Online Supplement

Appendix A. Figure

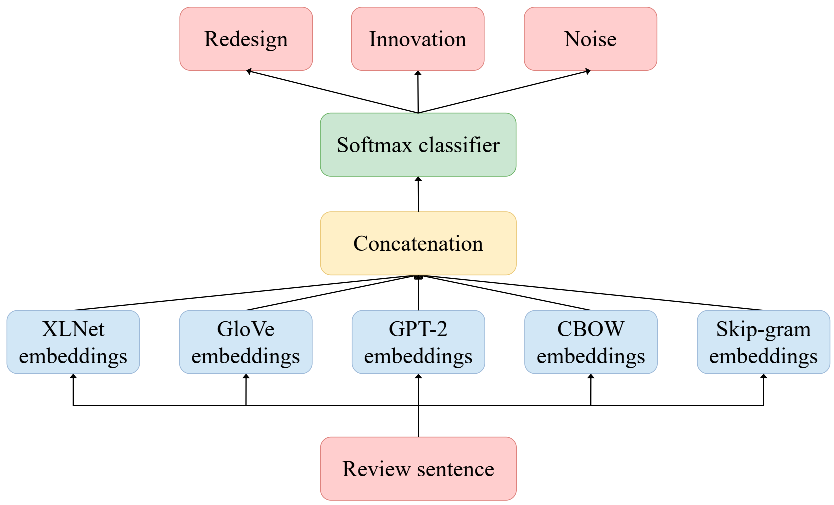


Figure A1 Framework of MEM for product redesign and innovation review classification.

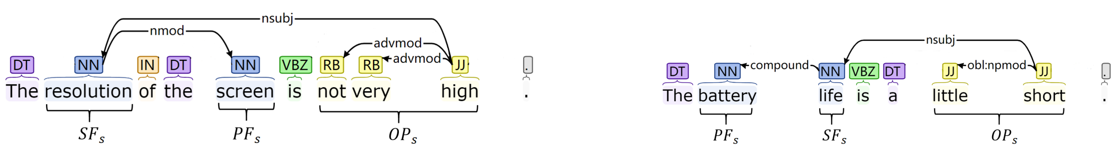


Figure A2 Two rules for extracting seed feature triples.

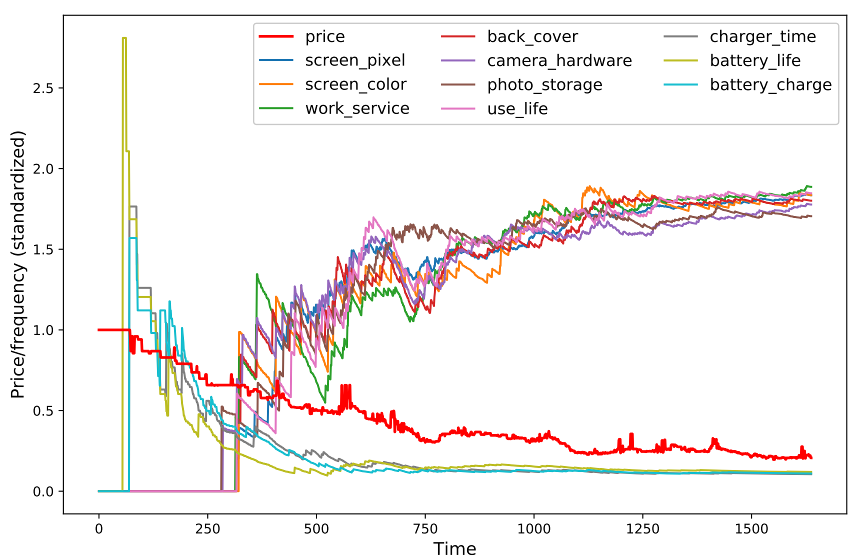


Figure A3 Frequency (TF-IDF) of the most relevant features and price change over time.

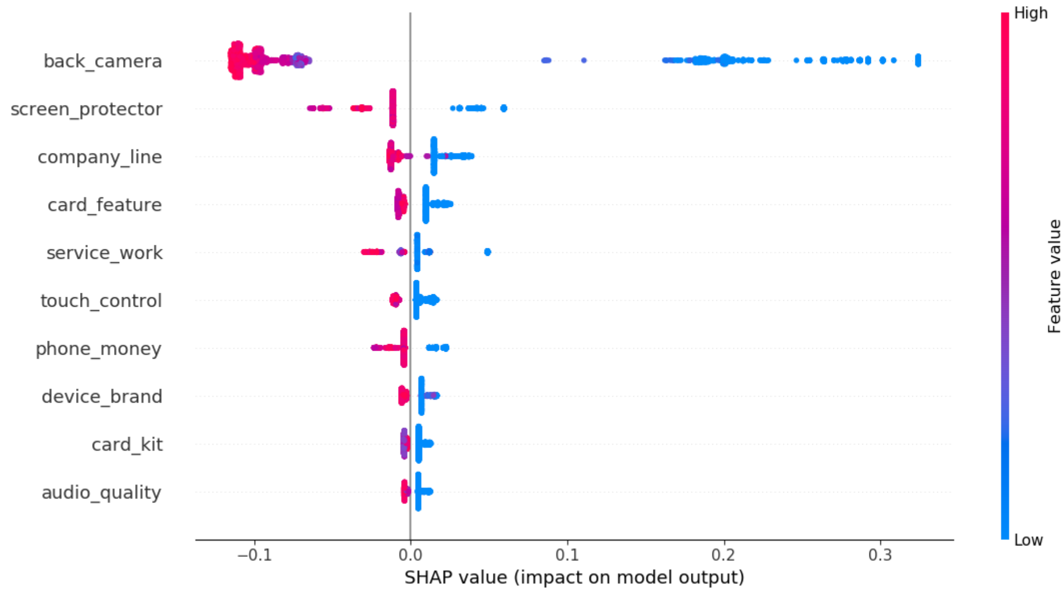
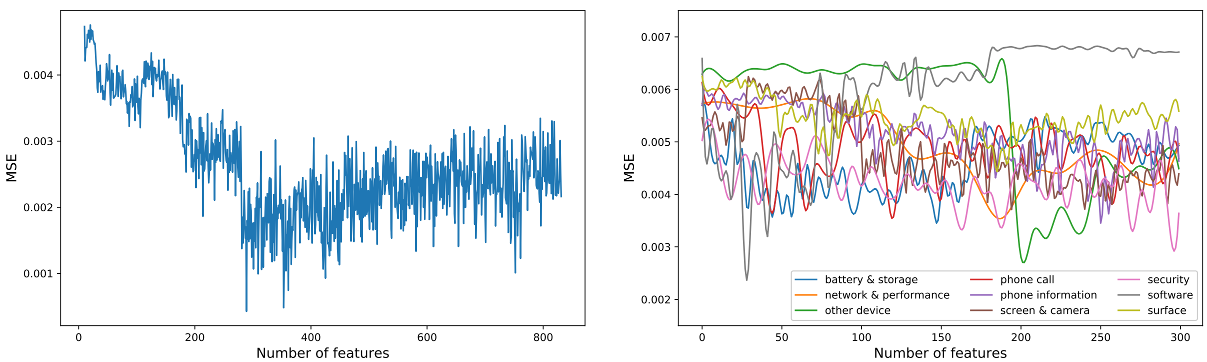
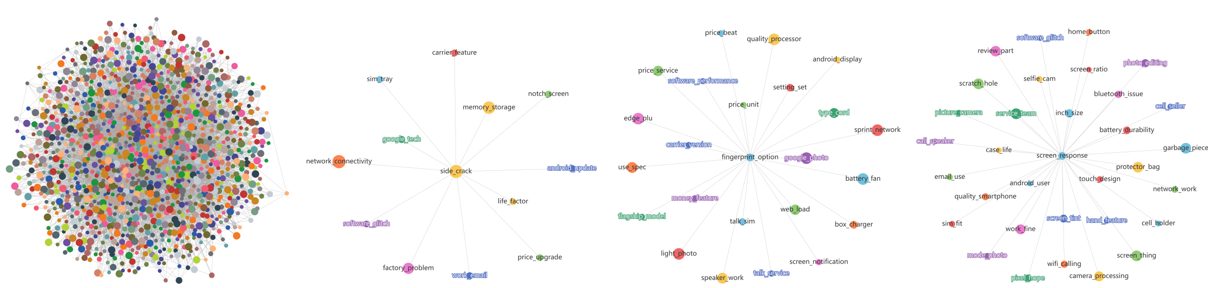


Figure A4 **Distribution of SHAP values for different features.**



(a) MSE vs. the number of all features (b) MSE vs. the number of features in each topic

Figure A5 **Changes of MSE when adding features to the SVR model in sequence.**



(a) All features (b) Centered on "side_crack" (c) Centered on "fingerprint_option" (d) Centered on "screen_response"

Figure A6 **Competitiveness network of all relevant redesign features obtained by the PUBG algorithm.**

Appendix B. Table

Table B1 Fifteen seed feature triples related to redesign with the highest frequencies.

PF_s	SF_s	OP_s	Frequency
battery	life	great	551
display	color	beautiful	380
battery	time	long	158
money	value	great	143
camera	quality	excellent	135
battery	life	amazing	118
sim	card	new	93
battery	life	better	92
customer	service	great	91
money	value	good	88
product	price	great	86
battery	life	terrible	74
buck	bang	best	74
battery	life	awesome	66
battery	life	poor	64

Table B2 Importance, performance, and competitiveness of ten redesign features.

Feature	Importance		Performance		Competitiveness		Category
back_camera	0.1351	High	0.2814	Low	0.0811	Low	$C8$
camera_pixel	0.0313	Low	0.5200	High	0.0214	Low	$C5$
memory_space	0.0237	Low	0.3503	Low	0.1339	High	$C2$
touch_control	0.0127	Low	0.4727	Low	0.1108	Low	$C6$
screen_color	0.0369	Low	0.5427	High	0.1125	Low	$C5$
corner_scratch	0.0463	Low	0.4408	Low	0.1446	High	$C2$
case_color	0.0754	High	0.5760	High	0.1792	High	$C3$
camera_sound	0.0408	Low	0.5062	Low	0.2046	High	$C2$
charger_cord	0.0831	High	0.7942	High	0.1775	High	$C3$
box_right	0.0132	Low	0.6472	High	0.1559	High	$C1$

Table B3 Importance, performance, and competitiveness of ten innovation features.

Feature	Importance		Performance		Competitiveness		Category
back_cover	0.0331	Low	0.8692	High	0.0325	Low	$C5$
touch_screen	0.0495	Low	0.6439	High	0.0832	Low	$C5$
box_cover	0.0937	High	0.3791	Low	0.1462	High	$C4$
speaker_volume	0.0112	Low	0.7236	High	0.1264	High	$C1$
edge_screen	0.0777	High	0.5243	Low	0.2136	High	$C4$
picture_quality	0.1296	High	0.7267	High	0.0654	Low	$C7$
camera_setup	0.0425	Low	0.5146	Low	0.1576	High	$C2$
screen_pixel	0.0371	Low	0.2364	Low	0.1176	High	$C2$
camera_clarity	0.0632	High	0.4147	Low	0.0964	Low	$C8$
battery_quality	0.0829	High	0.7372	High	0.1298	High	$C3$

Table B4 Product improvement strategies with different improvement preference ($K = 10$).

	$w_I = 0.6, w_P = 0.2, w_O = 0.2$	$w_I = 0.2, w_P = 0.6, w_O = 0.2$	$w_I = 0.2, w_P = 0.2, w_O = 0.6$
Redesign features	screen_response	wifi_connection	price_option
	camera_pixel	memory_capacity	heat_dissipation
	screen_image	wifi_signal	charge_time
	charge_time	storage_space	touch_response
	screen_resolution	operating_system	fingerprint_recognition
	sound_quality	sim_card	battery_life
		top_speaker	software_update
		app_support	
Innovation features	back_camera	shape_option	edge_screen
	bluetooth_connection	back_fingerprint	camera_app
	device_weight		charge_wireless
	screen_protector		

Appendix C. Performance evaluation of each phase of the proposed framework

C.1. Performance comparison of product feature extraction

To illustrate the performance and effectiveness of the product feature extraction methods in this study, a lexicon-based, syntax and dependency relation–based, and machine learning–based representative approaches are selected for comparison. The details of the comparison are as follows (comparison results are shown in Table C4).

C.1.1. Lexicon-based approaches. The lexicon-based approach extracts product features based on the frequency of the words and a predefined lexicon (Zhan et al. 2019, Liu et al. 2020, Sridhar et al. 2012, Hu et al. 2020). Hu et al. (2020) extracted nouns and noun phrases from the reviews, and then manually summarized the product features by word frequency. Following Hu et al. (2020)'s method, the extracted features related to product redesign are shown in Table C1. Compared with the MIMFE method, there are two shortcomings of the lexicon-based approach: (1) the opinions corresponding to the product feature words are missing, which makes it difficult to calculate the sentiment scores of the feature words accurately; (2) the subordination relationship between most feature words cannot be determined. For example, "camera" and "app" are two different features in Table C1, but the MIMFE method can determine that "app" is an attribute of "camera" (i.e., "camera_app").

C.1.2. Syntax and dependency relation–based approaches. The syntax and dependency relation–based approach extracts product features based on phrasing and connections relations among words in review text (Archak et al. 2011, Yan et al. 2015, Yang et al. 2022). Representatively, Yang et al. (2022) proposed a novel supervised deep topic modeling approach called sDTM, which leverages the dependency relationship with words in the text to enhance the topic modeling performance. Table C2 shows the product redesign features extracted by sDTM. In Table C2, the product features extracted by sDTM are all one word, which is not conducive to understanding the product features. Therefore, sDTM suffers from the same shortcomings as Section C.1.1.

Table C1 Product redesign features extracted by the approach of Hu et al. (2020).

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7
price point	stars	plenty	imperfections	tech	type	ios
market	seconds	improvement	error	space	terms	earphones
store	weeks	support	awesome	updates	wear	verizon store
purchases	times	quality	lack	tool	version	verizon network
fast shipping	period	great value	crap	service provider	size	battery life
premium	speed	perfect	flaw	storage	things	gsm
dollars	owner	gold	risk	screens	sides	battery capacity
great price	return policy	brightness	disappointment	software	talk	bluetooth
delivery	mother	glad	garbage	screen protector	shape	battery backup
transfer	wife	thanks	nothing	videos	world	wifi
cost	kids	friends	complaints	camera	reviews	app
price	sprint	picture quality	error	stylus	touch	headphones
exchange	party	great product	tmobile	security	signal	battery health
vendor	years	worth	waste	youtube	wrap	wireless
customers	life	gift	damages	screen	setup	battery
shipping	upgrade	screen quality	waste	touchscreen	water	verizon sim card
money	sister	mint condition	fault	speakers	texts	gps
great buy	replacement	health	defects	settings	system	bixby button
sellers	today	recognition	mistake	touch screen	voice	verizon phone
price range	switch	protector	junk	technology	usage	power button

Table C2 Product redesign features extracted by the approach of Yang et al. (2022).

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9
unlocked	price	screen	screen	love	battery	charger	sim	camera
verizon	bought	fingerprint	scratches	amazing	life	box	card	android
network	fast	battery	battery	recommend	happy	charging	note	features
locked	samsung	reader	brand	awesome	excellent	original	verizon	apps
carrier	worth	touch	buying	perfect	condition	charge	easy	quality
sprint	loved	fine	reviews	brand	expected	cable	set	nice
return	wife	recognition	price	absolutely	purchase	fast	unlocked	fast
video	shipping	hand	renewed	highly	day	port	version	pictures
tmobile	money	charge	buy	buy	arrived	headphones	wifi	bixby
store	gift	camera	refurbished	perfectly	time	usb	data	photos
mobile	loves	bad	protector	product	charge	scanner	calling	lot
unlock	husband	music	bought	condition	days	cord	water	size
received	son	time	purchase	quality	perfect	smartphone	transfer	app
buy	range	waste	happy	processor	product	plug	memory	display
connect	specs	feels	condition	issues	us	speed	switch	day
time	daughter	slow	scratch	complaints	hours	the	defective	feature
os	delivery	black	perfect	recommended	pleased	accessories	device	takes
bought	audio	fixed	glass	seller	brand	included	straight	easy
activate	wonderful	worse	recommend	satisfied	advertised	sucks	wireless	updates
item	holding	hold	seller	beautiful	earlier	generic	signal	user

C.1.3. Machine learning–based approaches. Based on the machine learning entity and relationship extraction task, product features and corresponding opinions can be extracted simultaneously

Table C3 Product redesign features extracted by the approach of Zhang et al. (2022b).

Topic 1	Topic 2	Topic 3	Topic 4
camera_good	screen condition_terrible	delivery speed_excellent	night vision_impressive
taking pictures_good	service provider_stupid	max capacity_great	balance_sensational
primary camera_too good	vibrate button_broken	speeds_fantastic	gesture controls_chaotic
camera lens_perfect	left speaker_doesn't work	costumer service_brilliant	maps screen_cluttered
photo lens_clear	speaker_horrible	batter like_excellent	brain power_hyped
cosmetic_blemishes	client_nervous	relation_perfect	night view_amazing
boom service_instant	networks_most difficult	ground_excellent	plastic mess_unbreakable
appearance_brand new	first store_wasn't compatible	privacy settings_great	black screen_greatreal
buttons_work	top reviews_negative	sand_awesome	bargin_huge
camera surround_circular	earphones_don't work	runs_great	picture-takers_heavy
color_really pops	ear speaker_failing	production_great	seniors_outrageous
gift_best	filter_misaligned	charging speed_awesome	stereo audio_amazing
reviews_legit	speaker quality_distorted	funtion_good	list_almost unbelievable
software update_fixed that	front speaker_stopped working	machine_fantastic	turd_complete
functions_work	haptic engine_broke	mictest levels_fine	os_amazing
equipment_unlocked	top speaker_staticky	performer_robust	dog portraits_amazing
system_just crashed	speakers_sound terrible	call screening_fantastic	memory bay_expandable
memory_wiped	maps_don't work	system resources_great	transition_challenging
flipphone_old	face id_quit working	multi tasking_great	websites_heavy
case_undented	bluetooth_didn't work	aspects_great	functionalities_so many

(Zhang et al. 2022b,a). Zhang et al. (2022b) proposed a state-of-the-art feature extraction model based on two-dimensional tables and deep learning, and the product redesign feature extraction results of this model are shown in Table C3. In Table C3, product features and opinions are extracted simultaneously (aspect_opinion), but the subordination between features is not intuitive, although some features contain multiple words. In addition, since the results extracted by this method are similar in form to the feature triples, we also compared the number of extracted product features, and the results are shown in Table C4, which indicates that the MIMFE method extracts more product features, since the potential information hidden in the reviews is considered. In summary, the product feature extraction method proposed in this paper shows the best performance.

Table C4 Performance comparison of different methods for extracting product features.

	Subordination	Fine-grained sentiment scores	Potential information	Number of features extracted
Hu et al. (2020)	✘	✘	✘	—
Yang et al. (2022)	✘	✘	✘	—
Zhang et al. (2022b)	✘	✔	✘	25868
MIMFE	✔	✔	✔	42567

C.2. Experiment on sales volume prediction using product feature attributes

Previous studies (Zhang et al. 2022b, 2021, Archak et al. 2011) have shown that most users browse online reviews before making a purchase, and the product feature attributes mentioned in the reviews directly influence the user's purchase behavior and thus sales (Geva et al. 2017, Ghose et al. 2019). Therefore, a set of useful product attributes should be more helpful in predicting product sales. Based on the above analysis, we designed an experiment to compare the predictive power of product feature attributes proposed by different studies on sales, so as to illustrate the usefulness of the product attributes constructed in this paper. First, we crawled sales and review data of 172 smartphones on amazon.com and classified the phones into three categories according to sales (200–1000,1000–5000,5000+) as labels for the classification task. Next, five studies for comparison are selected and the corresponding product feature attribute values are calculated. Finally, considering the high dimensionality of the feature attributes and small data volume, we choose SVR as the classifier to compare the prediction effect of different product feature attributes on sales volume. The comparison results are shown in Table C5. In Table C5, the product feature attributes constructed in this study outperformed other studies in predicting sales volume, which illustrates the usefulness of importance, performance and competitiveness. In addition, the ablation experiments in Table C5 show that the removal of competitiveness reduces the prediction performance, which indicates the necessity of introducing competitiveness.

Table C5 The predictive effect of product feature attributes of different literature on sales volume.

	Product feature attributes	Macro-precision	Macro-recall	Macro-F1
Bi et al. (2019)	Importance and performance.	0.71	0.81	0.76
Zhang et al. (2021)	Frequency of mentions and feature performance impact on consumer satisfaction.	0.65	0.74	0.70
Hu et al. (2020)	Impact, performance and impact-asymmetry.	0.75	0.82	0.78
Archak et al. (2011)	Frequency and performance.	0.70	0.76	0.73
Zhang et al. (2022b)	Main effects, interaction effects and performance.	0.75	0.80	0.77
This paper	Importance, performance and competitiveness.	**0.78**	**0.85**	**0.82**
	Importance and performance.	0.75	0.79	0.77

C.3. Tracking study

This study intends to verify the validity of the model by tracking the improvement of the next generation of new products in the three target products (Apple iPhone 8, NIKE Tanjun Sneakers and Canon EOS 200D). For Apple iPhone 8 and Canon EOS 200D, we select 8 and 9 new products of the same series for comparison, where the new products of Apple iPhone 8 include Apple X, Apple XR, Apple XS, etc. and Canon EOS 200D includes EOS M200, EOS Rebel SL3, EOS 250D, etc. Then, we used the product comparison functions of versus.com and zol.com (well-known electronic product review websites in the America and China) to determine the redesign and innovative features of different new products for the target products. For the NIKE Tanjun Sneakers, we select 13 new products of the same series for comparison (including Nike Air Max 270, Nike Tanjun Racer, Nike Tanjun Chukka, etc.) and determined the redesign and innovative features of the new product by manually comparing the product brochures. To evaluate the effectiveness of different NPD frameworks, we defined the improvement ratio (IRA) as a metric to evaluate the degree of new product improvement, as shown in Equation (A1).

$$IRA = \frac{\text{Number of product features actually improved}}{\text{Number of product features recommended for improvement}} \times 100\% \tag{A1}$$

Table C6 shows the improvement of some new generation Apple phones on different product features ($K = 10$) and IRA. We assume that the right product improvements will lead to an increase in sales.

Therefore, when choosing the comparator, low-volume products are removed and IRA can be used to roughly measure the effectiveness of different product improvement strategies. The IRA performance of different methods on each dataset is shown in Table C7, in which the framework of this paper outperforms the other methods in most cases.

Table C6 Improvements of some new generation products for Apple iPhone 8 in different product features.

	Apple 8 plus	Apple X	Apple XR	Apple XS	Apple 11
price_option	✘	✘	✘	✔	✔
heat_dissipation	✘	✘	✔	✔	✔
charge_time	✘	✘	✘	✘	✔
touch_response	✘	✘	✘	✔	✔
fingerprint_recognition	✘	✔	✔	✔	✔
battery_life	✔	✔	✔	✔	✔
software_update	✘	✔	✔	✔	✔
edge_screen	✘	✘	✘	✘	✘
camera_app	✘	✔	✔	✔	✔
charge_wireless	✘	✘	✔	✔	✔
IRA	10%	40%	60%	80%	90%

Table C7 Effectiveness (average IRA) comparison of different NPD frameworks.

		Apple iPhone 8	NIKE Tanjun Sneakers	Canon EOS 200D
Zhang et al. (2021)	$K = 10$	63.8%	51.5%	57.8%
	$K = 15$	59.2%	46.2%	52.6%
	$K = 20$	**56.9%**	44.6%	44.4%
Zhang et al. (2019)	$K = 10$	46.3%	51.5%	61.1%
	$K = 15$	49.2%	44.1%	51.1%
	$K = 20$	41.9%	40.8%	51.7%
Zhang et al. (2022b)	$K = 10$	63.8%	46.9%	63.3%
	$K = 15$	51.7%	47.7%	57.0%
	$K = 20$	47.5%	40.4%	**56.1%**
This paper	$K = 10$	**71.3%**	**60.0%**	**68.9%**
	$K = 15$	**61.7%**	**52.8%**	**61.5%**
	$K = 20$	54.4%	**46.9%**	55.0%

Appendix D. Additional case studies

Nelson (1970) categorized products as search products and experience products. For search products, customers can easily understand their quality through online search (e.g., electronics), while experience products have attributes that are difficult to assess before purchase (e.g., clothes), and smartphones are a typical search product. In addition to product type, different e-commerce platforms also affect the effect of NPD. Therefore, two additional datasets are constructed in this study to illustrate the generalizability of the proposed framework: (1) shoe dataset from amazon.com (experience products); and (2) camera dataset from jingdong.com (search products).

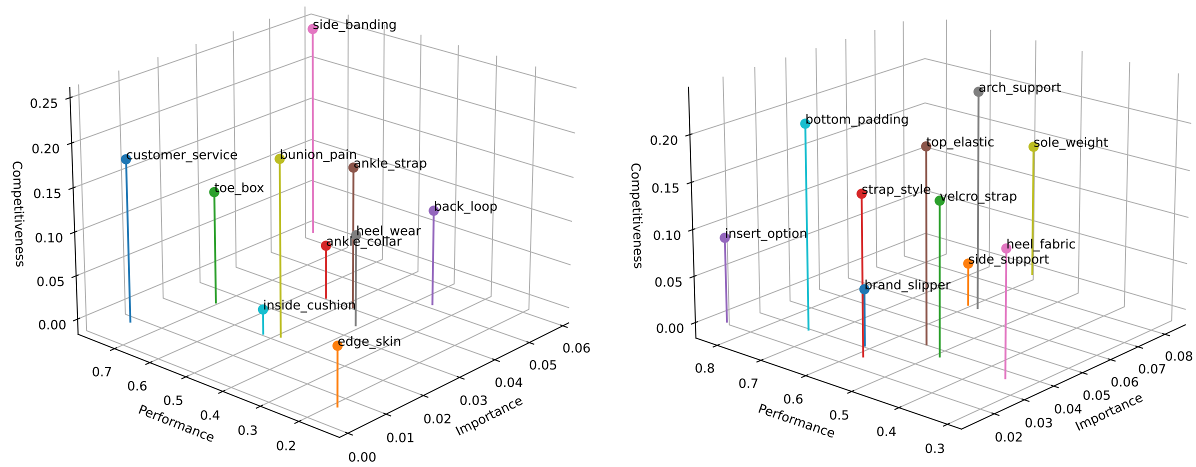
D.1. Case study of shoe dataset

As of April 2023, we have collected 98,884 online reviews about running shoes. After eliminating some invalid data, a total of 84,675 reviews were retained for analysis. Some of the redesign features extracted through the process of Section 3.1 and 3.2 are shown in Table D1. We obtained eleven topics of product redesign features as follows: marketing, components, insole & sole, size & weight, performance, comfort, material & quality, scene, style, price and service (because of space constraints, only the first five are shown in Table D1). As a result, we extracted 201 primary features and 1468 secondary features (i.e., 1468 product redesign features). Similarly, for reviews related to innovation, 78 primary features and 657 secondary features (i.e., 657 product innovation features) were extracted.

Taking the product NIKE Tanjun Sneakers as an example, we determined 545 redesign and 122 innovation features, respectively. Then, similar to Section 4.2, we obtained the importance, performance, and competitiveness of the product redesign and innovation features. Figure D1 shows the IPCA plot for the 10 redesign and innovation features, and Table D2 and Table D3 list the location of each redesign and innovation feature. Furthermore, to validate the effectiveness of the MSCS algorithm, we compared the performance of the different algorithms on the shoe dataset (the parameters remain the same as in Section 4.5). The specific comparison results shown in Figure D2 demonstrate the performance advantage of the MSCS algorithm. Finally, the product features to be improved can be obtained by the MSCS algorithm, and then according to the category (redesign or innovation), specific product improvement strategies can be developed as shown in Table D4.

Table D1 Product redesign features and corresponding topics extracted from the shoe dataset.

Topic	marketing	components	insole & sole	size & weight	performance
Product features	brand_purchase	toe_box	cushion_inside	width_foot	balance_foam
	brand_name	heel_room	bottom_glue	size_fit	balance_fit
	activity_support	toe_side	sole_material	wide_fit	body_support
	model_version	heel_material	cushion_feel	size_foot	body_workout
	design_pair	ankle_support	arch_support	weight_sock	stability_feeling
	show_lace	tongue_loop	bottom_grip	size_option	super_comfortable
	show_quality	string_lace	sole_rubber	area_padding	injury_run
	design_part	inside_padding	arch_runner	length_support	impact_work
	activity_place	knee_surgery	cohesion_style	width_snug	impact_wear
	venture_support	bunion_box	sole_support	size_box	speed_lace
Number of PFs	23	26	19	8	10
Number of SFs	115	234	148	84	61



(a) Product redesign (b) Product innovation

Figure D1 IPCA plot of product (a) redesign and (b) innovation features (shoe dataset).

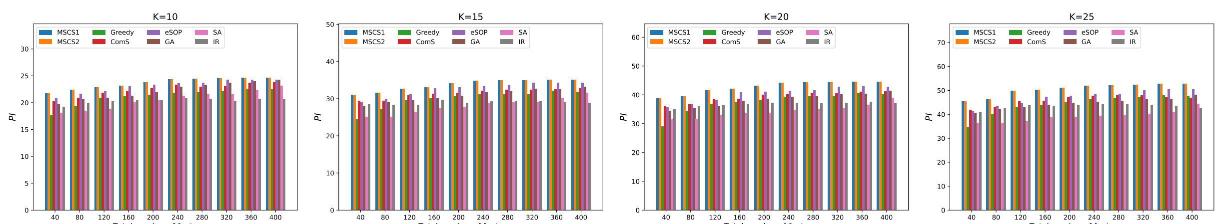


Figure D2 Effect comparison of different feature selection algorithm (shoe dataset).

Table D2 Importance, performance, and competitiveness of ten redesign features (shoe dataset).

Feature	Importance		Performance		Competitiveness		Category
customer_service	0.0034	Low	0.7163	High	0.1846	High	$C1$
edge_skin	0.0019	Low	0.1418	Low	0.0665	Low	$C6$
toe_box	0.0201	Low	0.6540	High	0.1279	Low	$C5$
ankle_collar	0.0353	High	0.5015	High	0.0610	Low	$C7$
back_loop	0.0462	High	0.3169	Low	0.1084	Low	$C8$
ankle_strap	0.0353	High	0.4278	Low	0.1608	High	$C4$
side_banding	0.0577	High	0.7602	High	0.2435	High	$C3$
heel_wear	0.0296	High	0.3646	Low	0.1031	Low	$C8$
bunion_pain	0.0168	Low	0.4394	Low	0.1998	High	$C2$
inside_cushion	0.0157	Low	0.4756	Low	0.0279	Low	$C6$

D.2. Case study of camera dataset

We collected 93,124 online reviews about digital cameras in Chinese from Jingdong.com, which were posted between January 2019 and May 2023. After translating them into English and eliminating some invalid data, a total of 81,276 reviews are retained for analysis. Some of the extracted product redesign features are shown in Table D5. We obtained ten topics of product redesign features as follows: screen & photography, service, control & memory, marketing, design & material, battery & portability, lens &

Table D3 Importance, performance, and competitiveness of ten innovation features (shoe dataset).

Feature	Importance		Performance		Competitiveness		Category
brand_slipper	0.0239	Low	0.5688	High	0.0599	Low	$C5$
side_support	0.0577	High	0.5611	High	0.0452	Low	$C7$
velcro_strap	0.0298	Low	0.4461	Low	0.1623	High	$C2$
strap_style	0.0189	Low	0.5393	Low	0.1694	High	$C2$
insert_option	0.0143	Low	0.8130	High	0.0904	Low	$C5$
top_elastic	0.0332	Low	0.4965	Low	0.2071	High	$C2$
heel_fabric	0.0296	Low	0.3083	Low	0.1335	Low	$C6$
arch_support	0.0575	High	0.5390	Low	0.2321	High	$C4$
sole_weight	0.0827	High	0.5727	High	0.1417	Low	$C7$
bottom_padding	0.0227	Low	0.6861	High	0.2176	High	$C1$

Table D4 Product improvement strategies (shoe dataset).

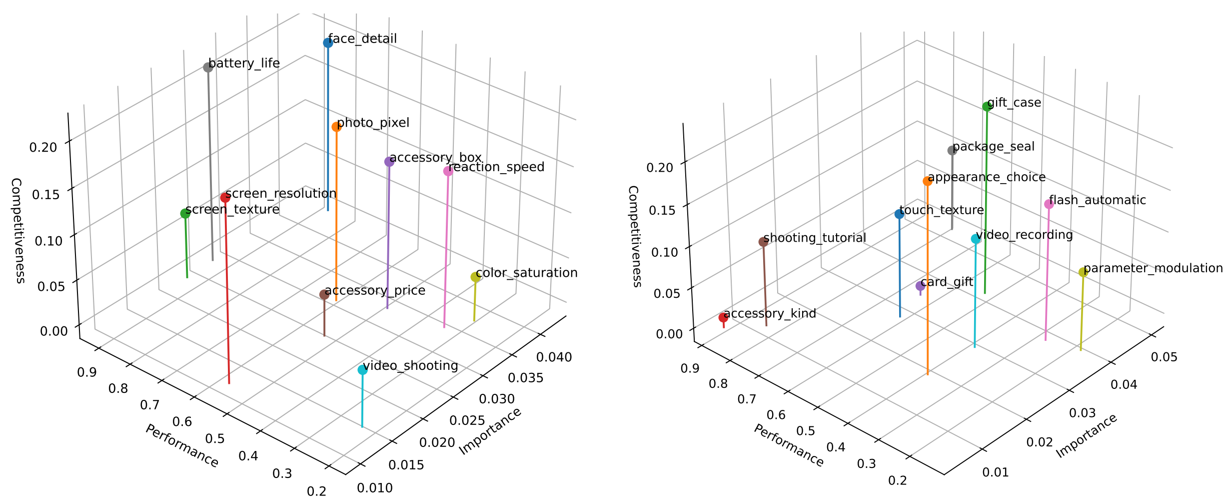
	$K = 10$	$K = 15$		$K = 20$	
Redesign features	back_loop	back_loop	color_choice	back_loop	toe_blow
	ankle_strap	ankle_strap	rock_walking	ankle_strap	price_option
	heel_wear	pair_fit	quality_stitching	slip_resistance	toe_touch
	arch_support	bottom_grip	rubber_bouncy	color_choice	rock_walking
	bottom_grip	plastic_odor	side_wear	support_water	sock_fit
	color_choice	color_choice		pair_fit	weight_distribution
	distance_hike	foot_pain		toe_pain	side_wear
	foot_pain	price_option		bottom_grip	sole_weight
	foot_side	quality_control		rubber_bouncy	strap_style
Innovation features	bottom_padding	bottom_padding		style_pair	insole_padding

shutter, overall condition, usage scenarios and software (because of space constraints, only the first four are shown in Table D5). As a result, we extracted 212 primary features and 1879 secondary features (i.e., 1879 product redesign features). Similarly, for reviews related to innovation, 83 primary features and 753 secondary features (i.e., 753 product innovation features) were extracted.

Taking the product Canon EOS 200D as an example, we determined 403 redesign and 158 innovation features, respectively. After obtaining the importance, performance and competitiveness of product redesign and innovation features, we construct IPCA plot of 10 redesign and innovation features in Figure D3 and list their locations in Table D6 and Table D7. In addition, we compare the performance of different algorithms on the camera dataset, as shown in Figure D4, illustrating the performance advantages of the MSCS algorithm. Finally, the product improvement strategies obtained by the MSCS algorithm are shown in Table D8.

Table D5 Product redesign features and corresponding topics extracted from the camera dataset.

Topic	control & memory	design & material	screen & photography	usage scenarios
Product features	reaction_speed	metal_texture	image_color	night_scene
	operation_speed	packaging_seal	screen_quality	cat_photo
	parameter_setting	design_style	pixel_size	street_portrait
	sensitivity_range	box_seal	photo_detail	scenery_shooting
	operation_button	custom_button	screen_side	night_exposure
	memory_card	packaging_design	video_production	gift_accessory
	touch_focu	bag_quality	portrait_photography	streaming_effect
	operating_menu	shoulder_strap	head_aperture	phone_transmission
	autofocu_effect	hand_grip	skin_color	usage_instruction
	shake_control	box_cover	photo_texture	scene_setting
Number of PFs	30	35	24	39
Number of SFs	263	247	317	345



(a) Product redesign (b) Product innovation

Figure D3 IPCA plot of product (a) redesign and (b) innovation features (camera dataset).

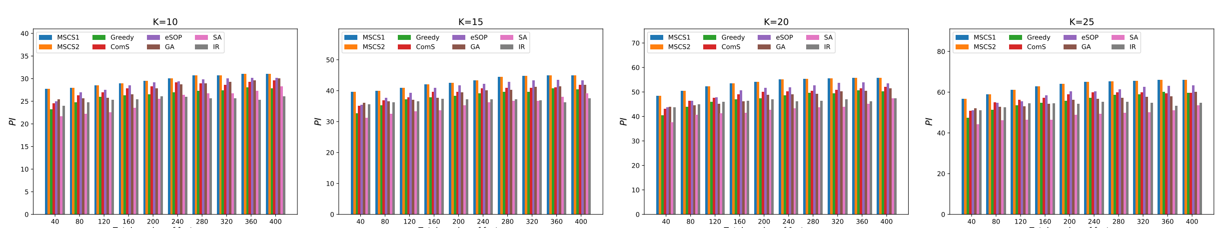


Figure D4 Effect comparison of different feature selection algorithm (camera dataset).

Table D6 Importance, performance, and competitiveness of ten redesign features (camera dataset).

Feature	Importance		Performance		Competitiveness		Category
face_detail	0.0422	High	0.8444	High	0.1895	High	$C3$
photo_pixel	0.0289	High	0.5714	High	0.1905	High	$C3$
screen_texture	0.0219	Low	0.9056	High	0.0722	Low	$C5$
screen_resolution	0.0100	Low	0.5391	Low	0.1984	High	$C2$
accessory_box	0.0313	High	0.4586	Low	0.1614	High	$C4$
accessory_price	0.0230	Low	0.4975	Low	0.0459	Low	$C6$
reaction_speed	0.0324	High	0.3078	Low	0.1707	High	$C4$
battery_life	0.0262	Low	0.9098	High	0.2135	High	$C1$
color_saturation	0.0354	High	0.2715	Low	0.0489	Low	$C8$
video_shooting	0.0131	Low	0.2037	Low	0.0619	Low	$C6$

Table D7 Importance, performance, and competitiveness of ten redesign features (camera dataset).

Feature	Importance		Performance		Competitiveness		Category
touch_texture	0.0252	Low	0.5779	High	0.1254	High	$C1$
appearance_choice	0.0163	Low	0.3563	Low	0.2277	High	$C2$
gift_case	0.0390	High	0.4899	Low	0.2258	High	$C4$
accessory_kind	0.0049	Low	0.8733	High	0.0121	Low	$C5$
card_gift	0.0319	High	0.6067	High	0.0111	Low	$C7$
shooting_tutorial	0.0096	Low	0.8001	High	0.1023	Low	$C5$
flash_automatic	0.0352	High	0.2382	Low	0.1638	High	$C4$
package_seal	0.0498	High	0.7565	High	0.0995	Low	$C7$
parameter_modulation	0.0366	High	0.1448	Low	0.0948	Low	$C8$
video_recording	0.0265	Low	0.3487	Low	0.1309	High	$C2$

Table D8 Product improvement strategies (camera dataset).

	$K = 10$	$K = 15$	$K = 20$	
Redesign features	image_quality	image_quality	image_quality	phone_transmission
	imaging_datum	face_detail	screen_operation	bluetooth_transmission
	face_detail	screen_operation	logistic_speed	beginner_choice
	screen_operation	logistic_speed	response_time	
	photography_manual	user_interface	shake_performance	
	portrait_shooting	accessory_price	color_saturation	
	service_attitude	response_time	delivery_package	
	battery_life	post_production	battery_life	
		operation_interface	button_sound	
		shake_performance	expres_delivery	
		appearance_design	shooting_autofocu	
		color_saturation	telephoto_angle	
Innovation features	video_recording	operation_manual	box_seal	eye_recognition
	accessory_box	video_recording	operation_manual	card_gift
		card_gift	beginner_course	

Appendix E. Comparison of all embedding methods

Word2Vec and GloVe are the most popular traditional embedding methods, however there are various contextual embedding methods that perform well, here we choose the six that work best, namely GPT-2, XLNet, BERT, RoBERTa, ALBERT and T5. RoBERTa is an improvement of BERT model, which uses more data and longer training time to pre-train with a more optimized training strategy to improve the performance. ALBERT is a lightweight version of BERT that reduces the number of model parameters and improves the efficiency and performance of the model through techniques such as parameter sharing and embedding layer parameter decomposition. T5 is a Transformer-based text generation model that can be applied not only to various downstream tasks (e.g., text classification, named entity recognition, etc.), but also directly to text generation tasks (e.g., translation, summarization, etc.). The dimensions and performance of the different embedding methods on the classification tasks of product redesign- and innovation-related reviews are shown in Table E1. The best performing contextual embedding methods are GPT-2 and XLNet, therefore, GPT-2 and XLNet are chosen as the base methods for MEM, in addition to the traditional embedding CBOW, Skip-gram and GloVe. It is worth noting that these five embedding methods are chosen to obtain better performance, but due to the limitation of the number of combinations and dimensions, whether other embedding combinations can achieve the best performance remains to be further explored in future studies.

Table E1 Dimensions and performance of the different embedding methods.

	Model	Dimensions	AUC	Precision	Recall	F1-value
Traditional embedding	CBOW	100	0.67	0.68	0.69	0.68
	Skip-gram	100	0.69	0.69	0.72	0.70
	GloVe	100	0.70	0.69	0.68	0.69
Contextual embedding	GPT-2	768	**0.73**	**0.74**	**0.76**	**0.75**
	XLNet	768	0.71	**0.72**	**0.73**	**0.73**
	BERT	768	0.70	0.69	0.72	0.71
	RoBERTa	768	**0.72**	0.71	0.72	0.72
	ALBERT	128	0.68	0.69	0.70	0.69
	T5	512	0.71	0.71	0.68	0.70

Appendix F. Proof of Theorem 1

Proof. Let $I = \{e_1, e_2, ..., e_m\}$ denote the set of elements, $\{I_1, I_2, ..., I_n\}$ denote the collection of subsets (some elements of I are contained in I_n), and integer K denote the number of subsets selected. The objective of solving the MC problem is to select K subsets to maximize the number of elements contained in these subsets. We can transform any instance of MC into a specific instance of MPI. First, elements and subsets can be represented as product features and competitive feature sets, respectively. Let ARF_k denote a subset of AR, and the competitiveness of TB_k with features in ARF_k ranks in the top ten (i.e., top ten competitive features of TB_k). In this case, each subset I_k is represented by a competitive features set ARF_k corresponding to feature TB_k. For a product feature, if it is repeated in TB, we remove it to ensure that no duplicate elements are in TB, i.e., $\forall TB \in TR$, $|TB|_f = K$. Then, we set $w_I = w_P = 0$ and $w_O = 1$ and suppose that the competitiveness of feature in $LF^K_{q_A}$ to AR_{q_A} is w, i.e., $Com(TB_i, AR_{q_A}) = w$ $(TB_i \in LF^K_{q_A}, q_A = 1, 2, ..., Q_A)$. As a result, we obtain that $OC^K_{q_A} = w$ if $|LF^K_{q_A}| > 0$ and $OC^K_{q_A} = 0$ if $|LF^K_{q_A}| = 0$. Therefore, we can design an instance of MPI. This instance can clearly be constructed in polynomial time.

In this particular instance of MPI, we have $PI^K = OC^K = \sum_{q_A=1}^{Q_A} OC^K_{q_A}$. Thus, the objective is to maximize the number of features whose competitiveness to feature in TB ranks in the top ten, i.e., maximize the number of elements in set $\{ARF_1 \cup ARF_2 \cup ... \cup ARF_K\}$. Therefore, both instances have the same objective, which is to select K competitive features set (subsets) to maximize the number of features (elements) contained in them. Thus, a direct correspondence is present between the solution of the MPI instance and the solution of the MC instance, which is clearly a simple transformation in polynomial time.

In sum, MC is reducible to MPI. MPI problem is NP-hard. □

Appendix G. Pseudocode of MSCS algorithm

Algorithm G1 Multistage combined search algorithm (MSCS).

Input: Redesign and innovation features set of all product AR, redesign and innovation features set of product to be improved TR, the number of features to be improved K, initial search breadth λ, optimized search breadth γ, search depth θ.

Output: Integer solution vector $x_{optimal}$.

function MSCS$(AR, TR, K, \lambda, \gamma, \theta)$
    $S_{initial}, S_{optimal} \leftarrow$ SpaceGenerator$(AR, TR, K, \lambda, \gamma)$
    $x_{initial} \leftarrow$ GreedySolver$(AR, TR, K, S_{initial})$
    $x_{optimal} \leftarrow x_{initial}$
    for $\theta' = 1$ to θ **do**
        $x_{last} \leftarrow \mathbf{0}$
        while $x_{optimal} \neq x_{last}$ **do**
            $x_{last} \leftarrow x_{optimal}$
            $x_{optimal} \leftarrow$ FindOptimal$(AR, TR, x_{last}, S_{optimal}, \theta')$
        end while
    end for
    return $x_{optimal}$
end function
function SpaceGenerator$(AR, TR, K, \lambda, \gamma)$
    $x_{continuous} \leftarrow$ InitialSolver(AR, TR)
    $x_{sort} \leftarrow$ Sort$(x_{continuous})$ ▷ The function of Sort(x) is to arrange the elements in x in descending order
    $S_{initial}, S_{optimal} \leftarrow \emptyset$
    $T_{initial} \leftarrow Max(\lceil \lambda Q_T \rceil, K)$
    $T_{optimal} \leftarrow T_{initial} + \lceil \gamma(Q_T - T_{initial}) \rceil$
    for $i = 1$ to $T_{initial}$ **do**
        $S_{initial} \leftarrow S_{initial} \cup$ Index$(x_{continuous}, x^i_{sort})$ ▷ x^i denote the ith element in vector x, Index(x, e) denote the position of element e in vector x
    end for
    for $i = T_{initial} + 1$ to $T_{optimal}$ **do**
        $S_{optimal} \leftarrow S_{optimal} \cup$ Index$(x_{continuous}, x^i_{sort})$
    end for
    return $S_{initial}, S_{optimal}$
end function
function GreedySolver$(AR, TR, K, S_{initial})$
    $x_{initial} \leftarrow \mathbf{0}$
    for $k = 1$ to K **do**
        $C \leftarrow \emptyset$
        for $i \in S_{initial}$ **do**
            $x_{temporary} \leftarrow x_{initial}$

        if $x^i_{temporary} = 0$ **then**
            $x^i_{temporary} \leftarrow 1$
            $C \leftarrow C \cup x_{temporary}$
        end if
    end for
    $x_{initial} \leftarrow \text{argmax}_{x \in C} PI^K_{AR,TR}(x)$
end for
return $x_{initial}$
end function
function FINDOPTIMAL($AR, TR, x_{last}, S_{optimal}, \theta'$)
$S_1 \leftarrow \{i | x^i_{last} = 1\}$
$CB_1 \leftarrow$ COMBINATION(S_1, θ') ▷ COMBINATION(S, θ) denote the set of all combinations of θ elements in S
$CB_{optimal} \leftarrow$ COMBINATION($S_{optimal}, \theta'$)
$CB_1 \leftarrow$ COMBINATION(S_1, θ')
$C \leftarrow \emptyset$
for $B_1 \in CB_1$ **do**
    $x_{temporary} \leftarrow x_{last}$
    for $i \in B_1$ **do**
        $x^i_{temporary} \leftarrow 0$
    end for
    for $B_{optimal} \in CB_{optimal}$ **do**
        $x'_{temporary} \leftarrow x_{temporary}$
        for $j \in B_{optimal}$ **do**
            $x'^j_{temporary} \leftarrow 1$
        end for
        if $|x'_{temporary}|_f = K$ **then**
            $C \leftarrow C \cup x'_{temporary}$
        end if
    end for
end for
$x_{optimal} \leftarrow \text{argmax}_{x \in C} PI^K_{AR,TR}(x)$
return $x_{optimal}$
end function

Appendix H. Consider production constraints

Following Zhang et al. (2019), engineering budget, lead time, and technical risk were determined to be the main constraints. The engineering budget indicates the investment in product redesign and innovation. Let EB_k denote the cost of redesigning or innovating feature TB_k and MLT_K represent the maximum lead time of the feature in TB. Technical risk is defined as quality fluctuations due to improvements, including changes in reliability, durability, and serviceability, which can be obtained through evaluation by design engineers. Let CR_k, CD_k, and CS_k denote the reliability, durability, and

Algorithm G2 Personalized unidirectional bipartite graph algorithm (PUBG).

Input: Transition possibility matrix PM, average performance of features $\overline{Per} = \{\overline{Per}_1, \overline{Per}_2, ..., \overline{Per}_Q\}$, damping factor α.

Output: Competitiveness of all features $\overline{C}$.

function PUBG($PM, \overline{Per}, \alpha$)
    $C \leftarrow \emptyset$
    for $q = 1$ to Q **do** $\triangleright$ select features to compare competitiveness
        $d_q \leftarrow (0, 0, ..., 0)$ $\triangleright$ initialize the personalized vector
        for $i = 1$ to $P + Q$ **do**
            if $i \leqslant P$ **then**
                $d_q^i \leftarrow 1$
            else if $i = P + q$ **then**
                $d_q^i \leftarrow \overline{Per_q}$
            end if
        end for
        $C_q \leftarrow$ RW(PM, d_q, α)
        $C \leftarrow C \cup \{C_q\}$
    end for
    $\overline{C} = \text{mean}(C)$ $\triangleright$ calculate the final competitiveness vector by averaging
    return $\overline{C}$, C $\triangleright$ return the competitiveness and competitiveness network
end function
function RW(PM, d_q, α)
    $C_q \leftarrow \{\frac{1}{P+Q}, \frac{1}{P+Q}, ..., \frac{1}{P+Q}\}$ $\triangleright$ initialize the competitiveness vector of F_q
    repeat
        $C_q^* \leftarrow C_q$
        $C_q \leftarrow \alpha \times PM \times C_q + (1 - \alpha) \times d_q$ $\triangleright$ random walk
    until $|C_q - C_q^*| < \varepsilon$ $\triangleright$ judge whether the stable state is reached
    return C_q
end function

serviceability change of feature TB_k, respectively. In addition, for complex products such as smartphone and cameras, where improvements to different product components and subsystems are relevant (i.e., product structure matching (Gokpinar et al. 2010)), additional costs are incurred. One example is adapting the charger when improving smartphone battery and charging. Let EB'_K and MLT'_K denote the additional cost and production lead time associated with other feature adjustments when improving feature TB_k. Besides the above factors, market information such as the latest designs and fashion trends of competing products also affects the NPD process. Therefore, designers may proactively identify specific product features for improvement to cope with the latest changes in the market. Let MB denote the predetermined product features (the number should be less than K). Therefore, the MPI problem can be updated as follows:

$$Max\ PI^K = w_I \sum_{k=1}^{K} Imp(TB_k) - w_P \sum_{k=1}^{K} Per(TB_k) + w_O OC^K$$
$$s.t. \begin{cases} \sum_{k=1}^{K} EB_k + EB'_K \leqslant PC_1 \\ MLT_K + MLT'_K \leqslant PC_2 \\ \sum_{k=1}^{K} (CR_k + CD_k + CS_k) \leqslant PC_3 \\ MB \subset TB \end{cases} \tag{A2}$$

where PC_1, PC_2, and PC_3 are production constraints.

It is worth noting that although the objective of Equation (A2) is to maximize the product improvement index PI^K rather than the product profit, as in Section 5.1, w_I can be used to some extent as the relative contribution of each feature pair to the profit, and the objective of Equation (A2) can be made to move closer to profit maximization by setting a larger w_I. Furthermore, the experiments presented in Appendix C.2 show that the product feature attributes involved in Equation (A2) (importance, performance and competitiveness) outperform other studies in predicting sales, indicating that boosting PI^K aligns with increasing product sales. Therefore, although PI^K does not directly represent profit, maximizing PI^K is consistent with maximizing profit. After adding these constraints to Algorithm G1 in Appendix G (e.g., to line 59), we can still use the MSCS algorithm to solve the problem. The product improvement strategy developed using this method can better satisfy the needs of actual production. Future work can improve the MSCS algorithm for different constraints.

References

Archak N, Ghose A, Ipeirotis PG (2011) Deriving the pricing power of product features by mining consumer reviews. *Management Science* 57(8):1485–1509.

Bi JW, Liu Y, Fan ZP, Zhang J (2019) Wisdom of crowds: Conducting importance-performance analysis (IPA) through online reviews. *Tourism Management* 70:460–478.

Geva T, Oestreicher-Singer G, Efron N, Shimshoni Y (2017) Using forum and search data for sales prediction of high-involvement projects. *Mis Quarterly* 41(1):65–+, URL http://dx.doi.org/10.25300/misq/2017/41.1.04.

Ghose A, Ipeirotis PG, Li BB (2019) Modeling consumer footprints on search engines: An interplay with social media. *Management Science* 65(3):1363–1385, URL http://dx.doi.org/10.1287/mnsc.2017.2991.

Gokpinar B, Hopp WJ, Iravani SMR (2010) The impact of misalignment of organizational structure and product architecture on quality in complex product development. *Management Science* 56(3):468–484.

Hu F, Li HX, Liu Y, Teichert T (2020) Optimizing service offerings using asymmetric impact-sentiment- performance analysis. *International Journal of Hospitality Management* 89:102557.

Liu XM, Wang GA, Fan WG, Zhang ZJ (2020) Finding useful solutions in online knowledge communities: A theory-driven design and multilevel analysis. *Information Systems Research* 31(3):731–752.

Nelson P (1970) Information and consumer behavior. *Journal of Political Economy* 78(2):311–329.

Sridhar K, Bezawada R, Trivedi M (2012) Investigating the drivers of consumer cross-category learning for new products using multiple data sets. *Marketing Science* 31(4):668–688.

Yan ZJ, Xing MM, Zhang DS, Ma BZ (2015) Exprs: An extended pagerank method for product feature extraction from online consumer reviews. *Information & Management* 52(7):850–858.

Yang Y, Zhang KP, Fan YY (2022) sDTM: A supervised bayesian deep topic model for text analytics. *Information Systems Research* Doi:https://doi.org/10.1287/isre.2022.1124.

Zhan YZ, Tan KH, Huo BF (2019) Bridging customer knowledge to innovative product development: a data mining approach. *International Journal of Production Research* 57(20):6335–6350.

Zhang CX, Xu ZS, Gou XJ, Chen SX (2021) An online reviews-driven method for the prioritization of improvements in hotel services. *Tourism Management* 87:104382.

Zhang L, Chu XN, Xue DY (2019) Identification of the to-be-improved product features based on online reviews for product redesign. *International Journal of Production Research* 57(8):2464–2479.

Zhang Z, Wei X, Zheng XL, Li QD, Zeng DD (2022a) Detecting product adoption intentions via multiview deep learning. *Informs Journal on Computing* 34(1):541–556.

Zhang ZL, Yang KJ, Zhang JZ, Palmatier RW (2022b) Uncovering synergy and dysergy in consumer reviews: A machine learning approach. *Management Science* Doi:https://doi.org/10.1287/mnsc.2022.4443.