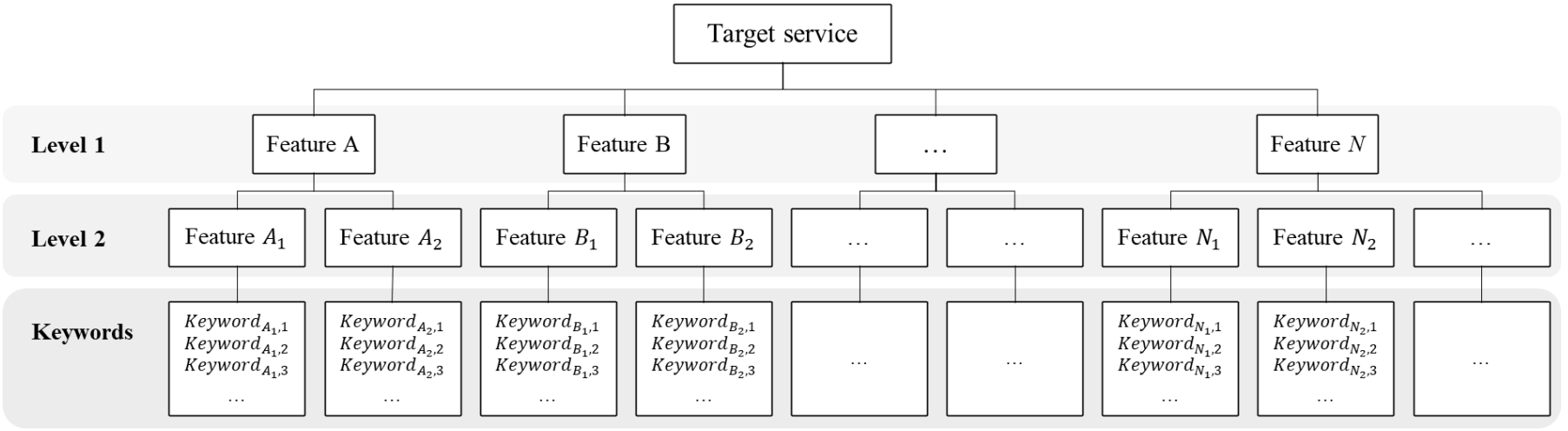
- Step 1: Data collection and pre-processing
  - Collecting the review date and review content
  - Pre-processing the review content via tokenization, stop-words removal, POS tagging, and lemmatization
  - Part of the customer review database

No.	Date	Reviewer ID	Rate	Title	Review content (raw data)	Review content (after pre-processing)	Version
1	2017-07-31	CPJ***	3	A good pastime	It's Angry Birds, but on steroids. The graphics and sound effect make it fun to pass the time playing.	[(angry, JJ), (bird, NN), (steroid, JJ), (graphic, JJ), (sound, NN), effect, NN), (make, VBP), (fun, NN), (pas, NN), (time, NN), (play, NN)]	2.14.0
…	…	…	…	…	…	…	…
2010	2017-12-25	Frogit to***	5	Frogit	It's addictive I can't stop playing I have so much fun great game!	[(addictive, JJ), (cant, NN), (stop, VB), (play, NN), (much, JJ), (fun, NN), (great, JJ), (game, NN)]	2.17.2

** Part of speech tags and descriptions: CD=cardinal number, DT=determiner, JJ=adjective, NN=noun (singular), PRP=personal pronoun, RB=adverb, RBR=adverb (comparative), VB=verb (base form), VBD=verb (past tense), VBP=verb (non-3rd person singular present), VBZ=verb (3rd person singular present), WRB=wh-adverb

Case Study on the Mobile Game Service Quality Monitoring

- Step 2: Construction of a service feature hierarchy with keyword dictionary
  - Service feature hierarchy



Case Study on the Mobile Game Service Quality Monitoring

- Step 2: Construction of a service feature hierarchy with keyword dictionary
  - Service feature hierarchy employed in this study

Service feature (level 1)	Description	Service feature (level 2)	Description	Ciurumelea et al. (2017)	Mcllroy et al. (2016)	Khalid et al. (2015)	Maalej and Nabil (2015)	Fu et al. (2013)
Compatibility	Issues related to version of the OS or the specific phone device	Version	Issues related to update or mobile app version	✓	✓	✓		
		Hardware	Issues related to a specific mobile phone device of OS	✓	✓	✓		✓
Usage	Reports the things that are uncomfortable to use and things that user want to improve	Attribute requests	Issues related to additional attribute(s) or modification	✓	✓	✓	✓	
		Bug reporting	Issues related to unexpected bug		✓	✓	✓	
		Difficulty of game	Issues related to difficulty of mobile game	Added by the authors				
		Spam	Issues related to advertisement					✓
Resources	Mentions the memory or battery usage	Battery	Issues related to battery usage	✓	✓	✓		
		Memory	Issues related to memory usage	✓	✓	✓		
Pricing	Refers the licensing model, price of the app, or in-app purchase issues	Price	Issues related to the licensing model, price of the app, or in-app purchase	✓	✓	✓		✓
Protection	States the security issues or user privacy	Security	Issues related to security or lack of it	✓				
		Privacy	Issues related to permissions and privacy	✓	✓	✓		

Case Study on the Mobile Game Service Quality Monitoring

- Step 3: Identification of customer complaints
  - Part of the results of sentiment analysis

Review contents	Positive	Neutral	Negative	Compound	Classification
It's Angry Birds, but on steroids. The graphics and sound effect make it fun to pass the time playing.	0.268	0.645	0.087	**0.671**	Positive
Amazing and fun game I love it	0.79	0.21	0.0	**0.906**	Positive
I love playing the game it is so much fun and it is challenging it make your mind work thank you so much	0.446	0.554	0.	**0.913**	Positive
It's cute and the upgrades are nice	0.537	0.463	0.0	**0.7**	Positive
The graphics on this latest version are amazing! Even on my iPad 2.	0.254	0.746	0.0	**0.624**	Positive
…	…	…	…	**…**	…
App keeps freezing when you are playing, do you lose your archived levels. Very frustrating!	0.088	0.54	0.372	**-0.699**	Negative
Too hard to reach boss - run out of birds before getting to final room!	0.074	0.811	0.114	**-0.151**	Negative
Game has started freezing in the middle of the game. Is there a bug?	0.0	0.896	0.104	**-0.103**	Negative

  - Out of 2,010 customer reviews, a total of 640 customer reviews identified as customer complaints

Case Study on the Mobile Game Service Quality Monitoring

- Step 4: Development of a customer complaints chart via SPC
  - Three issues for developing the customer complaints chart using SPC
    - In terms of the use of statistic
      - Interpreting each customer complaint as non-conformity
      - Defining customer complaints index (CCI) by modifying the index used for measuring service performance (Chen and Yang, 2000; Rasouli and Zarei, 2016; Yang and Chen, 2000)

$$CCI_{i,t} = \frac{The\ number\ of\ negative\ customer\ reviews_{i,t}}{The\ total\ number\ of\ customer\ reviews_{i,t}} = \frac{N(NCR_{i,t})}{N(CR_{i,t})}$$

where $NCR_{i,t}$ represents the number of negative customer reviews for the *i* th service feature at the time period *t*

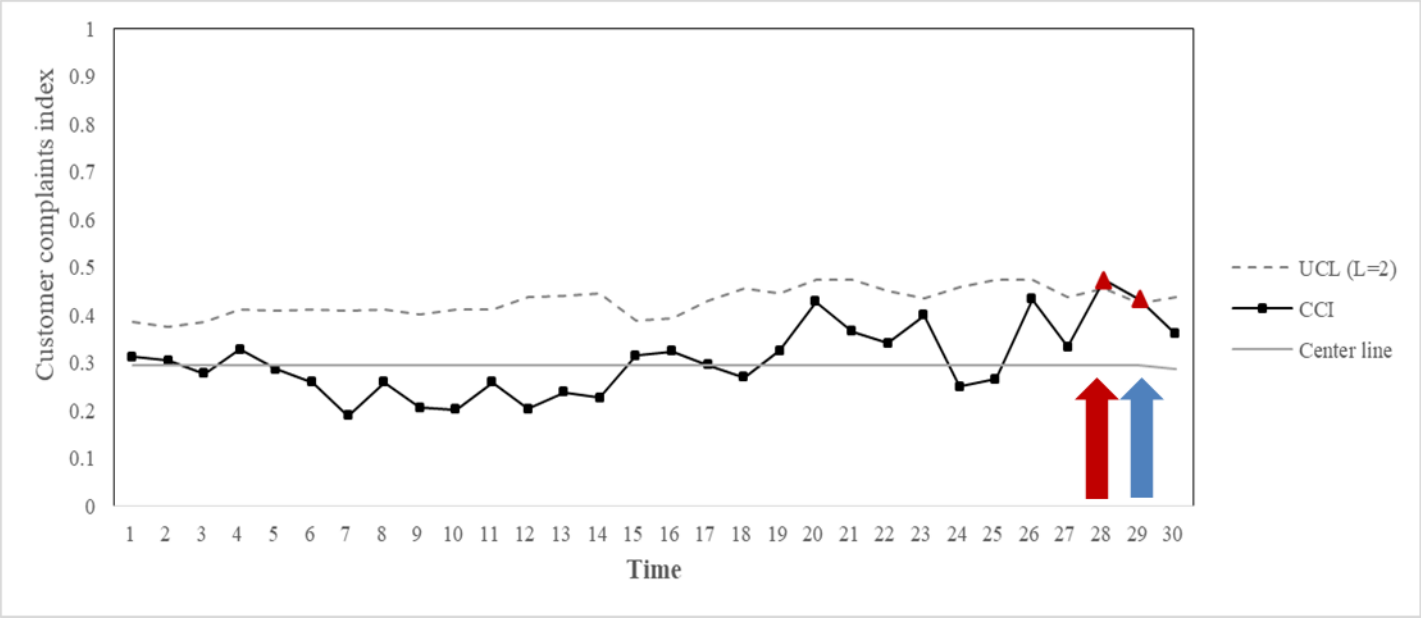
      - Defining the center line(CL) and upper control limit (UCL) as:

$$CL_i = \frac{\sum_{t=1}^{T} NCR_{i,t}}{\sum_{t=1}^{T} CR_{i,t}} \qquad UCL_{i,t} = CL_i + L\sqrt{\frac{CL_i(1-CL_i)}{CR_{i,t}}}$$

where *T* and *L* denote the number of time periods and sensitivity parameter, respectively

Case Study on the Mobile Game Service Quality Monitoring

- Step 4: Development of a customer complaints chart via SPC
  - Customer complaints chart for the overall status (time period=5days, L=2)

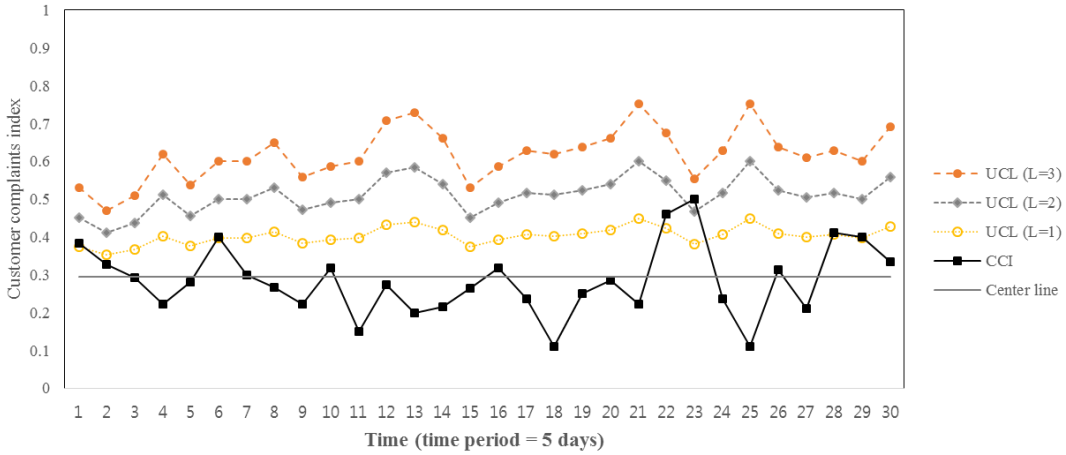


- Scheduled update to add a new attribute on December 11th, which caused the compatibility issue
- Out-of-control signals on December 13th-17th
- Minor update to solve the compatibility problem on December 20th

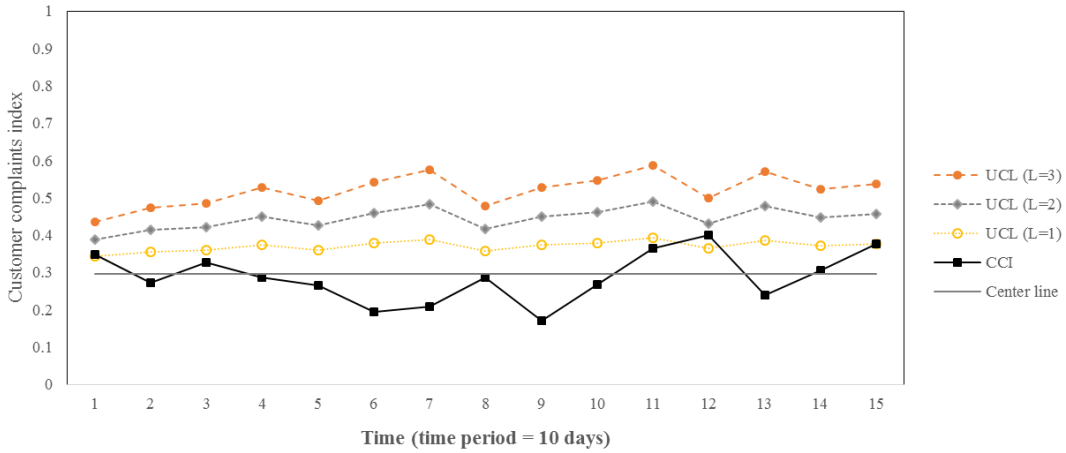
Case Study on the Mobile Game Service Quality Monitoring

- Customer complaints chart for the attribute request feature with different value of control parameters

(a)



(b)



Service feature (level 1)	Description	Service feature (level 2)
Compatibility	Issues related to version of the OS or the specific phone device	Version Hardware
Usage	**Reports the things that are uncomfortable to use and things that user want to improve**	**Attribute requests** Bug reporting Difficulty of game Spam
Resources	Mentions the memory or battery usage	Battery Memory
Pricing	Refers the licensing model, price of the app, or in-app purchase issues	Price
Protection	States the security issues or user privacy	Security Privacy

Further Readings Recommended

- Shin, J., Joung, J., and Lim, C., "Online Review Mining Meets Interpretable Machine Learning for Customer-oriented Service Quality Management," 2022. Article 6
- Kim, J. and Lim, C., "Customer Complaints Monitoring with Customer Review Data Analytics: An Integrated Method of Sentiment And Statistical Process Control Analyses," Advanced Engineering Informatics, Vol. 49, 101304, 2021. Article 7

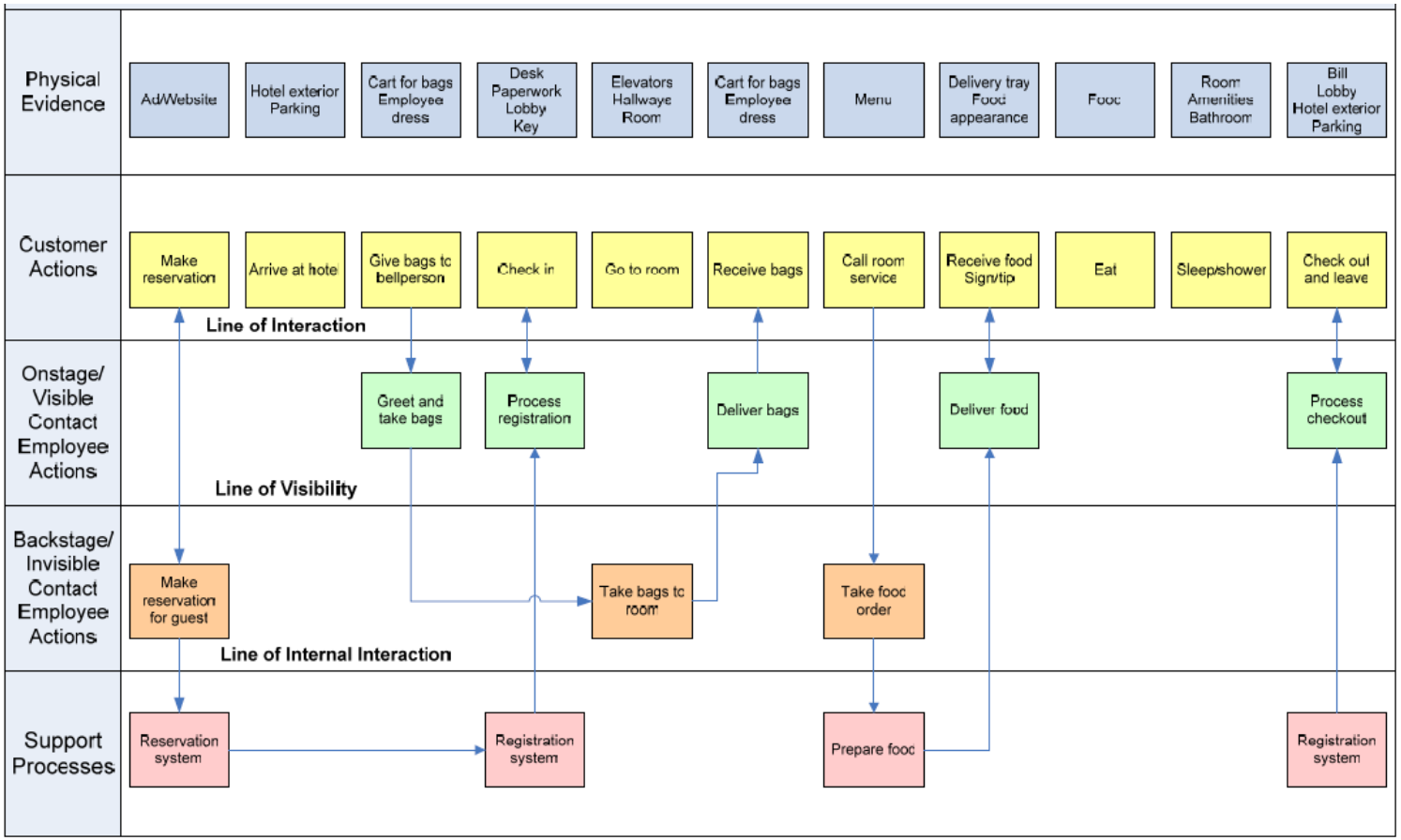
Assignment 4 (by 10.7 11:59 pm)

- By yourself, (1) complete the construction of the review-feature dataset on service quality of hotels in Singapore based on the practice demonstrated by the TA. Then, (2) using the review-feature matrix you constructed, develop a service quality prediction model for the hotels in Singapore (i.e., predict the customer's quality evaluation with review data). Do it all by yourself, and describe the analysis process and outcome in detail. Interpret the outcome (e.g., name the service features you identified, interpret the coefficient/importance values of service features to the quality ratings).
- (3) What other interesting machines can be developed using the review-feature matrix dataset you constructed? Describe your ideas in detail (e.g., describe the learning objective and process). Try to think your own creative, unique ideas! You have completed the basic review mining activities (tasks 1 and 2) as well as your own idea generation (task 3). Then, (4) describe how you can use your machine(s) to automate the monitoring, evaluation, and improvement of hotel service quality? Imagine you are working for a real hotel.
- Using a similar approach that you have practiced so far, (5) what other services can be improved using review mining machines or another intelligent machine that learns other types of raw data traces of service quality (e.g., customer behavior data)? Assume that you actually manage the quality of service in question. (6) How would you conduct this job in your own creative, unique way? What kinds of data and methods are you going to collect, analyze, and learn? Describe your service intelligence development plan in detail. If possible, visualize your plan clearly (e.g., draw an image, construct a mathematical model).
- Upload your code and a several paragraph essay on the tasks (1)~(6) in the Blackboard.

Blueprinting with Review Mining for Service Improvement

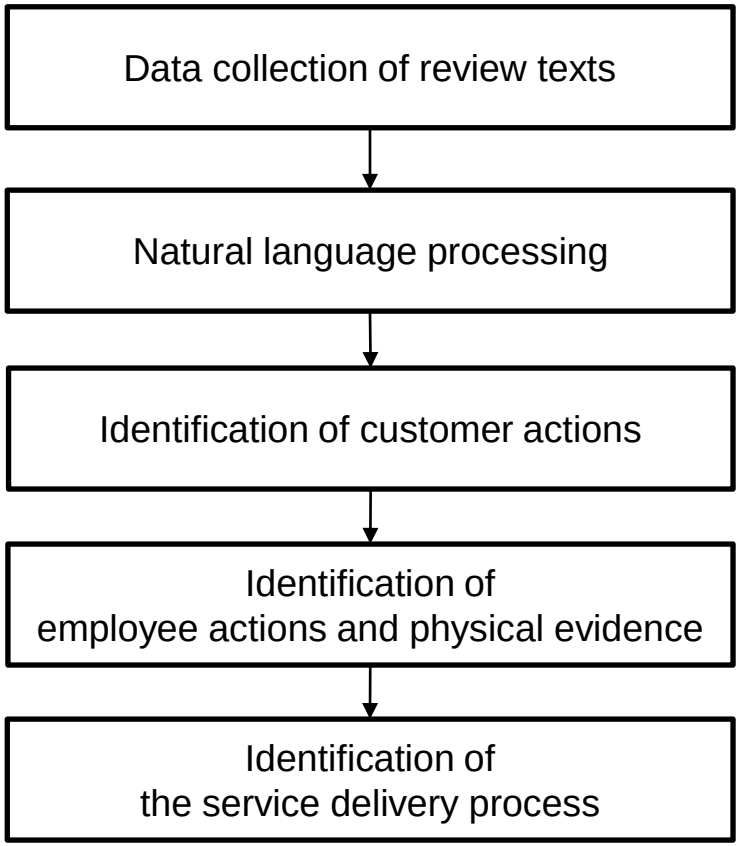
Service Blueprinting

- A customer-focused approach for service design and service improvement



A Blueprint for Overnight Hotel Stay Service

A Data-driven Method



Case Study on the Hotel Services in Seoul

Physical evidence

hotel

hotel room

key reception

staff

taxi
bus
train
subway
airport

Customer actions

find hotel → book hotel → check in → find staff → take vehicles

Employee actions

bring luggage

give room

help guest

provide service

take luggage

give information

call taxi

Core steps

Auxiliary steps

Restaurant buffet

eat breakfast

provide breakfast

front bill

check out

help bus

pool
lounge
sauna
gym

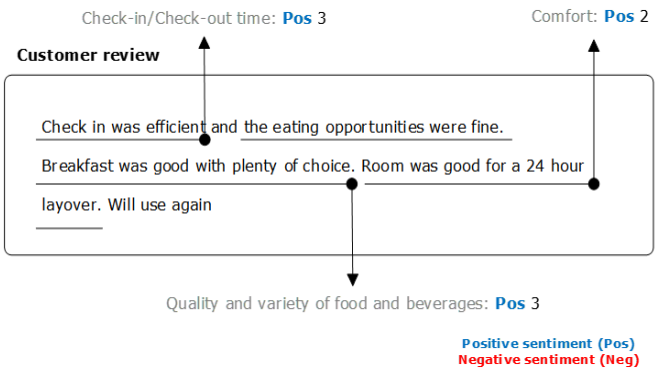
use facilities

With This Approach, We Can Measure/Improve the Service in Blueprint

Identify process-step-feature values in the service delivery process

check in	use facilities	get room	find staff	eat breakfast	take vehicles	check out
checkin check-in	pool lounge sauna gym	room	staff	breakfast	taxi bus train subway airport location transportation	checkout check-out

Keyword-based sentiment analysis



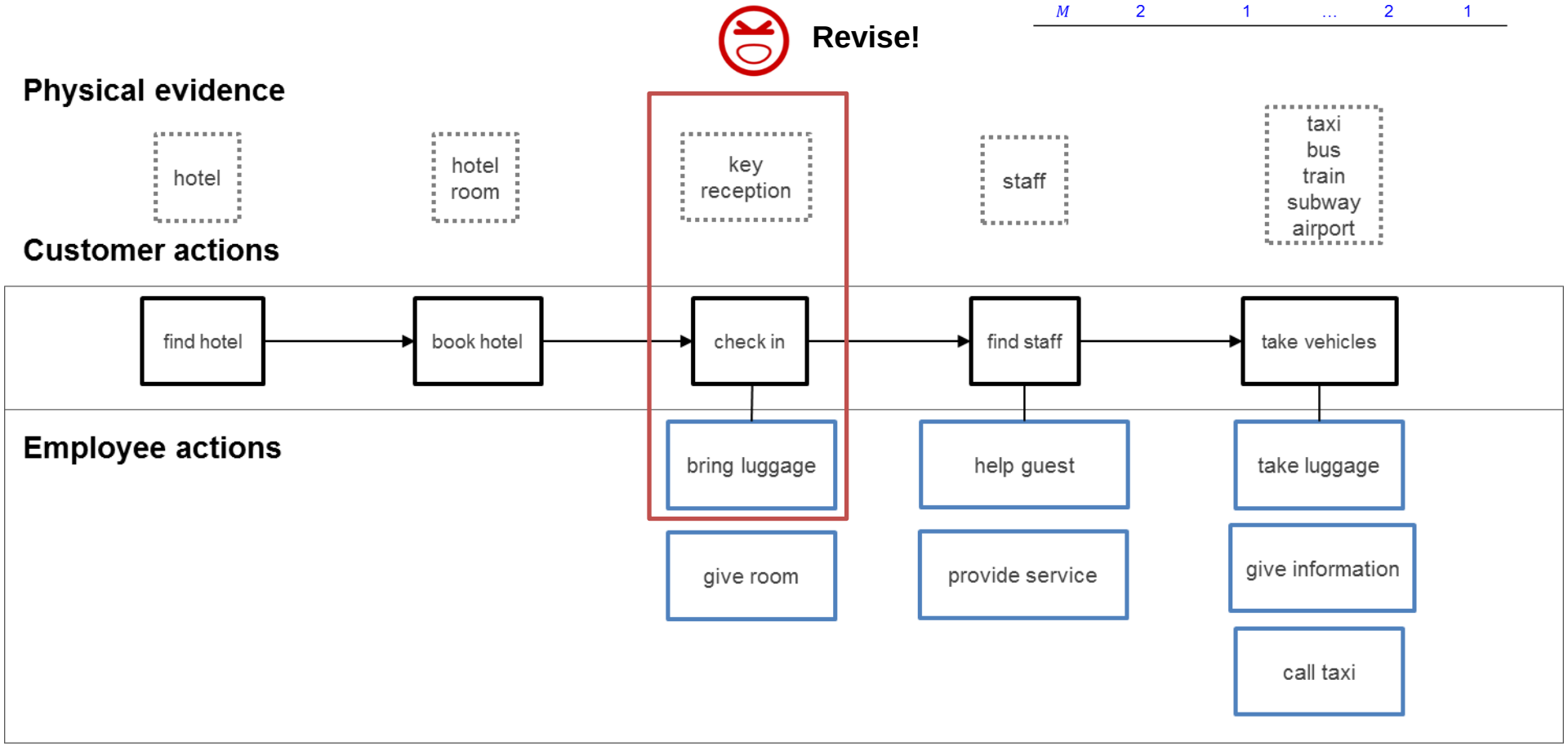
Process-step-feature importance calculation

	Check_in	*Use facilities*	…	*Check_out*	*Overall rating*
1	5	4	…	5	5
2	4	4	…	5	4
3	5	4	…	5	4
…	…	…	…	…	…
M	2	1	…	2	1

With This Approach, We Can Measure/Improve the Service in Blueprint

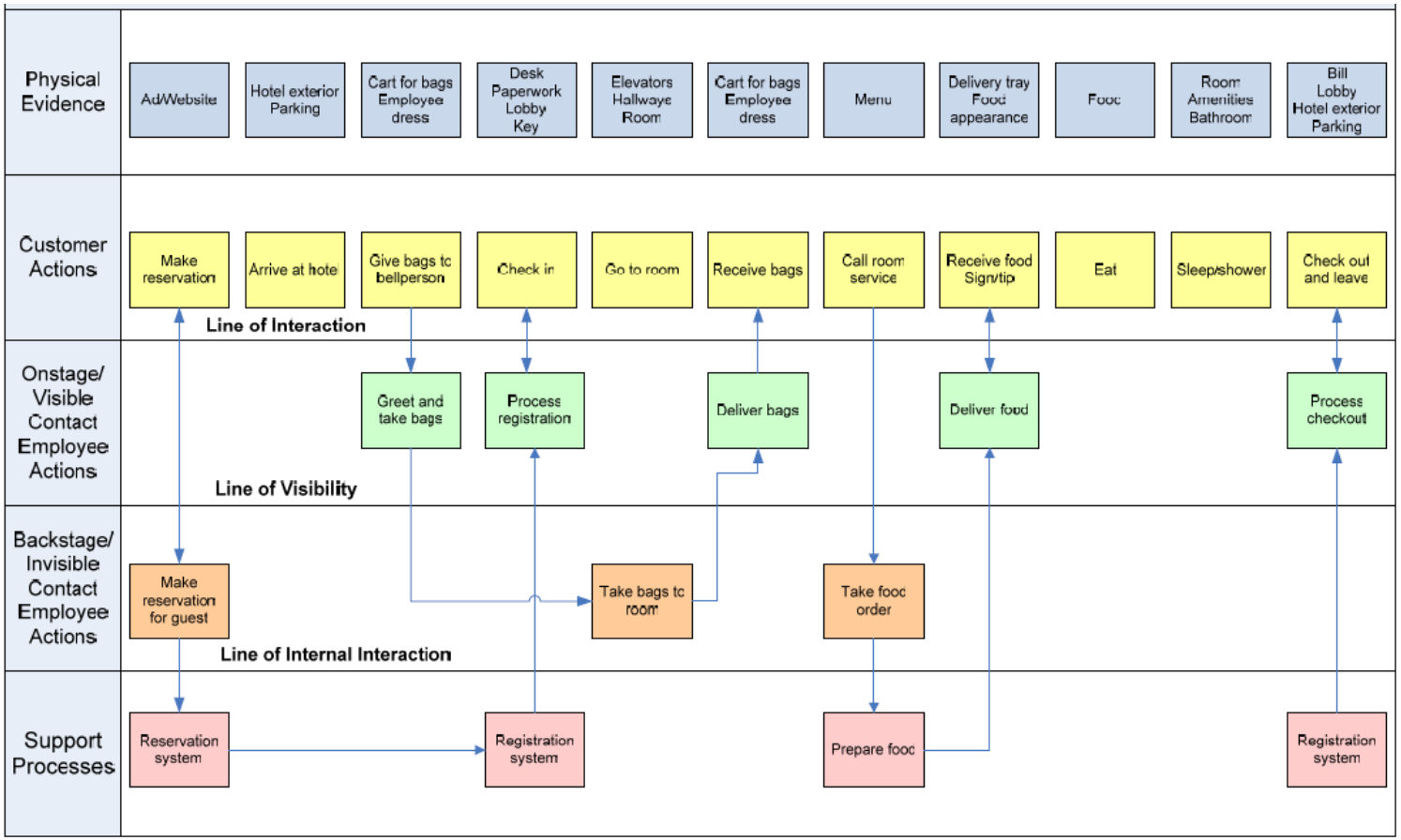
Rating = f(step$_1$, step$_2$, step$_3$…)

	Check_in	*Use facilities*	…	*Check_out*	*Overall rating*
1	5	4	…	5	5
2	4	4	…	5	4
3	5	4	…	5	4
…	…	…	…	…	…
M	2	1	…	2	1



With This Approach, We Can Measure/Improve the Service in Blueprint

- A customer-focused approach for service design and service improvement



A Blueprint for Overnight Hotel Stay Service

Customer-oriented Benchmarking of Service Firms

Clustering of Firms for Benchmarking: A Dataset Perspective

- Review-feature matrix

	Feature 1	Feature 2	Feature 3	…	Feature m-1	Feature m
Review 1	…	…	…	…	…	…
Review 2	…	…	…	…	…	…
Review 3	…	…	…	…	…	…
…	…	…	…	…	…	…
…	…	…	…	…	…	…
…	…	…	…	…	…	…
…	…	…	…	…	…	…
…	…	…	…	…	…	…
…	…	…	…	…	…	…
…	…	…	…	…	…	…
Review n-1	…	…	…	…	…	…
Review n	…	…	…	…	…	…

Clustering of Firms for Benchmarking: A Dataset Perspective

- Firm(Hotel)-review-feature matrix

		Feature 1	Feature 2	Feature 3	…	Feature m-1	Feature m
Firm 1	Review 1	…	…	…	…	…	…
…	…	…	…	…	…	…	…
Firm 2	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
Firm I-1	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…
Firm I	…	…	…	…	…	…	…
…	Review n	…	…	…	…	…	…

Clustering of Firms (Hotels) for Benchmarking

- Data collection & pre-processing
  - Employing the review content and the overall rating of customer reviews
    - Part of the customer review database

Firm No.	Review No.	Date of stay	Rating	Title	Review content (raw data)	Review content (Pre-processed data)
S2-a	1	2019-07	5	Lovely place!	This is a great place, friendly staff, conveniently located, safe and is highly recommended for anyone open to sharing. …	great place friendly staff conveniently locate safe highly recommend anyone open sharing …
S2-a	2	2019-06	5	Marvelous cabin territory	Thailand is a wonderful country and I get a kick out of the opportunity to explore each spot. A month prior I visited the said country and booked in Bed Station Hostel….	thailand wonderful country get kick opportunity explore spot month prior visit say country book bed station hostel …
S2-a	3	2019-06	5	Best hotel in Bangkok	Lovely hotel that is clean and has good amenities such as a gym, PlayStation and tv, films, books, cheap water, a bar and events …	lovely hotel clean good amenity gym playstation tv film book cheap water bar event …
…	…	…	...	…	…	…
S5-g	26932	2016-07	4	Good service staff	I get a good service from the stuff(especially Seangwanich), they are friendly and understanding. The room is nice and the facility is well maintained …	get good service stuff especially seangwanich friendly understanding room nice facility well maintain
S5-g	26933	2016-07	4	Coffee at the Bakery	This is my first time visiting at the Bakery at Phaza Athenee Bangkok Hotel. Overall decoration is so fine and nice that you feel relaxing with a cup of Illy coffee, Italian blended coffee. …	first time visit bakery phaza athenee bangkok hotel overall decoration fine nice feel relax cup illy coffee italian blend coffee …
S5-g	26934	2016-07	5	Perfect Building, Perfect Staff, …	The hotel is incredibly beautiful, having previously been a palace for a Thai Prince. The architecture is stunning, and the service matches. The staff is helpful, courteous, and professional. …	hotel incredibly beautiful previously palace thai prince architecture stun service match staff helpful courteous professional …

Clustering of Firms (Hotels) for Benchmarking

- Identification of customer-centric service features with keyword dictionary
  - Customer-centric service features employed in this example

	#	Features	Researcher (Year)					
			Bi et al. (2019)	Hu et al. (2019)	Guo et al. (2017)	Mankad et al. (2016)	Büschken & Allenby (2016)	Tssang & Qu (2000)
Data-driven (Topic modeling)	1	Food	√	√	√		√	√
	2	Service	√	√	√	√	√	√
	3	Room	√	√	√		√	√
	4	Facility	√	√	√			√
	5	Location	√	√	√	√	√	√
	6	Cleanliness	√	√				√
	7	Value	√	√	√	√		√
From literature	8	Wi-Fi/Internet	√	√		√	√	
	9	Security						√

Clustering of Firms (Hotels) for Benchmarking

- Measurement of service feature's importance and performance of the firms (hotels)(
  - Measuring service **performance** at the service feature-level
    - Identifying the linkages between a customer review sentence and service features by matching the keyword dictionary constructed in the prior step
    - Calculating four sentiment scores of the sentences via VADER
      - i.e., positive, neutral, negative, and compound score
    - Classifying customer review sentences for each of the relevant service features as positive or negative based on the compound score to reduce the measurement error
    - Example of review-performance matrix

Service #	Review #	Food	Service	…	Wifi	Security	Rating
Firm 1	Review 1	0	1	…	-1	0	5
	Review 2	1	-1	…	0	0	2
	…			…			…
Firm l	Review n	1	1	…	0	0	3

- 1 : Positive
- -1 : Negative
- 0 : Unmentioned

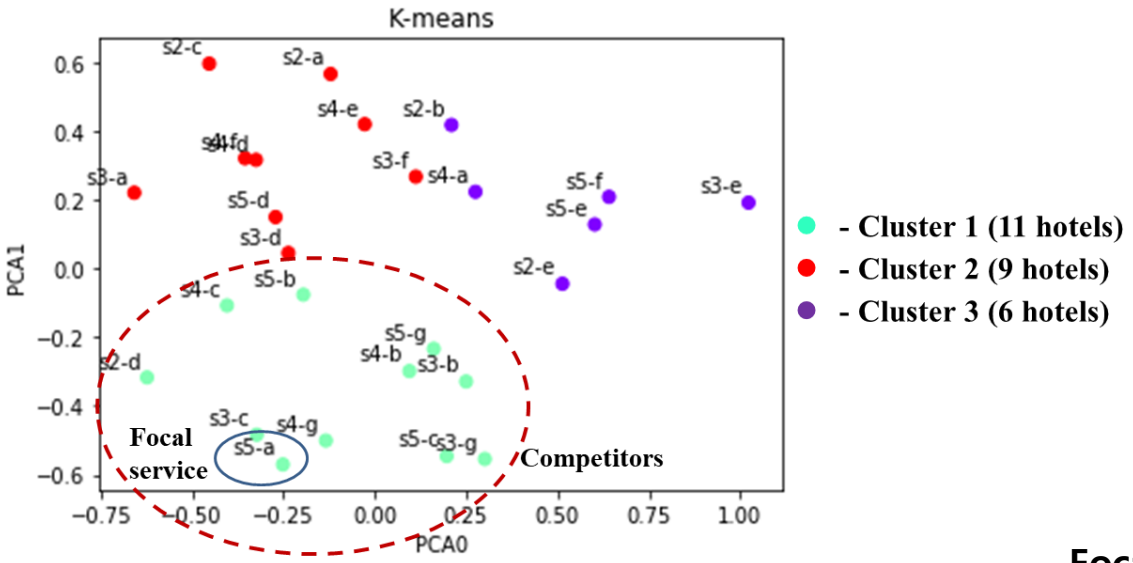
Clustering of Firms (Hotels) for Benchmarking

- Measurement of service feature's importance and performance
  - Measuring **importance** of the service features (cont'd)
    - Part of the service-feature importance matrix

Hotel	Coefficient values of the service features								
	Food	Service	Room	Facility	Location	Cleanliness	Value	Wi-Fi	Security
S2-a	0.309	0.634	0.383	0.000	0.234	0.618	0.274	1.000	0.571
S2-b	0.369	0.655	0.670	0.000	0.284	0.813	0.437	1.000	0.168
S2-c	0.301	0.483	0.608	0.171	0.000	0.531	0.192	0.726	1.000
S2-d	0.105	0.033	0.120	0.097	0.000	1.000	0.292	0.274	0.551
S2-e	0.000	0.788	0.567	0.805	0.020	1.000	0.792	0.591	0.021
S3-a	0.175	0.235	0.667	0.363	0.124	0.787	0.000	0.389	1.000
…	…	…	…	…	…	…	…	…	…
S5-a	0.075	0.328	0.211	0.213	0.000	1.000	0.332	0.035	0.200
S5-b	0.227	0.587	0.502	0.150	0.000	1.000	0.199	0.425	0.338
S5-c	0.346	0.523	0.533	0.321	0.081	1.000	0.550	0.020	0.000
S5-d	0.154	0.436	0.785	0.313	0.000	1.000	0.286	0.531	0.566
S5-e	0.516	0.812	0.694	0.612	0.373	1.000	0.556	0.774	0.000
S5-f	0.268	1.000	0.405	0.584	0.138	0.000	0.773	0.341	0.324
S5-g	0.320	0.504	0.521	0.275	0.159	1.000	0.425	0.428	0.000
Average	0.261	0.571	0.504	0.289	0.147	0.840	0.384	0.483	0.388

Clustering of Firms (Hotels) for Benchmarking

- Identification of the competitors and best practices as benchmark
  - Identifying the **competitors** via *K*-means clustering
    - Conducting the *k*-means clustering algorithm to identify competitor groups
    - Identified groups for 26 hotels in Bangkok

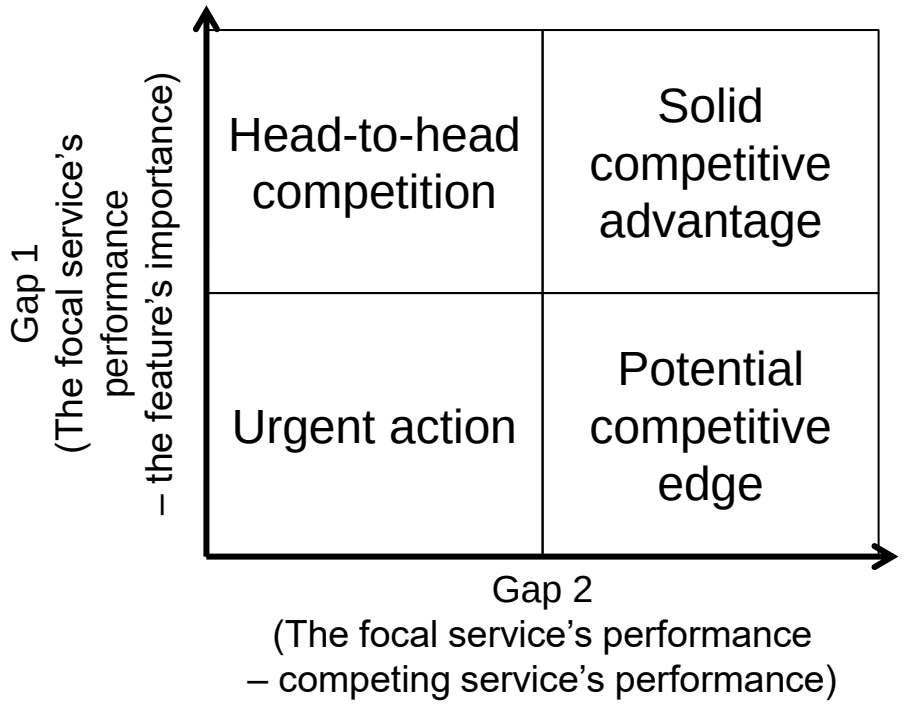


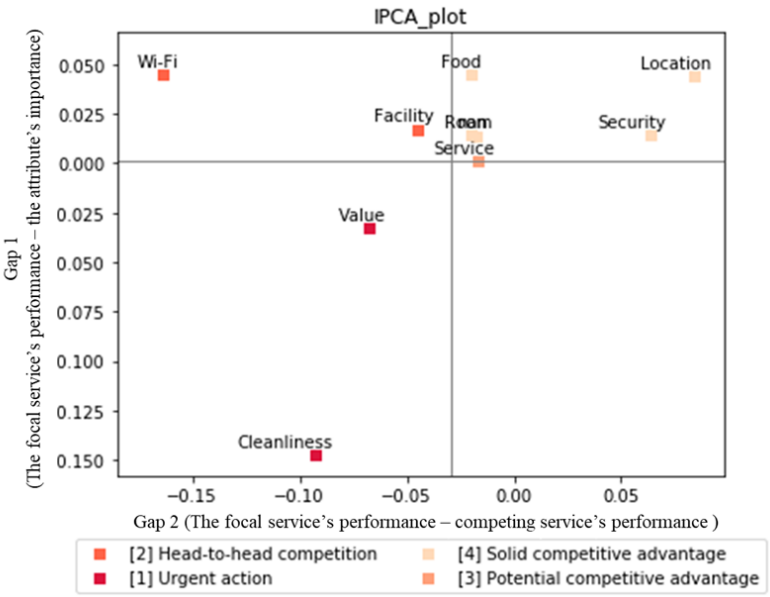
Focal hotel

Hotel	S2-d	S3-b	S3-c	**S3-g**	S4-b	S4-c	**S4-g**	S5-a	**S5-b**	S5-c	S5-g
Performance	0.45	0.459	0.445	0.839	0.426	0.543	0.724	0.604	0.623	0.492	0.564
Ranking	9	8	10	1	11	6	2	4	3	7	5

Clustering of Firms (Hotels) for Benchmarking

- Prioritization of service features and development of a strategic action plan for service improvement
  - Employing Importance Performance Competitor Analysis (IPCA)
    - Examining two types of gaps,
      - Gap 1 between the focal service's performance and the importance for the feature
      - Gap 2 between focal and competing benchmarks' performance scores for the feature

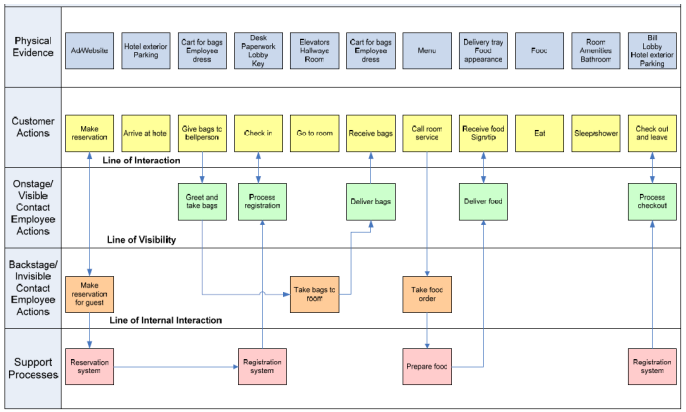


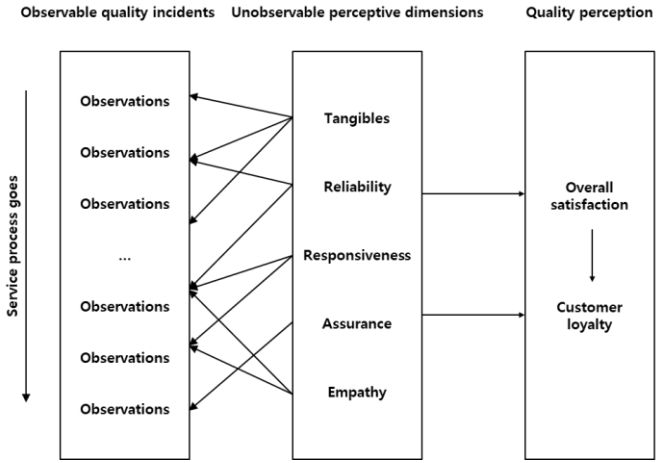


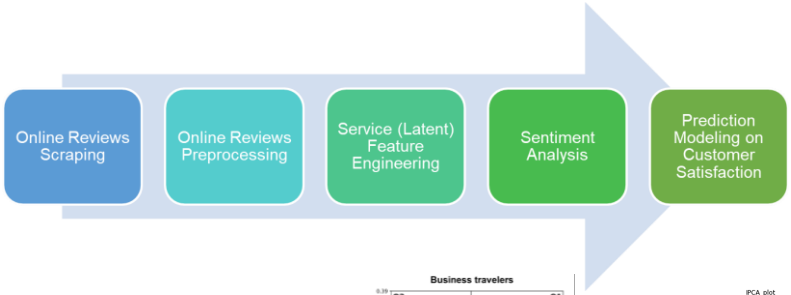
Assignment 4 (by 10.7 11:59 pm)

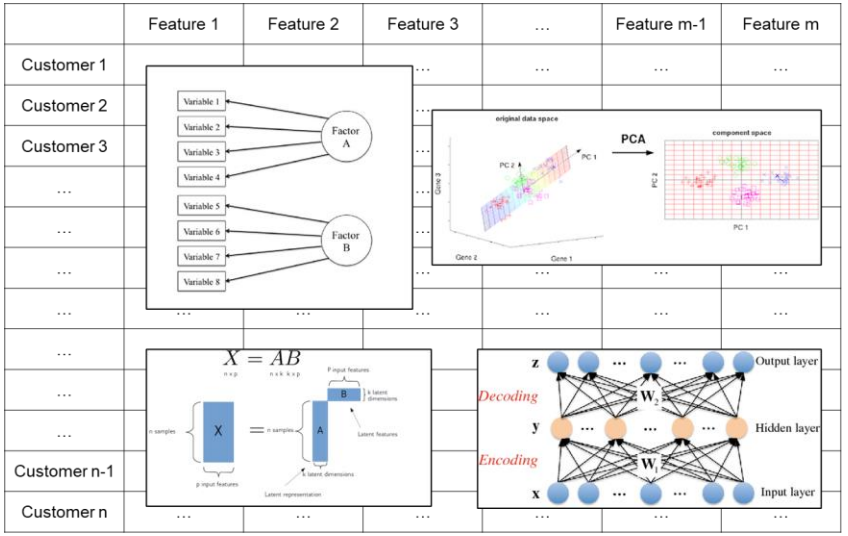
- By yourself, (1) complete the construction of the review-feature dataset on service quality of hotels in Singapore based on the practice demonstrated by the TA. Then, (2) using the review-feature matrix you constructed, develop a service quality prediction model for the hotels in Singapore (i.e., predict the customer's quality evaluation with review data). Do it all by yourself, and describe the analysis process and outcome in detail. Interpret the outcome (e.g., name the service features you identified, interpret the coefficient/importance values of service features to the quality ratings).
- (3) What other interesting machines can be developed using the review-feature matrix dataset you constructed? Describe your ideas in detail (e.g., describe the learning objective and process). Try to think your own creative, unique ideas! You have completed the basic review mining activities (tasks 1 and 2) as well as your own idea generation (task 3). Then, (4) describe how you can use your machine(s) to automate the monitoring, evaluation, and improvement of hotel service quality? Imagine you are working for a real hotel.
- Using a similar approach that you have practiced so far, (5) what other services can be improved using review mining machines or another intelligent machine that learns other types of raw data traces of service quality (e.g., customer behavior data)? Assume that you actually manage the quality of service in question. (6) How would you conduct this job in your own creative, unique way? What kinds of data and methods are you going to collect, analyze, and learn? Describe your service intelligence development plan in detail. If possible, visualize your plan clearly (e.g., draw an image, construct a mathematical model).
- Upload your code and a several paragraph essay on the tasks (1)~(6) in the Blackboard.

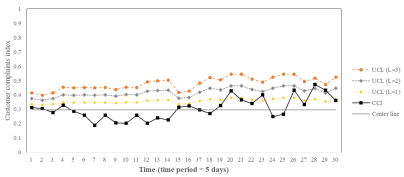
A Big Picture of Data-driven Service Quality Evaluation and Improvement

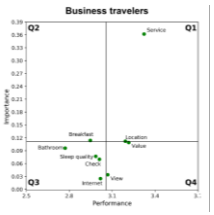


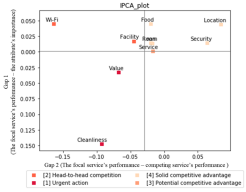












A Big Picture of Data-driven Service Quality Evaluation and Improvement

- What are the limitations of the multi-step framework of service quality evaluation?

  How can we address this research issue with advanced techniques of machine learning?

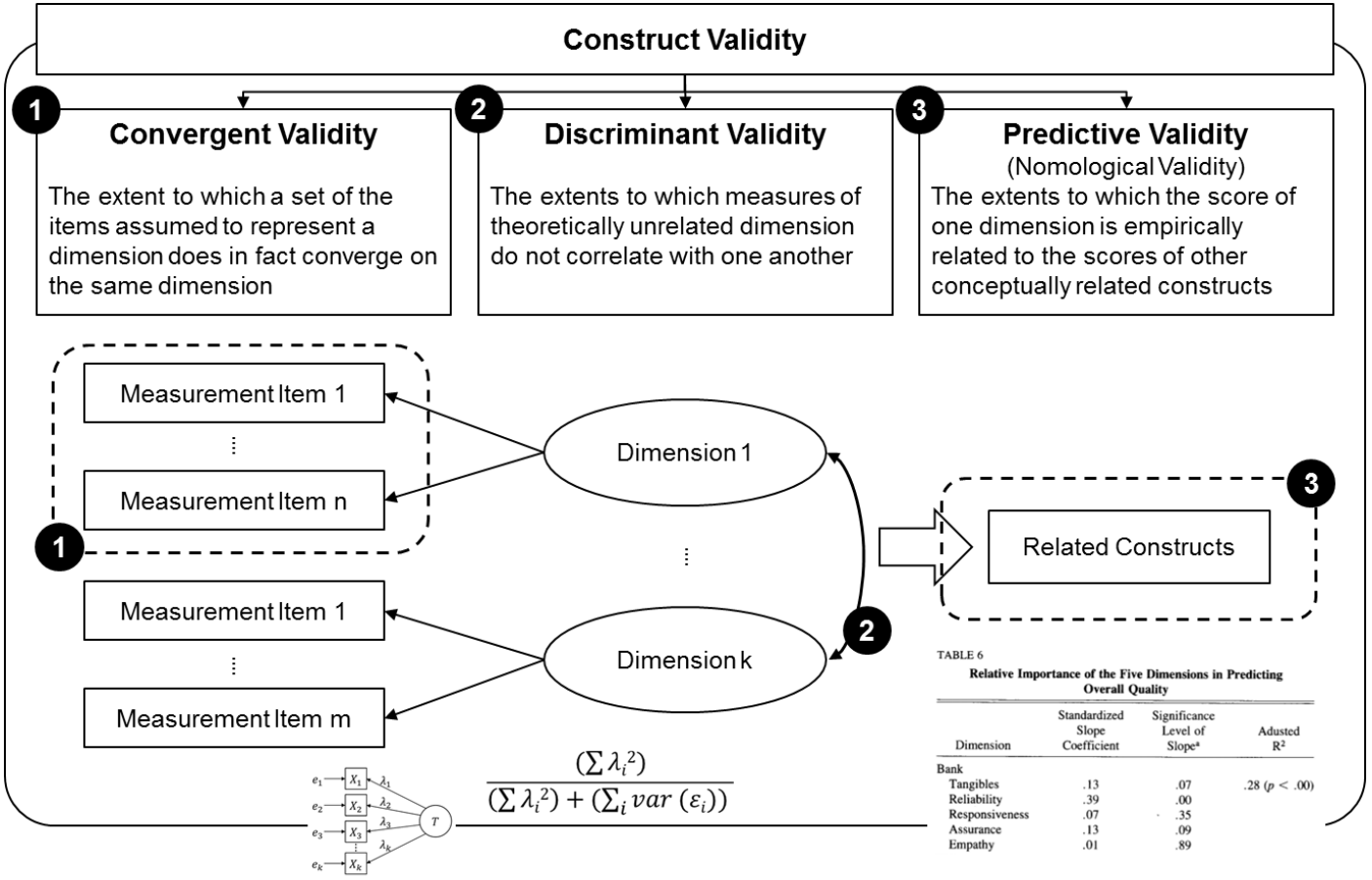


TABLE 6

Relative Importance of the Five Dimensions in Predicting Overall Quality

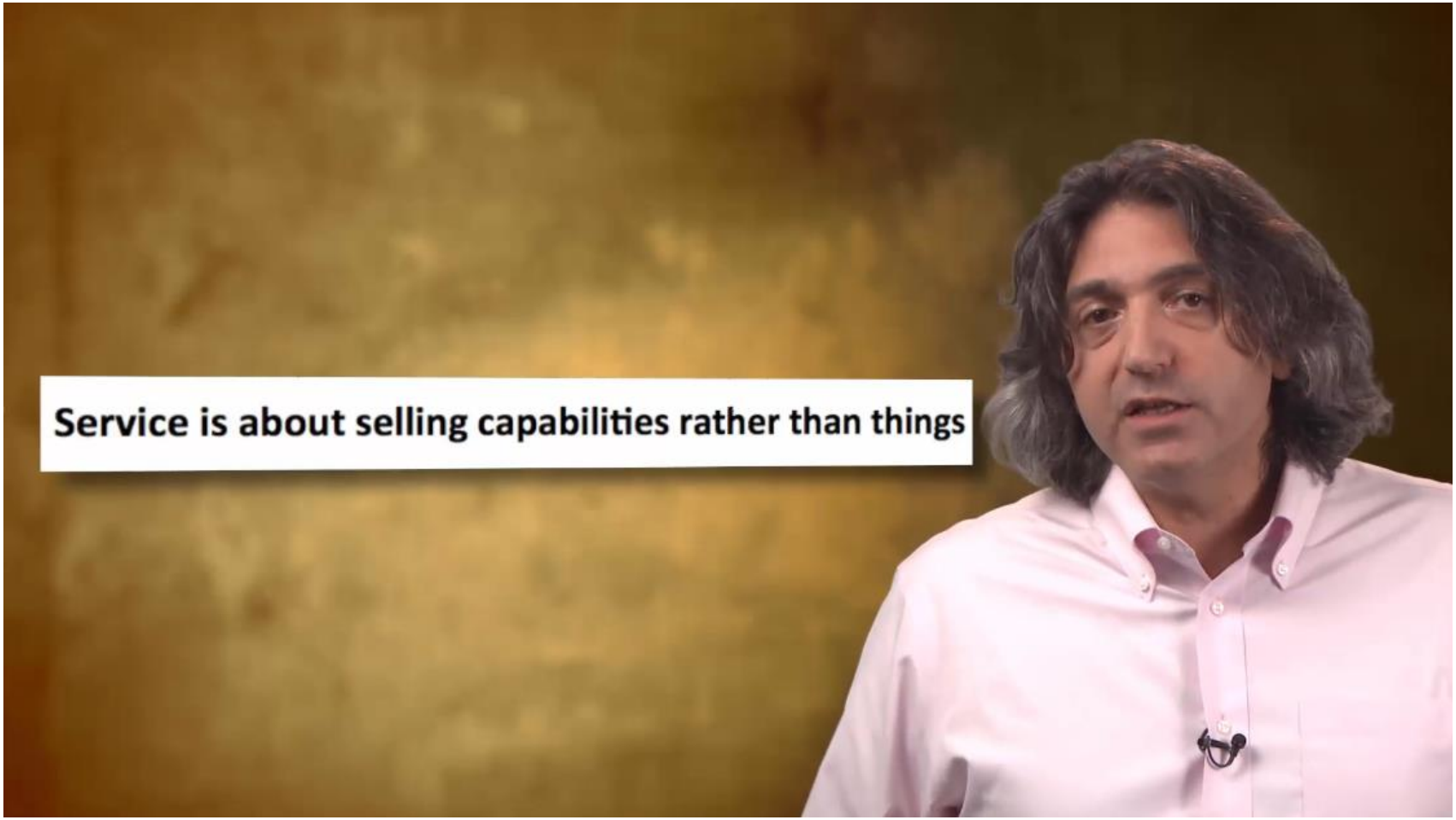
Dimension	Standardized Slope Coefficient	Significance Level of Slope[a]	Adusted R^2
Bank			
Tangibles	.13	.07	.28 ($p < .00$)
Reliability	.39	.00	
Responsiveness	.07	.35	
Assurance	.13	.09	
Empathy	.01	.89	

Reference: Peter (1981)

Service Intelligence Week 5.

[Service Customization and Customer Segmentation]

Service is Simply to Serve Customers: Help Tasks or to Do the Tasks

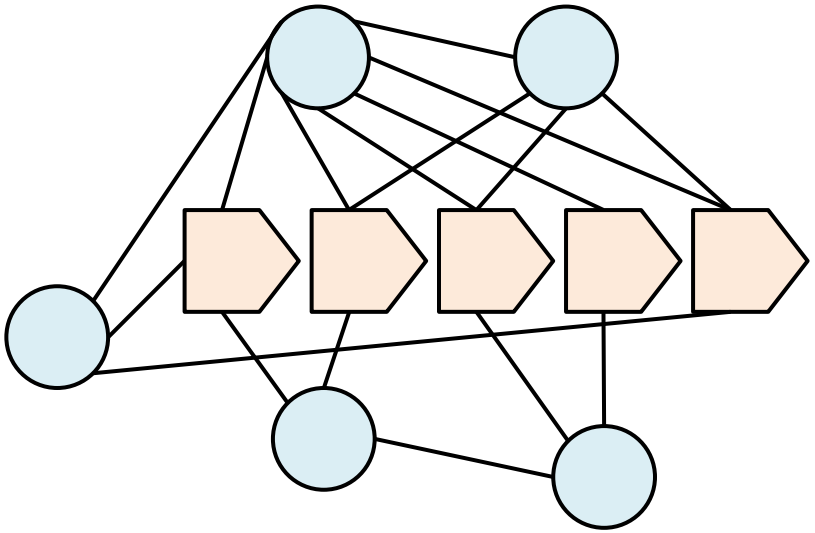


All Customers are Different

Customization is the Essence in Service Design and Delivery

consumer behaviors
(user behaviors)

service delivery
and experience



quality

relationship

value cocreation

technology and
self-service

innovation

design and
development

productivity

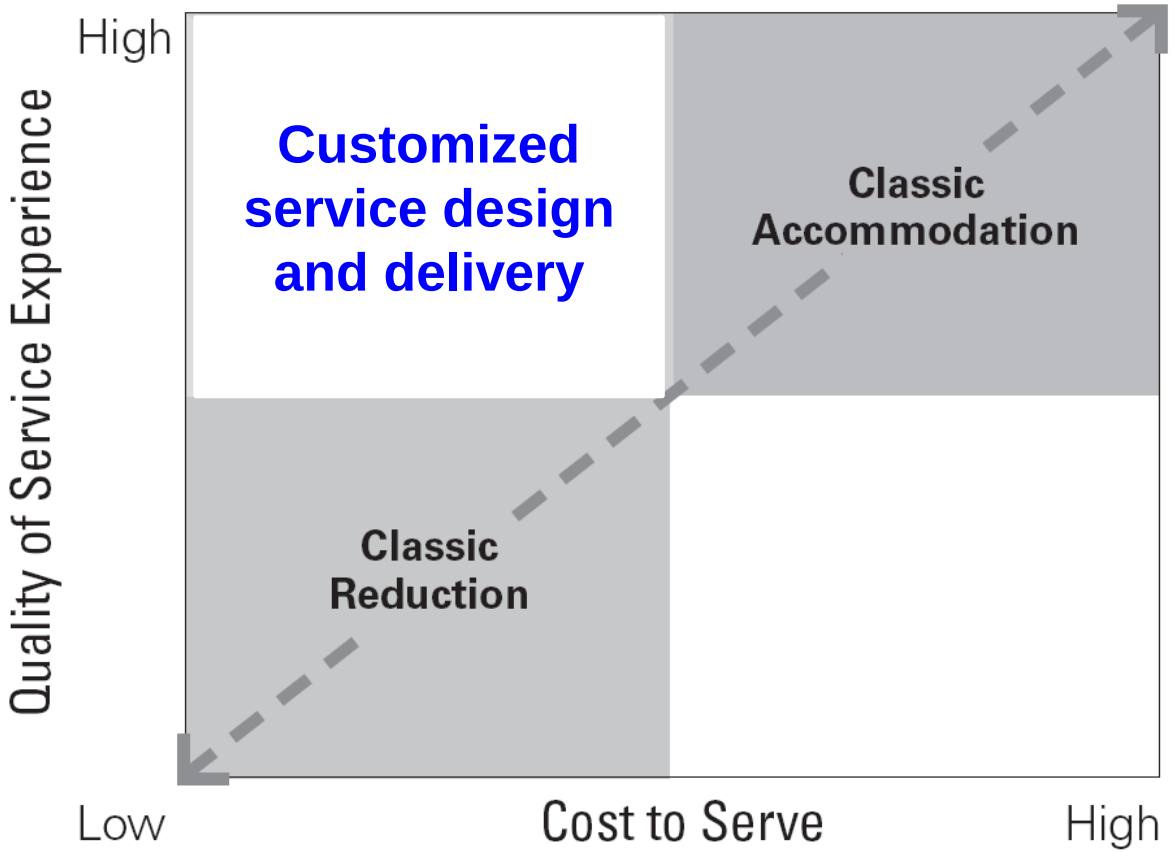
knowledge and
employee
management

cost and pricing

service operations
and supply chain
management

Customization is the Essence in Service Design and Delivery

- Service customization should aim to accommodate and control customer variability at the same time



Service Customization Requires Customer Segmentation

- Motorola case example of customer segmentation based on a survey approach

Step#2 — **Choose the segmentation criteria**

Step#3 — **Conduct cluster analysis**

Segment 1 Opportunities	Segment 2 Opportunities	Segment 3 Opportunities
4. Minimize the effort required to communicate discreetly	1. Minimize the number of messages that are misunderstood	5. Minimize the number of annoying incoming communications
7. Minimize the effort required to establish a record of the communication	2. Minimize the number of interruptions during a communication	6. Minimize the time it takes to confirm receipt of a communication
8. Minimize the number of communications that can be intercepted	3. Minimize the amount of interference encountered when communicating	10. Minimize the effort required to program the device
	9. Minimize the likelihood of making inadvertent changes to established settings	
	11. Minimize the effort to operate the device with gloves on	

Clustering algorithm

- Focused on the opportunity ratings given to the eleven selected outcomes
- Placed the respondents surveyed into a predetermined number of segments based on their responses

Step#4 — **Profile the clusters**

Reference: What customers want (Ulwick, 2005)

Service Customization Requires Customer Segmentation

- Motorola case example of customer segmentation based on a survey approach

Step#3 Conduct cluster analysis

Step#4 Profile the clusters

	Segment 1: Privacy	Segment 2: Emergency	Segment 3: Administrative
Outcomes Desired	• Discreet communications • Record of communications • Low interceptions	• Clear messages • Few interruptions • Lower interference • Low risk of inadvertent changes to settings • Easy to use with gloves on	• Few unimportant incoming calls • Quick receipt confirmation • Easy to program device
Characteristics	• Covert operations inside vehicle • Younger • High urban concentration	• Firefighters, police, security personnel • Often have to leave vehicle • Must maintain contact at all times	• Coast guard, locomotive engineers, etc • Rely on radio for their daily job • Perform admin tasks
Resulting Solution	• Enhanced encryption • A mechanism to prevent others from overhearing communications • Noiseless operation	• Voice command technology • Emergency locators • Modifications to permit use with gloves	• Easier-to-program radio • Mechanisms to ensure message receipt

Cluster interpretation

- Discovering segments of high potential growth
- Identifying demanding customer segments that may be willing to pay more for more elaborate solutions
- Identifying customer segments that are unattractive and should not be targeted

Reference: What customers want (Ulwick, 2005)

Customer Segmentation: A Dataset Perspective

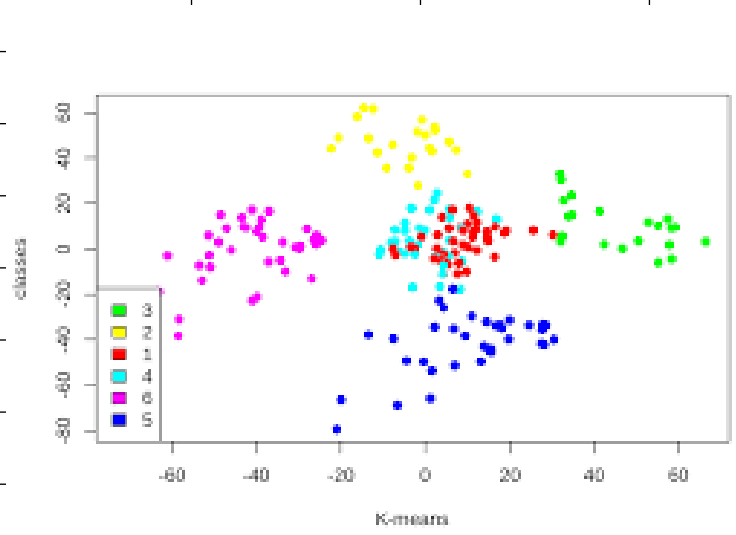
- Customer-feature matrix

	Feature 1	Feature 2	Feature 3	…	Feature m-1	Feature m
Customer 1	…	…	…	…	…	…
Customer 2	…	…	…	…	…	…
Customer 3	…	…	…	…	…	…
…	…	…	…	…	…	…
…	…	…	…	…	…	…
…	…	…	…	…	…	…
…	…	…	…	…	…	…
…	…	…	…	…	…	…
…	…	…	…	…	…	…
…	…	…	…	…	…	…
Customer n-1	…	…	…	…	…	…
Customer n	…	…	…	…	…	…

Customer Segmentation: A Dataset Perspective

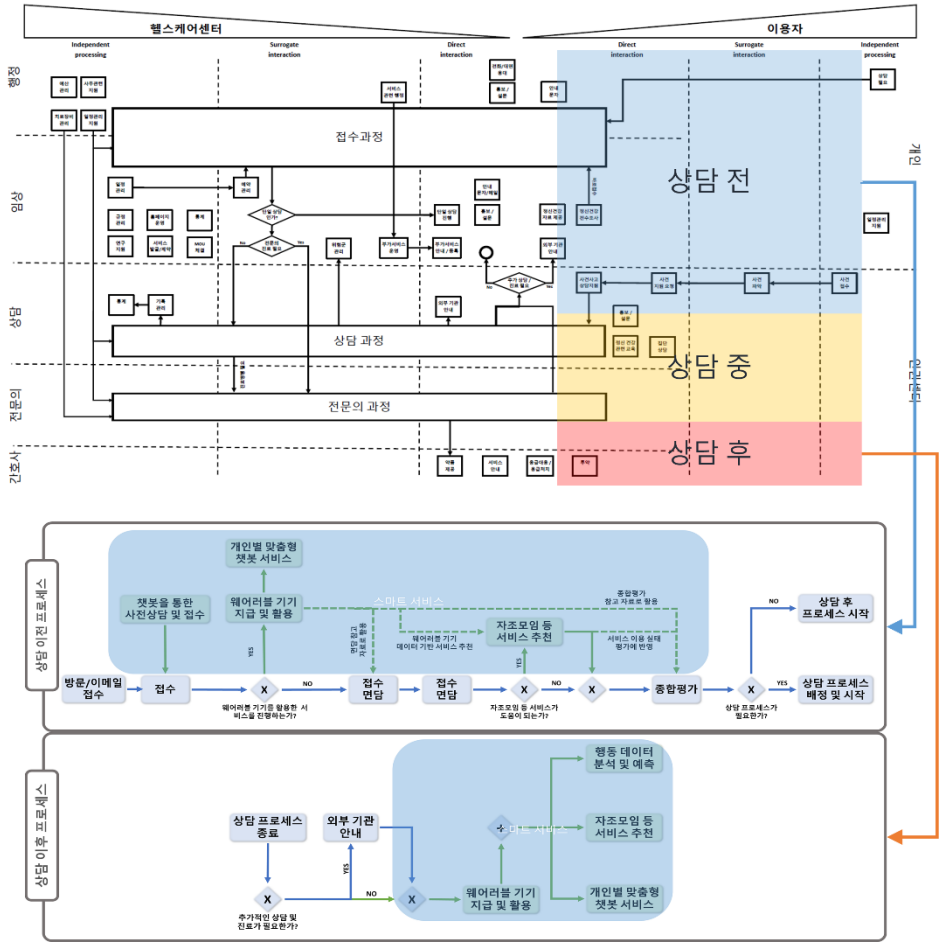
- Customer-feature matrix

	Feature 1	Feature 2	Feature 3	…	Feature m-1	Feature m
Customer 1	…	…	…	…	…	…
Customer 2	…	…	…	…	…	…
Customer 3	…					…
…	…					…
…	…					…
…	…					…
…	…					…
…	…					…
…	…					…
…	…					…
Customer n-1	…	…	…	…	…	…
Customer n	…	…	…	…	…	…



Customer Segmentation: Students Clustering for Counseling/Advising

Customer Segmentation: Students Clustering for Counseling/Advising



Customer Segmentation: Students Clustering for Counseling/Advising

Customer Segmentation: Students Clustering for Counseling/Advising

489 respondents from the 5 universities

	질문	1차설문	2차설문
1	너의 취미가 뭐야?	o	o
2	너에게 중요한 관계는 누구야?	o	o
3	그러면 너가 일 또는 공부를 하고 있는 혹은 일하게 될 공간에서 동료와의 관계는 어땠으면 좋겠어?	o	o
4	너가 설계하는 미래의 커리어가 있을까?	o	o
5	이번에는 좀 더 넓게, 너가 꿈꾸는 삶은 어떤 삶이야? 흔히 라이프스타일이라고들 하지!	o	o
6	그러면 너가 현재 어떤 것을 잘 하는지 궁금해 지는데, 혹시 난 이걸 좀 잘하는 거 같아 하는게 있을까?	o	o

What are the Main Difficulties in Students' Task Completion??

12	그러면 일을 처리할 때 너의 패턴이 있을거잖아, 일할 때 습관이라던지.. 그 중에 일을 잘 할수 있게 해 주는 요소가 있을까?	o	o
13	이번엔 해야 할 일을 미루게 하거나, 못하게 하는 요소는 어떤 게 있을까?	o	o
14	어떤 마음이, 어떤 성격이 해야할 일(활동)을 하는데 방해가 되는 거 같아?	o	o
15	일을 처리할 때의 너의 습관 중에서 일을 방해하는 요소가 있을 텐데 그게 뭐야?	o	x
16	너의 삶에서 가치있는 것들 3가지만 말해볼래?	o	o
17	그러면 이 가치들을 실현하기 위한 너의 삶의 목표를 세워서 말해줘!	o	x
18	[목표를 모두 수행했을 경우] 많이 힘들었을텐데, 모두 잘 이행했구나! 이렇게 잘 이행할 수 있었던 이유가 뭔거 같아?	o	x
19	[목표를 절반만 수행했을 때] 열심히 수행하려고 노력했는데, 맘대로 되지 않을 때도 있지 ㅠㅠ 모든 목표를 다 수행하지 못한 이유가 있을까?	o	x
20	[목표를 하나도 수행하지 못했을 때] 혹시 오늘 하루 힘든 일이 있는거야...? ㅠㅠ 가끔은 이렇게 쉬어가도 괜찮지만, 혹시 무슨일이 있는지 걱정	o	o
21	삶에서 이루고 싶은 것을 두가지 정도만 말해줄래?	x	o
22	위 목표를 이루기 위한 1년 후 이루고 싶은 목표를 세워보자!	x	o
23	그럼 한달 안에 이룰 수 있는 목표를 2가지만 말해줄래?	x	o
24	그러면 위 목표를 이루기 위해 내일까지 뭘 해야할까? 목표를 구체적으로 2가지 이상 적어줘!	x	o

Customer Segmentation: A Dataset Perspective

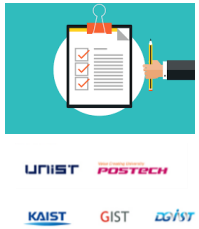
- Customer-feature matrix

	Word feature 1	Word feature 2	Word feature 3	…	Word feature m-1	Word feature m
Customer 1	…	…	…	…	…	…
Customer 2	…	…	…	…	…	…
Customer 3	…	…	…	…	…	…
…	…	…	…	…	…	…
…	…	…	…	…	…	…
…	…	…	…	…	…	…
…	…	…	…	…	…	…
…	…	…	…	…	…	…
…	…	…	…	…	…	…
…	…	…	…	…	…	…
Customer n-1	…	…	…	…	…	…
Customer n	…	…	…	…	…	…

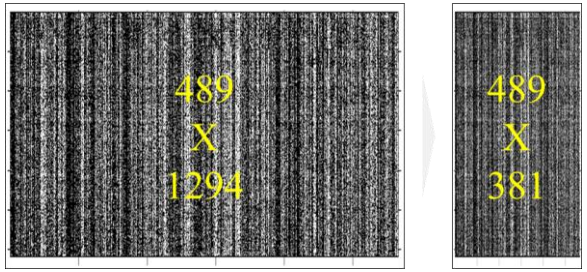
Customer Segmentation: Students Clustering for Counseling/Advising

1. Collection of text data on the topic in question
2. Semi-automatic semantic identification of significant word-features
3. Unsupervised learning of the data with selected word-features
4. Interpretation

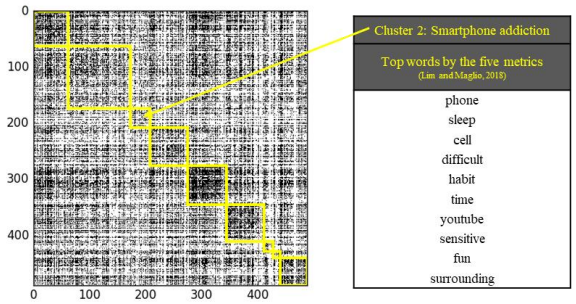
1



2



3



4

- 스트레스 (Stress)
- 완벽주의 (Perfectionism)
- 스마트폰 (Smartphone)
- 너무피곤 (Tiredness)
- 인간관계 (Relationship)
- 놀고싶어 (WantPlay)
- 귀차니즘 (Laziness)
- 자기안심 (Self-justification)
- 일이많음 (TooManyWorks)

Vector Space of a Text Dataset

- Represent the dataset as a vector space (a "document-term matrix")

	time	stress	sleep	lazy	thing	...	youtube
Student 1	3	3	2	0	1	...	0
Student 2	5	0	0	3	3	...	0
...	...	...	...	...	...	...	...
Student m	1	0	5	0	0	...	2

To find the significance of a term in a document

- Calculate a "TF-IDF" value and transform the vector space

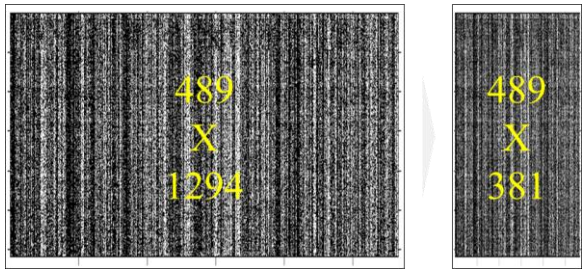
$tf-idf\ value\ after\ L2-normalization, when$
$tf-idf(t,d) = tf(t,d) \times (log \frac{1+n_d}{1+df(d,f)} + 1)$

	time	stress	sleep	lazy	thing	...	youtube
Student 1	0.291	0.120	0.071	0	0.142	...	0
Student 2	0.231	0	0	0.146	0.228	...	0
...	...	...	...	...	...	...	...
Student m	0.038	0	0.115	0	0	...	0.243

Identification of Significant Word-Features

- Three types of words in a corpus
  - (1) Case-sensitive words such as respondent's acronym: "LOL" "STAR" …
  - (2) Representative words of the overall topic: "task" "sleep" "stress" "lazy" …
  - (3) General words in surveys and English words: "my" "wrote" "within" …

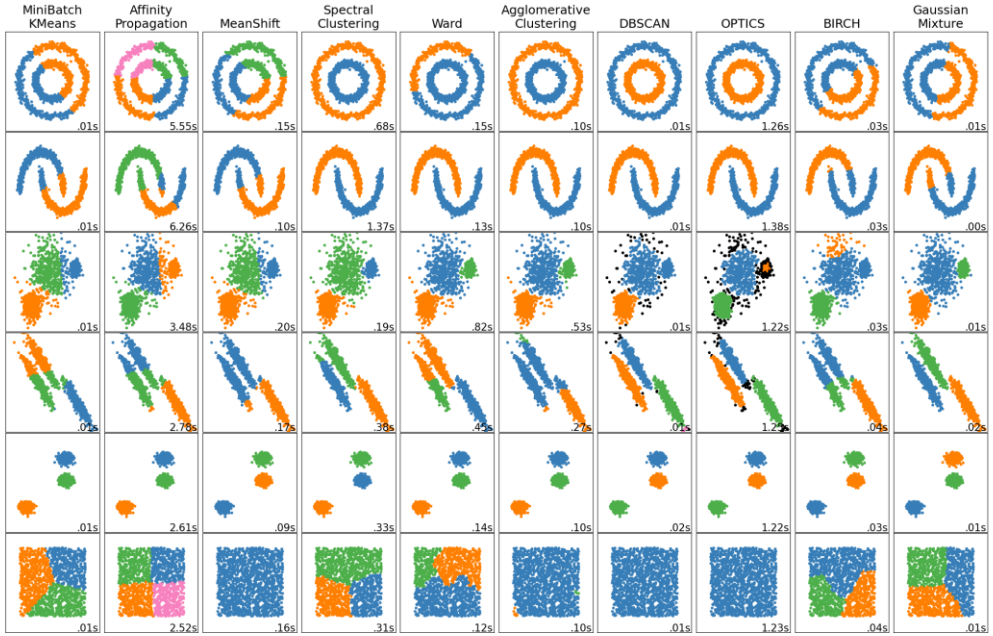
- Requirements of the selection of significant word-features
  - Include the type (2) words as much as possible
  - Exclude the types (1) and (3) words as much as possible



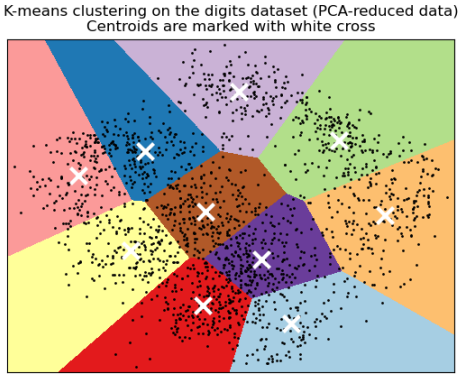
- Types (1) and (2) words have high TF-IDF values, while Type (3) words have low values

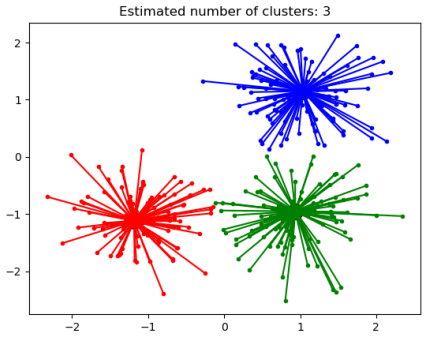
Reference: Lim, C. and Maglio, P., "Data-driven Understanding of Smart Service Systems through Text Mining," *Service Science*, Vol. 10, No. 2, 154-180, 2018.

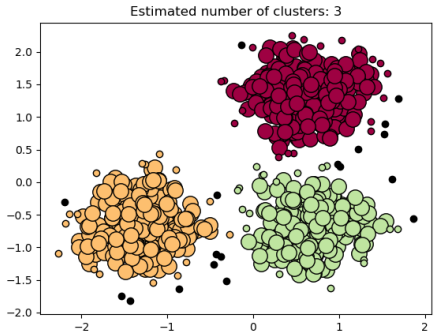
Clustering of the Student Data

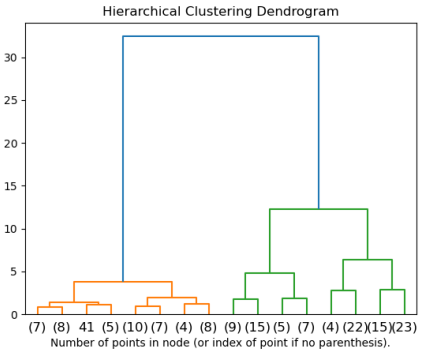


$$\sum_{i=0}^{n} \min_{\mu_j \in C} (||x_i - \mu_j||^2)$$



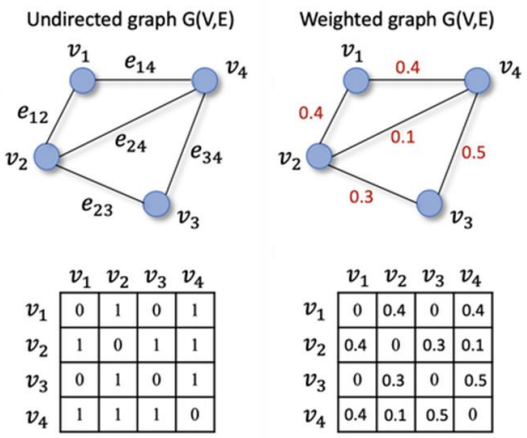






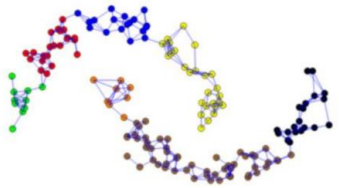
Reference: https://scikit-learn.org/stable/modules/clustering.html#hierarchical-clustering

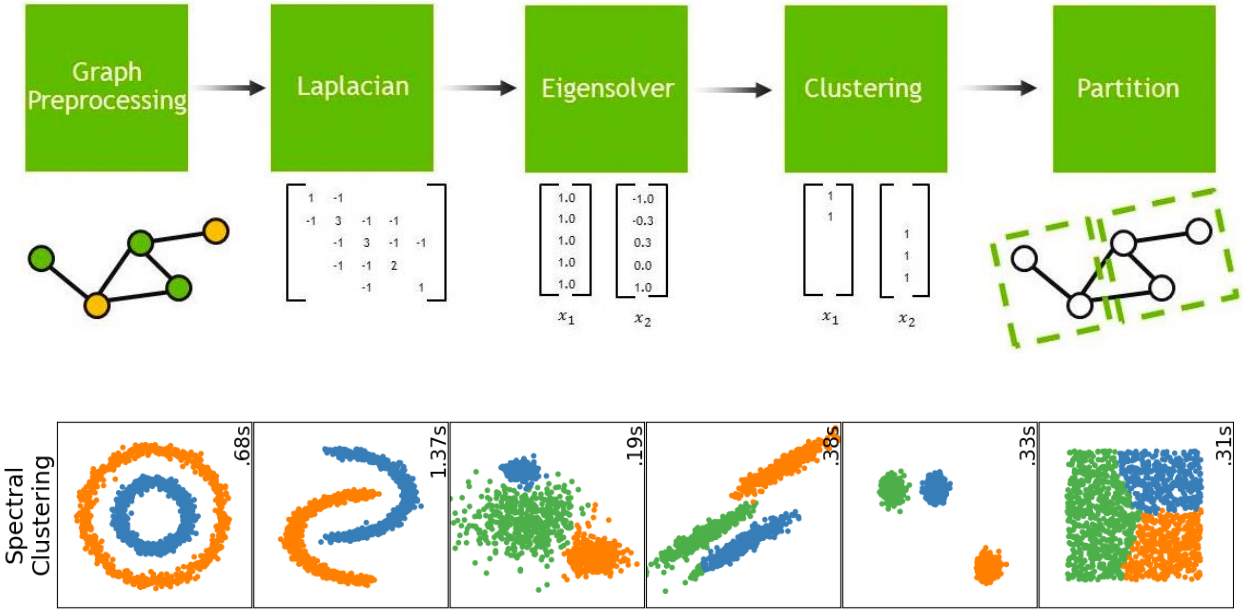
Spectral Clustering of the Student Data



	v_1	v_2	v_3	v_4
v_1	0	1	0	1
v_2	1	0	1	1
v_3	0	1	0	1
v_4	1	1	1	0

	v_1	v_2	v_3	v_4
v_1	0	0.4	0	0.4
v_2	0.4	0	0.3	0.1
v_3	0	0.3	0	0.5
v_4	0.4	0.1	0.5	0





- Euclidean distance
- Cosine similarity
- Jaccard coefficient
- Pearson correlation coefficient
- …

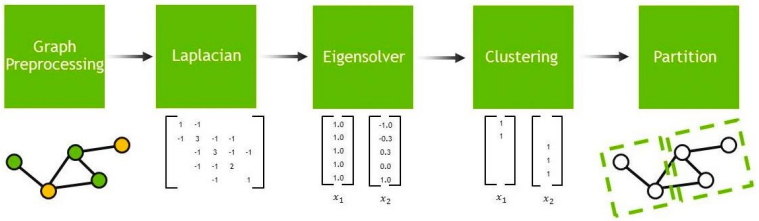
$$\mathrm{Sim}(u_a, u_b) = \frac{\sum_{i=1}^{i'} (r_{u_{a,i}})(r_{u_{b,i}})}{\sqrt{\sum_{i=1}^{n} (r_{u_{a,i}})}\sqrt{\sum_{i=1}^{n} (r_{u_{b,i}})}}$$

$$\mathrm{Sim}(u_a, u_b) = \frac{|I_{u_a}| \cap |I_{u_b}|}{|I_{u_a}| \cup |I_{u_b}|}$$

$$\mathrm{Sim}(u_a, u_b) = \frac{\sum_{i=1}^{i'} (r_{u_{a,i}} - r_{\overline{u_a}})(r_{u_{b,i}} - r_{\overline{u_b}})}{\sqrt{\sum_{i=1}^{i'} (r_{u_{a,i}} - r_{\overline{u_a}})^2} \cdot \sqrt{\sum_{i=1}^{i'} (r_{\overline{u_{b,i}}} - r_{\overline{u_b}})^2}}$$

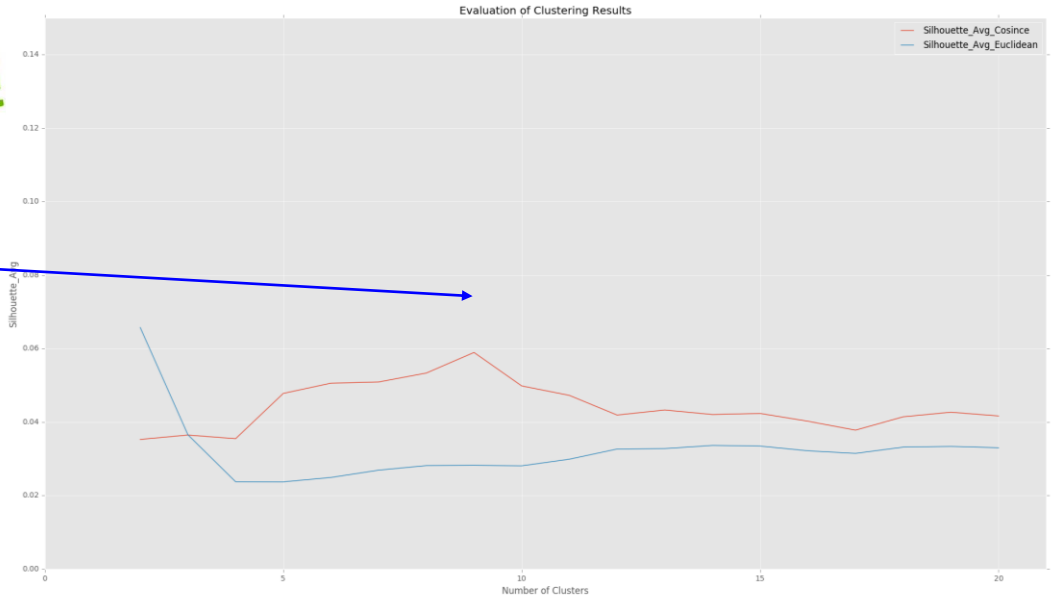
Spectral Clustering of the Student Data

- Silhouette Coefficient score graph (n_clusters = 2 to 20; 300 iterations for each)
  - The graph partitioning problem in a spectral clustering is an NP-hard problem
  - The clustering result and the corresponding Silhouette Coefficient score change every time



The mean Silhouette Coefficient scores based on the cosine and Euclidean distance are high at ‘n_clusters = 9’

After a manual interpretation, 9 was considered an optimal number of clusters



$$s(i) = \frac{(b(i) - a(i))}{(\max(a(i), b(i))}$$

Interpretation of the Clusters

Cluster 0. 스트레스 (Stress)

Cluster 1. 완벽주의 (Perfectionism)

Cluster 2. 스마트폰 (Smartphone)

Cluster 3. 너무피곤 (Tiredness)

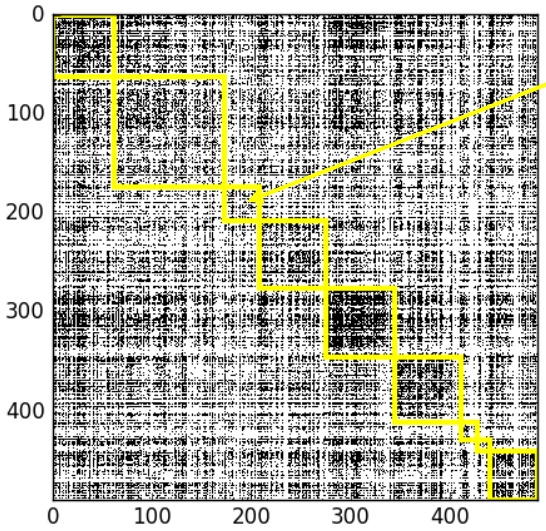
Cluster 4. 인간관계 (Relationship)

Cluster 5. 놀고싶어 (WantPlay)

Cluster 6. 귀차니즘 (Laziness)

Cluster 7. 자기안심 (Self-justification)

Cluster 8. 일이많음 (TooManyWorks)



Cluster 2: Smartphone addiction
Top words by the five metrics (Lim and Maglio, 2018)
phone
sleep
cell
difficult
habit
time
youtube
sensitive
fun
surrounding

Interpretation of the Clusters

Cluster 0. 스트레스 (Stress)

Cluster 1. 완벽주의 (Perfectionism)

Cluster 2. 스마트폰 (Smartphone)

Cluster 3. 너무피곤 (Tiredness)

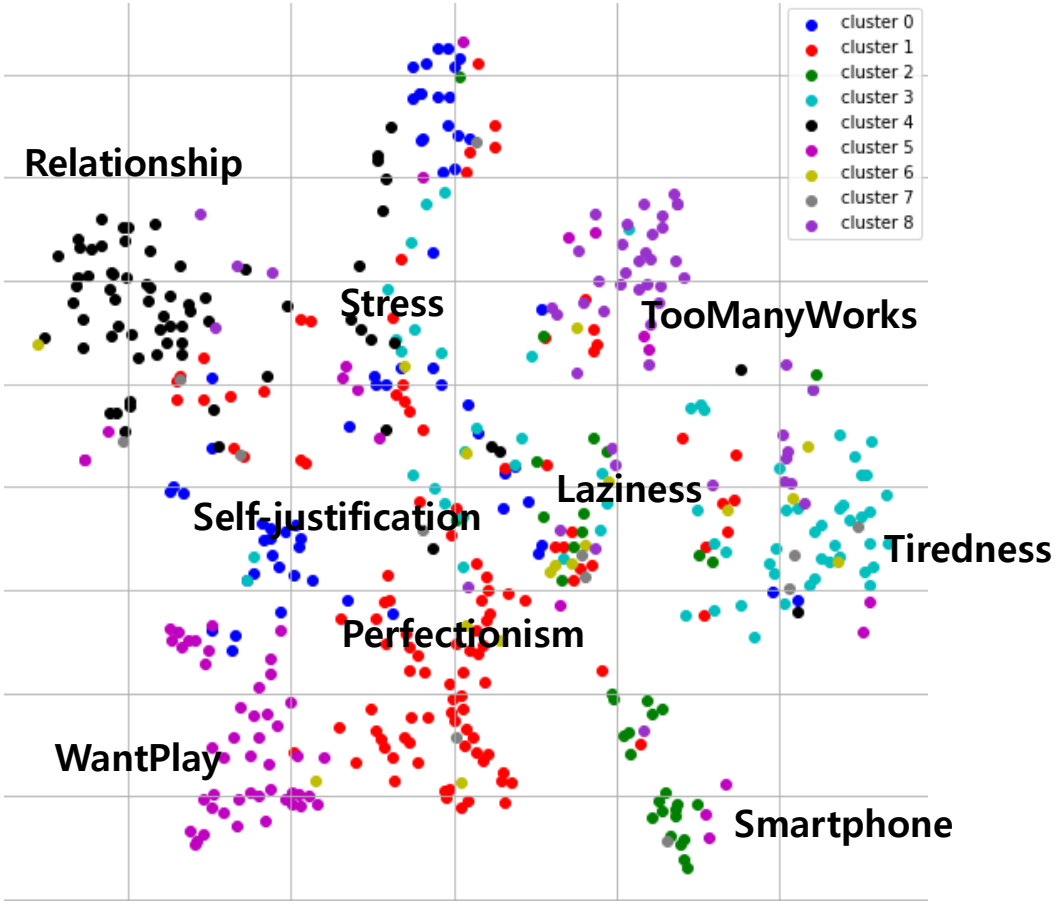
Cluster 4. 인간관계 (Relationship)

Cluster 5. 놀고싶어 (WantPlay)

Cluster 6. 귀차니즘 (Laziness)

Cluster 7. 자기안심 (Self-justification)

Cluster 8. 일이많음 (TooManyWorks)



Interpretation of the Clusters

Cluster 0. 스트레스 (Stress)

Cluster 1. 완벽주의 (Perfectionism)

Cluster 2. 스마트폰 (Smartphone)

Cluster 3. 너무피곤 (Tiredness)

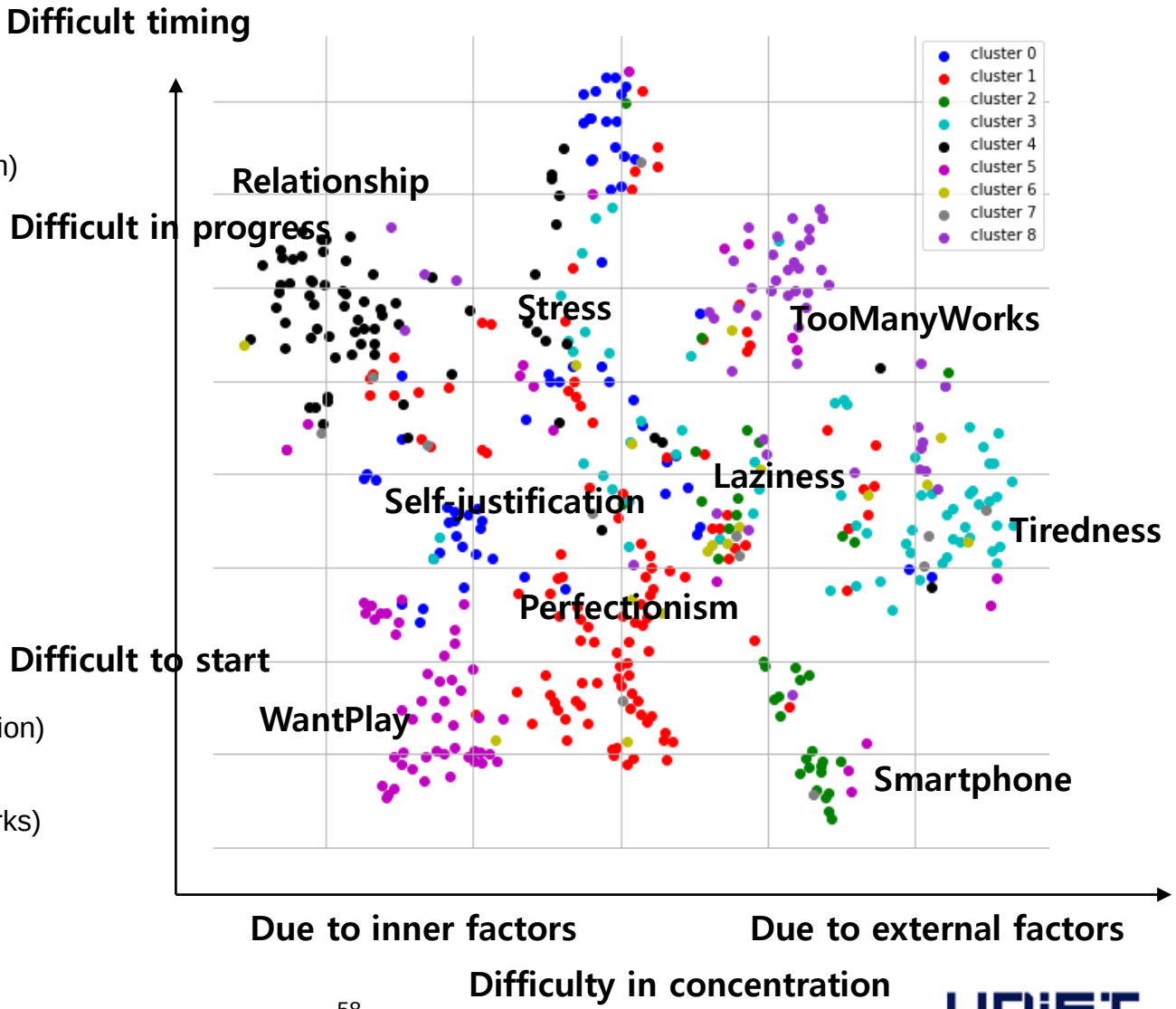
Cluster 4. 인간관계 (Relationship)

Cluster 5. 놀고싶어 (WantPlay)

Cluster 6. 귀차니즘 (Laziness)

Cluster 7. 자기안심 (Self-justification)

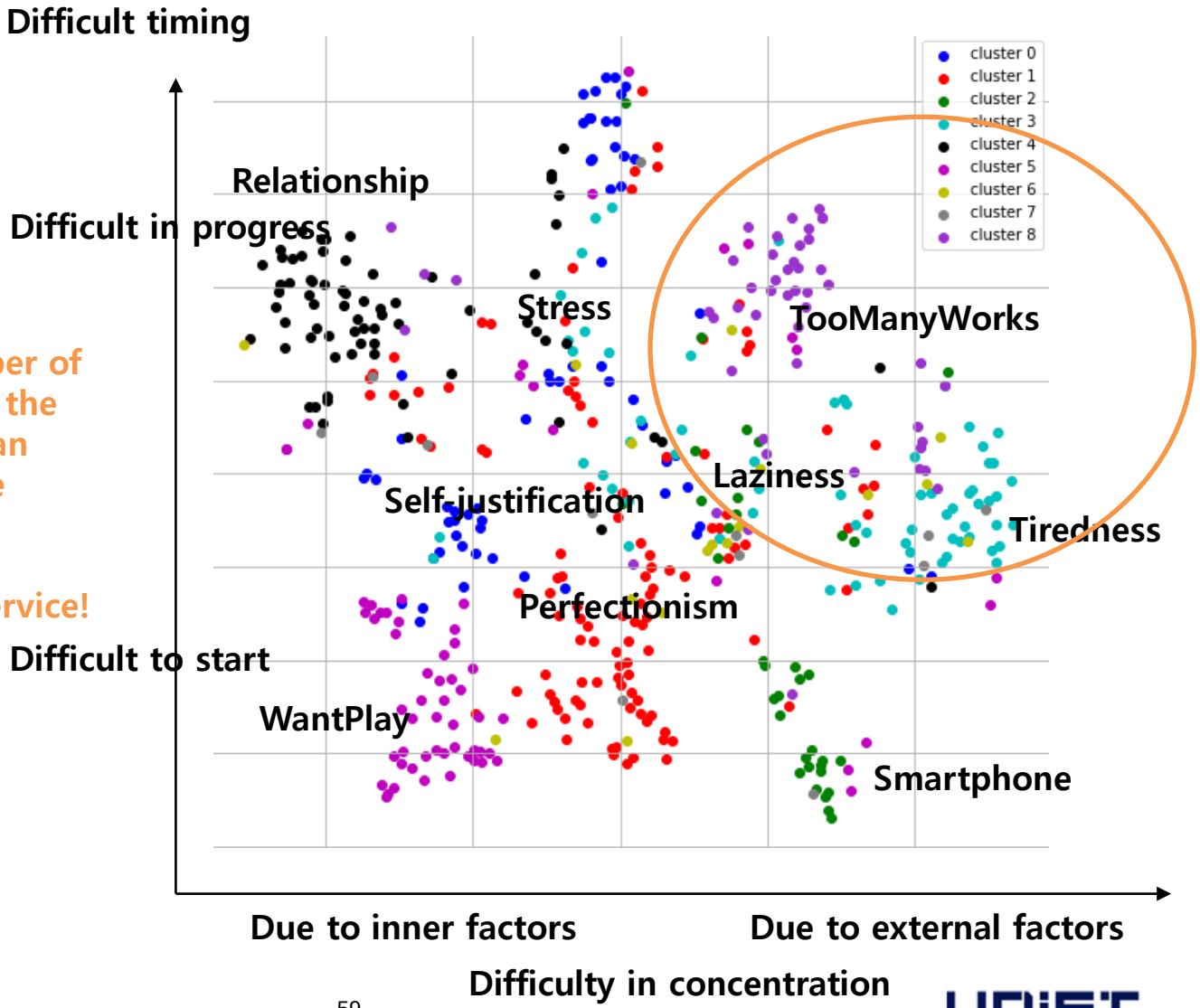
Cluster 8. 일이많음 (TooManyWorks)



Advising Service Customization to Student Segments

I would try to advise the student to reduce the number of his/her tasks by prioritizing the tasks, so that the student can focus fewer items and make some more time to focus.

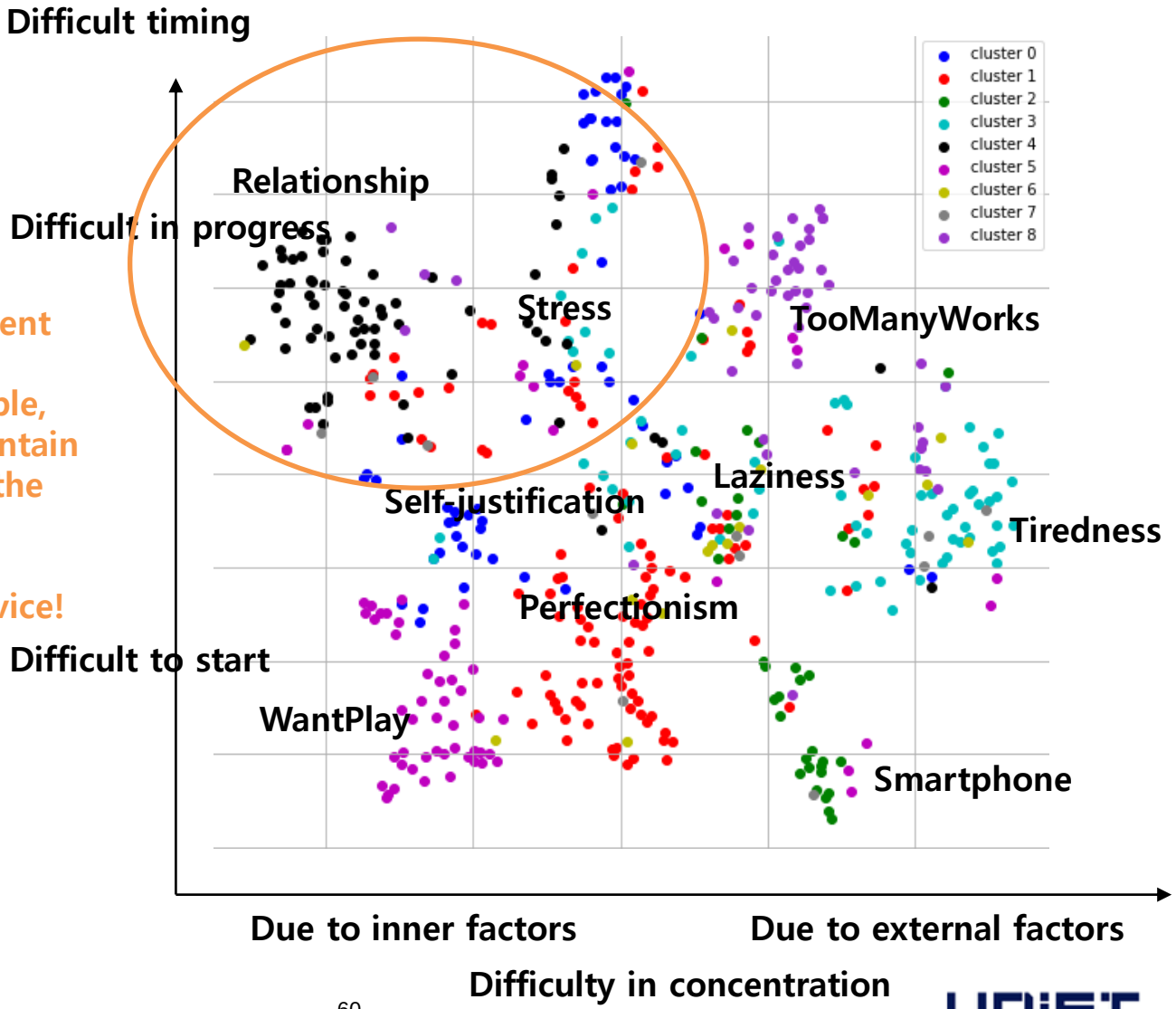
→ Prioritization coaching service!



Advising Service Customization to Student Segments

I would try to help the student become robust to the relationship with other people, so that the student can maintain his/her focus regardless of the relationship status.

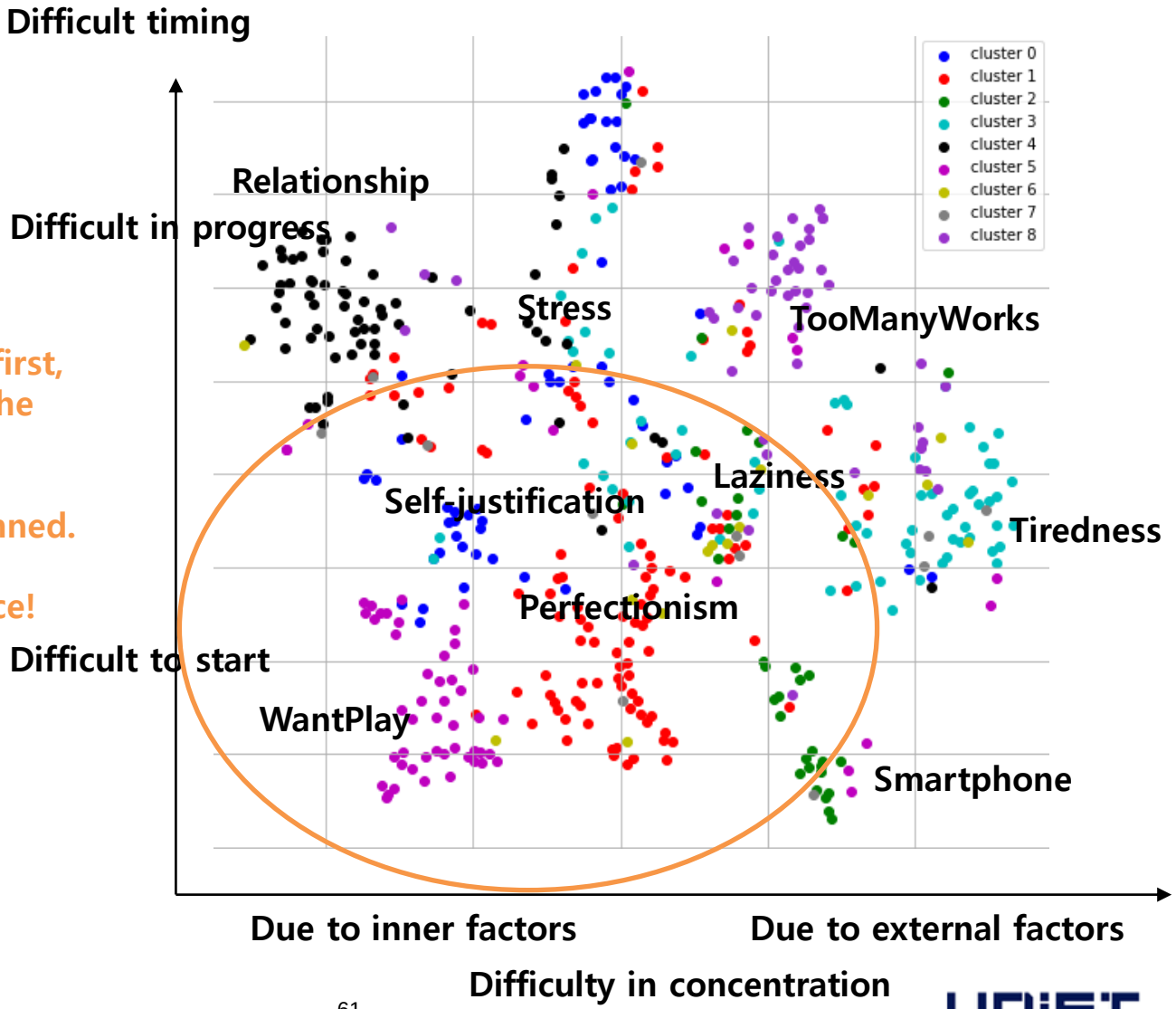
→ Robustness coaching service!



Advising Service Customization to Student Segments

I would try to advise the student to just start a task first, and then think later. Once the task is started, the task completion process can be properly estimated and planned.

→ Just-do-it coaching service!

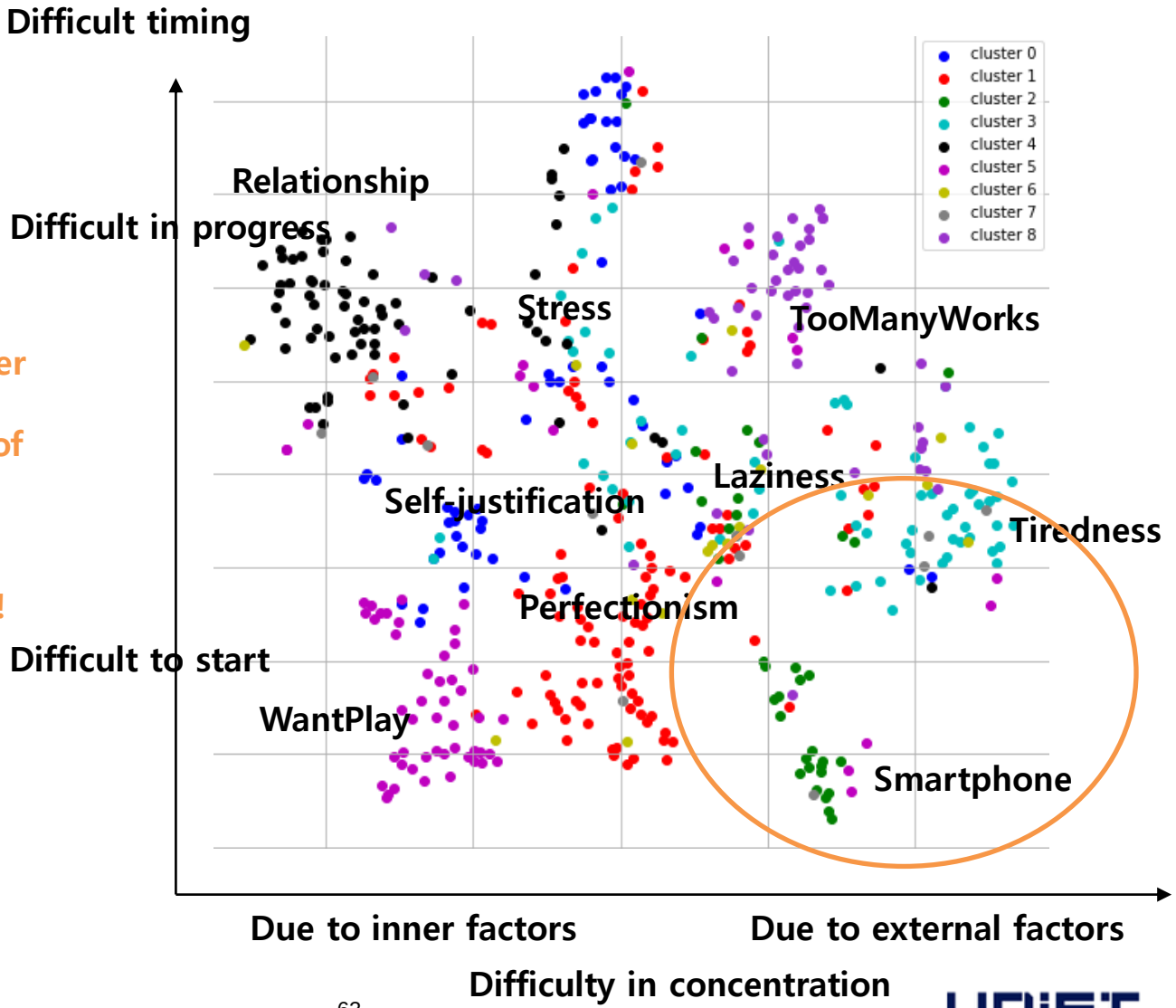


Clustering of Customers: Students Clustering for Counseling

Advising Service Customization to Student Segments

I would try to advise the student to search for another routine to escape from the addiction and habitual use of smartphone or other digital devices/services.

→ Routine coaching service!



The Essence is not Clustering but Segmentation and Customization

- Customer features can be engineered and embedded in various ways
- Similarities among customers can be measured in various ways
- Soft clustering can be used instead of hard clustering
- A sophisticated clustering algorithm may not be necessary in many cases
- Iteration of analyses-interpretations is critical, as we deal with "unsupervised" learning
- One service can be applied for multiple customer segments
- Clustering attempts should focus on understanding customer variabilities

	Feature 1	Feature 2	Feature 3	…	Feature m-1	Feature m
Customer 1	…	…	…	…	…	…
…	…	…	…	…	…	…
Customer n	…	…	…	…	…	…

Assignment 5 (This is a practice with no due date and evaluation)

- (1) Based on the lab demonstration by TA Jung, apply clustering algorithms to the provided dataset of LG Electronics System Air Conditional customers. Evaluate and compare the clustering results (e.g., using the Silhouette coefficient), and select the most accurate and interpretable result of clustering. You should describe the analytics process in detail.
- (2) You may have identified multiple clusters. Focus on the most clear and interesting clusters. Design MORE THAN THREE customized services to address the customer clusters. For example, when there are school-related customer clusters identified, you may want to design an Air Conditioning service specialized to schools. For another example, when there will be a customer cluster with specific maintenance issues frequently occur, you may want to offer an Air Conditioner Maintenance service specialized to this customer. You should describe in detail your services customized for your focal customer segments.
- (3) Assume you actually need to use your outcomes of customer segmentation and service customization for LG Electronics. Starting from your analytics outcomes and designed services, think how to further develop and operate your designed services. What kinds of data are further required to completely develop and operate your services? What kinds of intelligent machines can be further used to improve the efficiency and performance of your services? Describe your service development and implementation plan in detail.
- (4) Select one of the datasets you used or constructed in your previous assignments (e.g., item purchase data and customer review data), and apply clustering algorithms to the dataset. Evaluate and compare the clustering results (e.g., using the Silhouette coefficient), and select the most accurate and interpretable result of clustering. Finally, similar to the above practice, design and describe at least three options of service customization to your customer segments.
- (5) Think about your concerned or interested service around UNIST, in Ulsan, in your hometown, or any other interested service that require further customer segmentation and service customization. If you would actually conduct a study on customer segmentation and service customization, how would you conduct the research in your own creative, unique way? What kinds of data are you going to collect, analyze, and learn? What analytics methods are you going to use? Describe your service customization and development plan in detail. If possible, visualize your plan clearly (e.g., draw an image, construct a mathematical model). To facilitate your thinking, you may want to identify and review a paper or any other reference in the Internet, related to the service you are interested or concerned.
- Upload your code and a several paragraph essay on the tasks (1)~(5) in the Blackboard, IF YOU WANT.

Practice Demonstration by TA Jung:

See the Video Uploaded

Next Classes

- There is no class on September 28 (the school foundation day of UNIST)
- There is no class on October 3 (the national foundation day of Korea)
- See you on October 5 (Wednesday). We will start to discuss customer behavior data use