

Machine Learning and Artificial Intelligence Research in Tourism and Hospitality

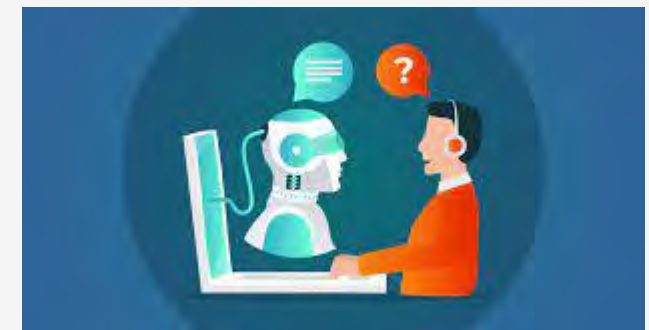
Yang Yang, Ph.D.

Associate Professor
Temple University

ML and AI matter

About 12% of other travel and tourism businesses (e.g. hospitality and entertainment) have embraced AI at scale with an annual growth rate of 6.5% (Huang, et. Al., 2021).

AI technologies can offer various advantages for both suppliers (e.g. improved productivity, efficiency and profitability) and consumers (e.g. convenient and personalized tourist experiences) (Samara et al., 2020).



ML and AI matter - data

Various types of big data available

GPS track of vehicles

Social media

Cell phone roaming

Text analytics became particularly popular to explore text data, and popular methods include sentiment analysis, topic mining, and document classification (Zhang, Qiao, Yang, & Zhang, 2020).



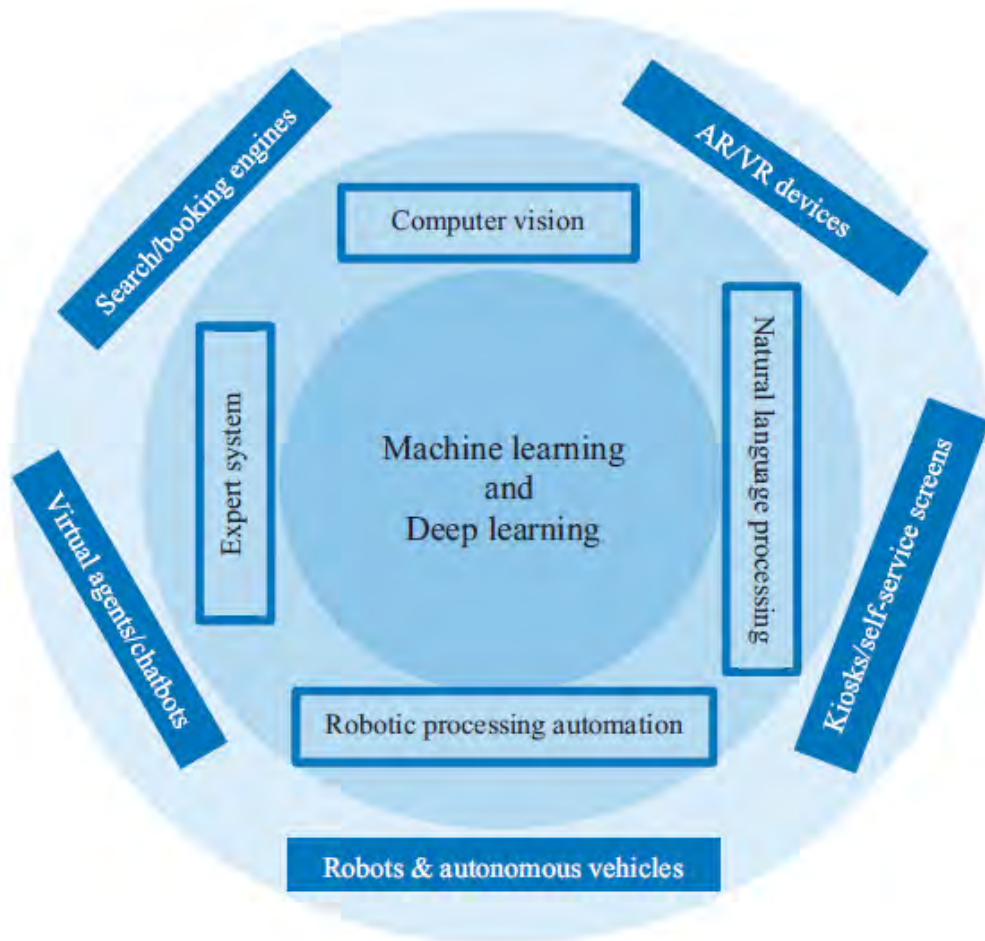
User Generate Content

Credit card transaction

**Search engine query/
web traffic**

Also, sophisticated artificial intelligence methods have been introduced to analyze big data in tourism (Zhang, et al., 2019).

ML and AI matter



Huang, et.al. (2021)

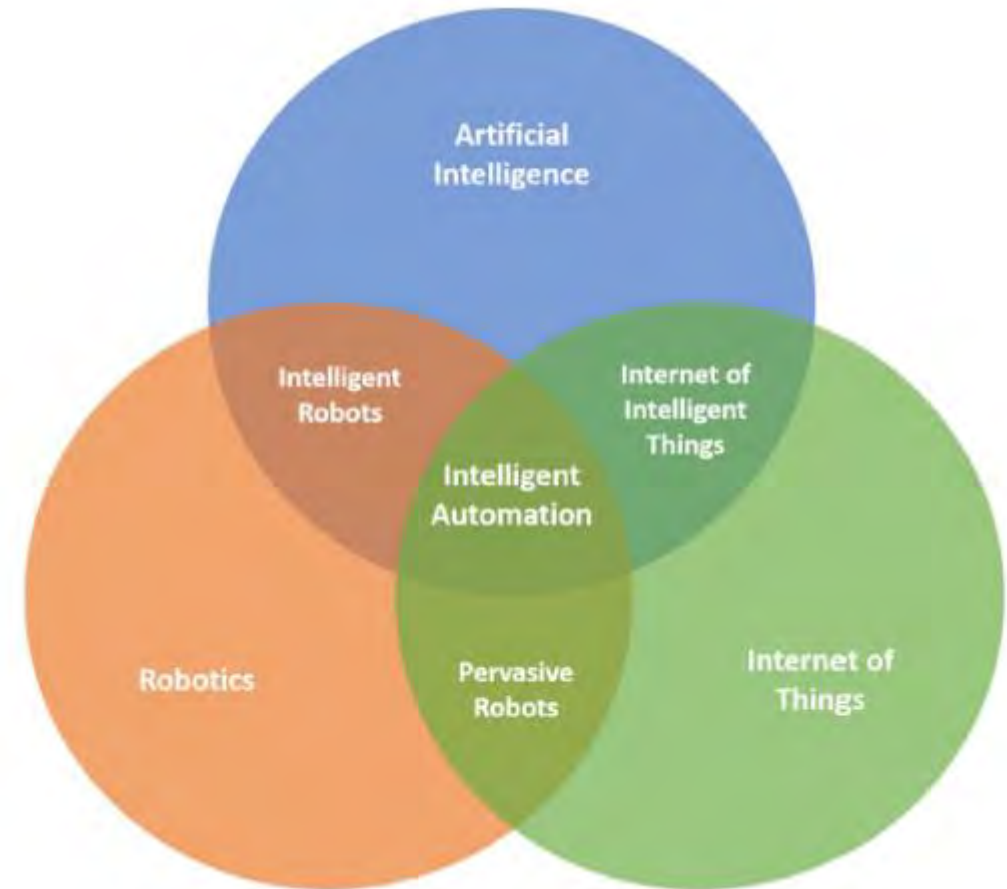


Fig. 1. Technological Framework of Intelligent Automation in Tourism.

lis (2020)

Research examples

1. Machine learning and GIS system: hotel location selection tool
2. Text-mining: sleep quality study
3. CNN: air quality study
4. Image analytics: avatar and review usefulness
5. Video analytics: Facebook videos on destination marketing
6. Experiment study: robotic involvement

1. Machine learning and GIS system



Contents lists available at ScienceDirect

International Journal of Hospitality Management

journal homepage: www.elsevier.com/locate/ijhosman

Hotel location evaluation: A combination of machine learning tools and web GIS

Yang Yang^{a,1}, Jingyin Tang^{b,2}, Hao Luo^{c,*}, Rob Law^{d,3}

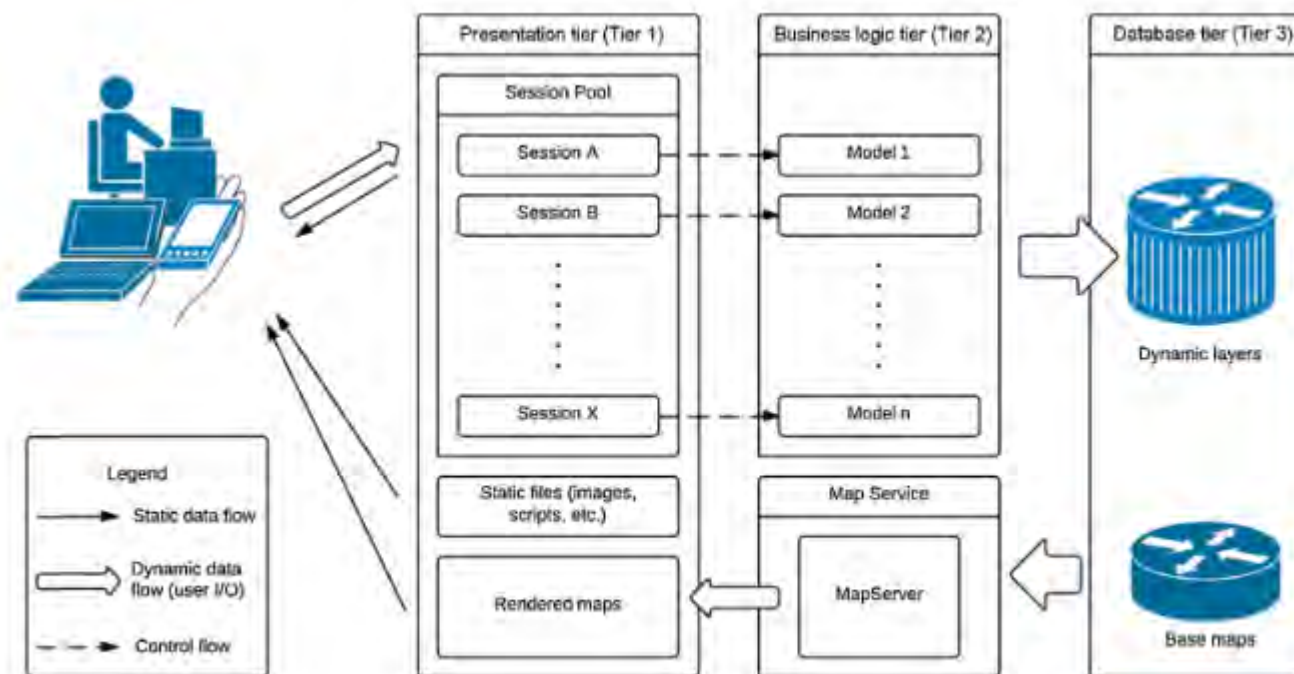
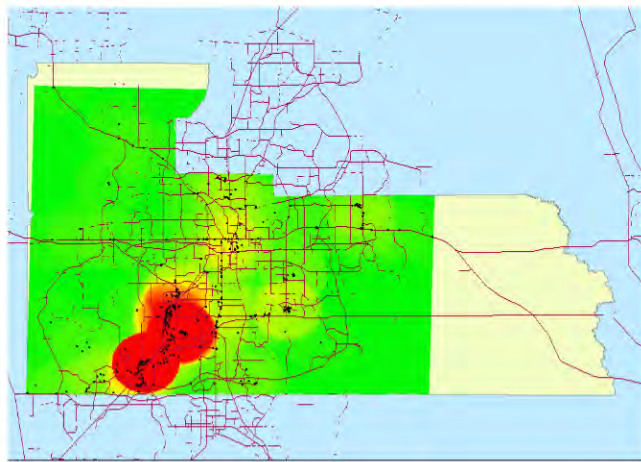


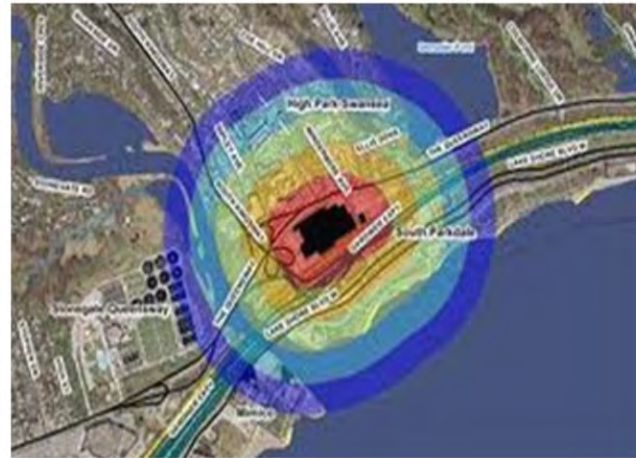
Fig. 1. Three-tier architecture of HoLSAT.

1. Machine learning and GIS system

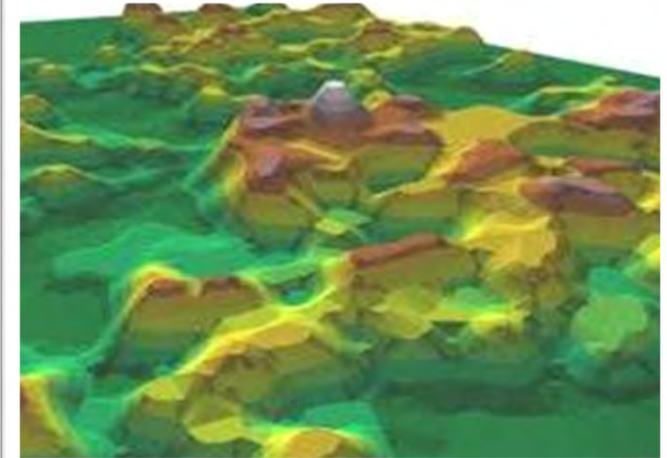
- Visualize, question, analyze, interpret, and understand data to reveal relationships, patterns, and trends



Hot-spot of hotel demand

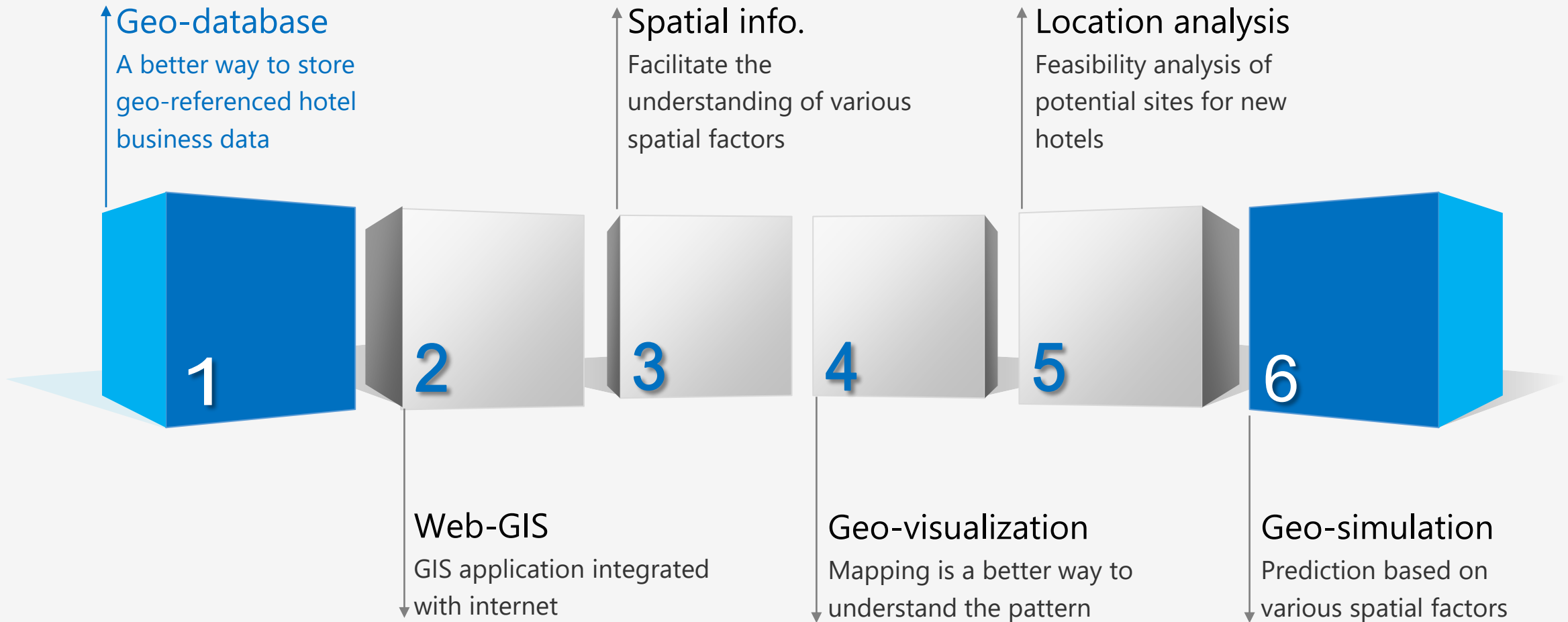


Accessibility measure



Hotel pricing surface

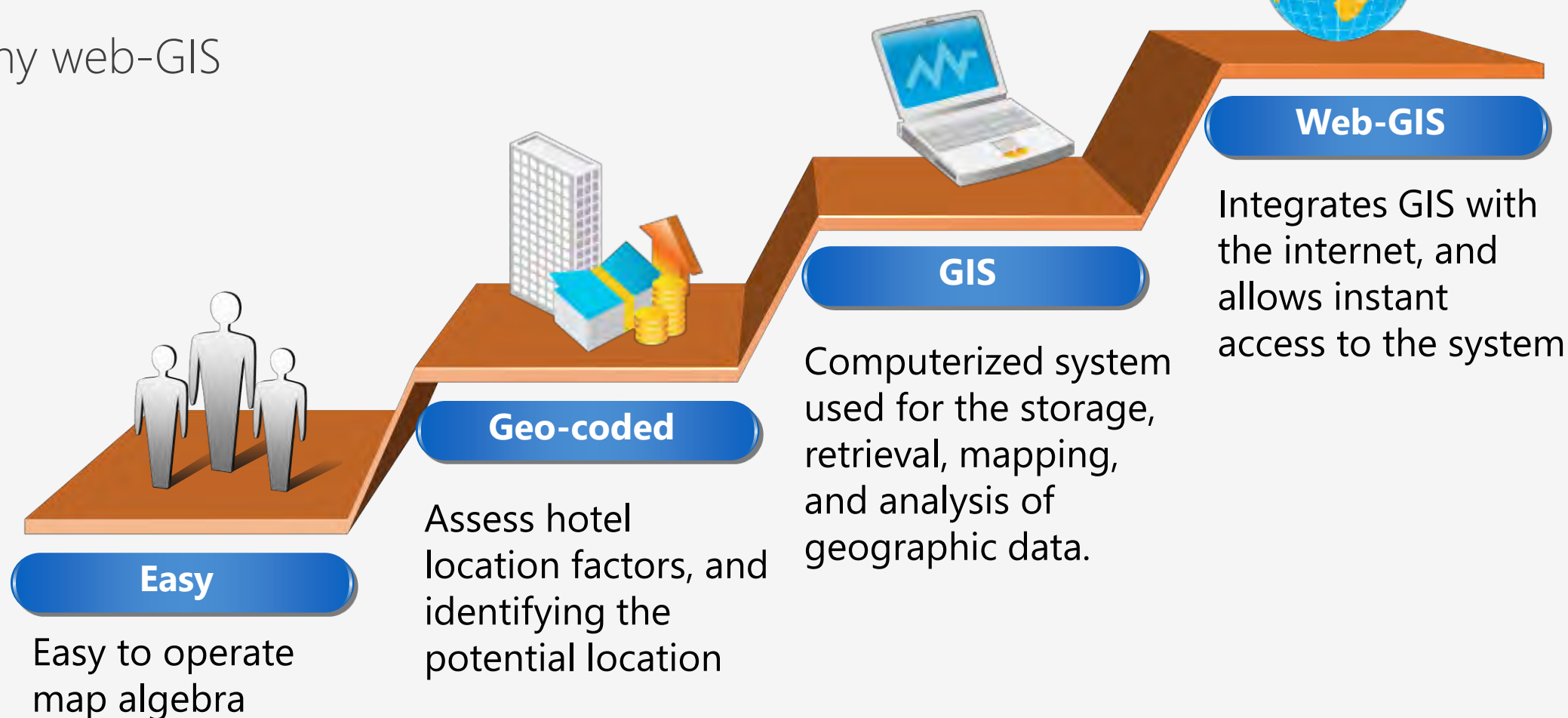
1. Machine learning and GIS system



1. Machine learning and GIS system



Why web-GIS



1. Machine learning and GIS system

Neural Network



1. Inspired by animal nervous systems
2. Identify patterns or structures in data
3. Approximate complex function relationships

$$y = f\left(\beta_0 + \sum_{j=1}^r \beta_j \cdot f(\beta_{0j} + \beta_j' \mathbf{x})\right)$$

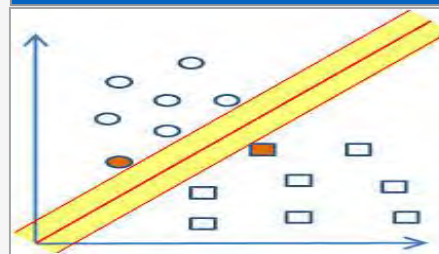
Projection Pursuit



1. Capture the non-linear relationship
2. Alleviates the multi-collinearity
3. Provides an ideal solution for over-fitting

$$y = \beta_0 + \sum_{j=1}^r f_j(\beta_j' \mathbf{x}) + \varepsilon$$

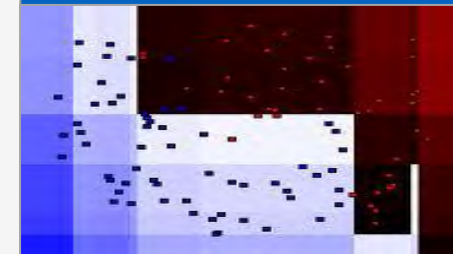
Support Vector



1. Robust and distribution-free
2. Solid statistical theory background
3. Capture non-linear relationship

$$y = \beta_0 + (\mathbf{w}' \cdot \Phi(\mathbf{x}))$$

Boosting Reg.



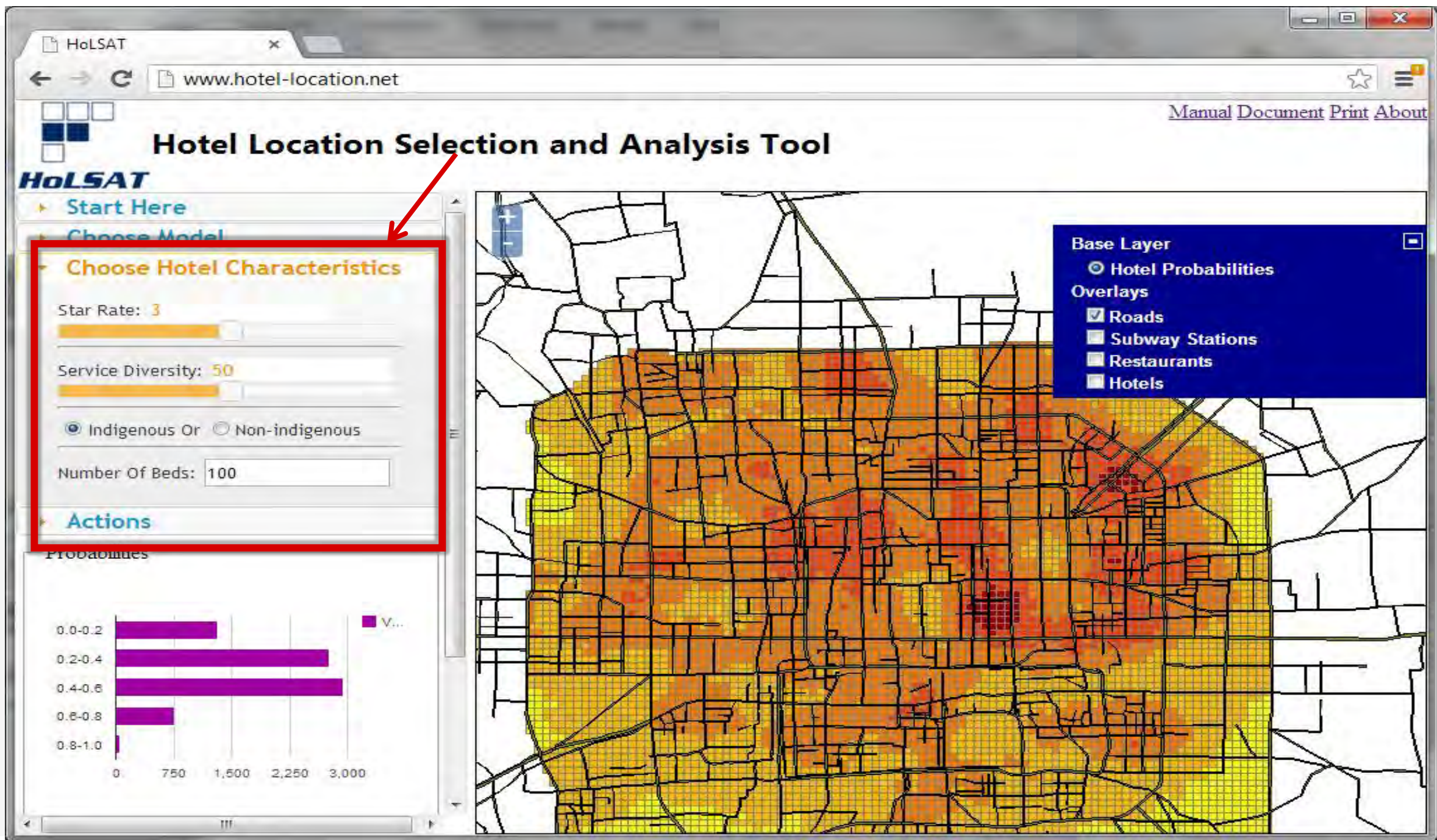
1. Free of a pre-specified functional relationship
2. Owns inherent advantages in predictive accuracy

$$y = F_0(\mathbf{x}) + \sum_{m=1}^M v \rho_m B_m(\mathbf{x})$$

1. Machine learning and GIS system

Table 3
Results of models with alternative business success indicators.

	Linear regression	Projection pursuit regression	Neural network	Support vector regression	Boosted regression
Labor productivity					
Pseudo- R^2	0.157	0.181	0.146	0.184	0.107
MAE (CV)	6.784	6.178	6.879	6.323	6.763
MSE (CV)	246.704	224.884	251.858	247.676	280.126
Occupancy rate					
Pseudo- R^2	0.153	0.207	0.179	0.237	0.225
MAE (CV)	7.920	8.107	8.017	8.373	8.505
MSE (CV)	99.231	103.418	100.422	110.498	113.054
DEA efficiency score					
Pseudo- R^2	0.184	0.206	0.182	0.262	0.149
MAE (CV)	0.0852	0.0891	0.0892	0.0922	0.1276
MSE (CV)	0.0119	0.0131	0.0131	0.0132	0.0239



2. Text-mining: sleep quality study



Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

International Journal of Hospitality Management

journal homepage: www.elsevier.com/locate/ijhm

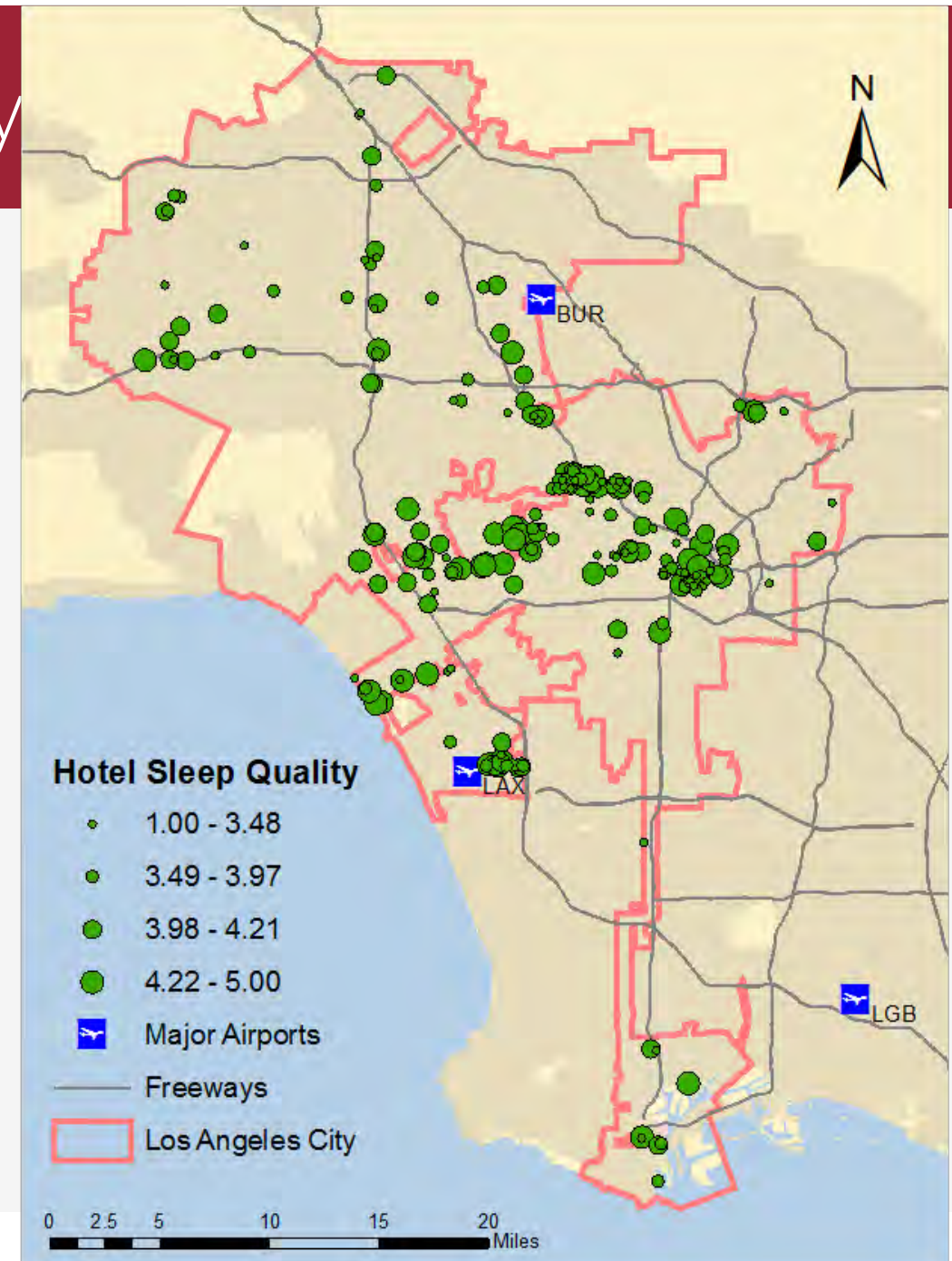
Sleepless nights in hotels? Understanding factors that influence hotel sleep quality


Zhenxing Mao^a, Yang Yang^{b,*}, Mingshu Wang^c

2. Text-mining: sleep quality study

Objective: investigate how unique travel- and hotel-related factors (facilitators and barriers) contribute to subjective sleep quality for travelers who stay overnight in hotels.

Method: we collected a large corpus of TripAdvisor review data for hotels in Los Angeles to explore and unveil how different factors help influence the level of sleep quality rated by reviewers.







Jakarta, Indonesia
 Level **6** Contributor
 903 reviews
 307 hotel reviews
 720 helpful votes


"We hope it is better next time"


 Reviewed 1 week ago

We looked forward to another delightful stay at Courtyard LAX after a great overnight in 2013. But the first shuttle van passed us up at Bradley Terminal and the air vent in our room 628 blew directly on to the bed. We were too tired to pack up and change rooms, so we wound up sleeping with the window open despite the noise of aircraft landing and taking off. Check-out in the morning was fast, and the shuttle bus took us right to the airport without waiting for other passengers.

Stayed February 2017, traveled as a couple

 Sleep Quality
  Cleanliness
 Service








Jakarta, Indonesia
 Level **6** Contributor
 903 reviews
 307 hotel reviews
 720 helpful votes

Level **6** Contributor

- TripAdvisor member since 2006
- 65+ man from Jakarta, Indonesia




Review distribution (900)

	Excellent (448)
	Very good (326)
	Average (106)
	Poor (21)
	Terrible (2)

1,705 Contributions
720 Helpful votes
287 Cities visited

[Message](#)
[Full profile](#)

Stayed February 2017, traveled as a couple

 Sleep Quality
  Cleanliness
 Service

2. Text-mining: sleep quality study

We propose a framework/typology to classify potential sleep factors among hotel guests into two meaningful domains, traveler characteristics and hotel characteristics



2. Text-mining: sleep quality study

Demographic characteristics

traveler_age

Traveler age: 1 = 18–24; 2 = 25–34; 3 = 35–49; 4 = 50–64; 5 = 65 and above

traveler_gender

Traveler gender: 1 = male; 2 = female.

Tripographic characteristics

traveler_type

Traveler type: 1 = couple travelers; 2 = business travelers; 3 = solo travelers;
4 = family travelers; 5 = travelers with friends

Incities

Log of number of different cities visited as shown in reviewer profile

Indistance

Log of geographical distance between the home city and Los Angeles (in km)

longitude_dif

Difference in longitude between the home city and Los Angeles

Hotel facilities

floors

Number of floors of the hotel's major building with guest accommodations

star

Star rating for the hotel on TripAdvisor

Hotel location

restaurant

Number of restaurants within 1 km of the hotel

freeway

An indicator of a freeway within 200 m of the hotel

airport

An indicator of an airport within 2 km of the hotel

canopy_cover

Tree canopy cover percentage within 500 m of hotel

Control variable

expertise

Reviewer's expertise class on TripAdvisor from 0 to 6.

2. Text-mining: sleep quality study

We searched the review content using the defined list of sleep-related keywords associated with hotel sleeping environment (Pallesen, et al., 2016).

We also manually excluded irrelevant reviews like those highlighting pool temperature and lighting problems when taking showers.

The results demonstrate the applicability and accuracy of using sentiment analysis to generate an evaluation measure of each aspect related to sleeping environment in a hotel setting.

Factor	Observations	Mean of score	Std. Dev. of score	Example
Pillow	373	0.306	0.270	<p>"Beds and pillows are glorious :)" (score = 0.500)</p> <p>"The one thing I hate is tons of pillows on the bed that are all down" (score = -0.176)</p>
Mattress	127	0.205	0.306	<p>"Very comfortable beds with mattress cushion covers" (score = 0.560)</p> <p>"The comfort level of the mattress was so poor that my sleep was terrible" (score = -0.548)</p>
Bedding	170	0.264	0.296	<p>"Very comfortable bedding" (score = 0.523)</p> <p>"The carpet is dingy shag and the bedding is outdated" (score = -0.400)</p>
Temperature	251	0.162	0.268	<p>"Room temperature - I admittedly like my room on the cold side but it seems as though the insulation of our room was weak and by mid-afternoon it got very warm" (score = 0.041)</p> <p>"The temperature was impossible to control" (score = -0.483)</p>
Noise	1337	0.146	0.233	<p>"I expected some street noise but was surprised how good the sound proofing was" (score = 0.431)</p> <p>"The noise was going on all night with doors being slammed and loud talking in the corridors" (score = -0.236)</p>
Lighting	225	0.227	0.285	<p>"It was nice to have a dimmer so we could have the light on at night without being too intrusive" (score = 0.313)</p> <p>"The noise is pretty bad and also the lights from the Holiday Inn sign creates a green glow into the room" (score = -0.217)</p>
Ventilation	39	0.095	0.242	<p>"On the upside, the ventilation in the room was terrific" (score = 0.407)</p> <p>"Poor ventilation which kept the bathroom hot and steamy after showering" (score = -0.250)</p>

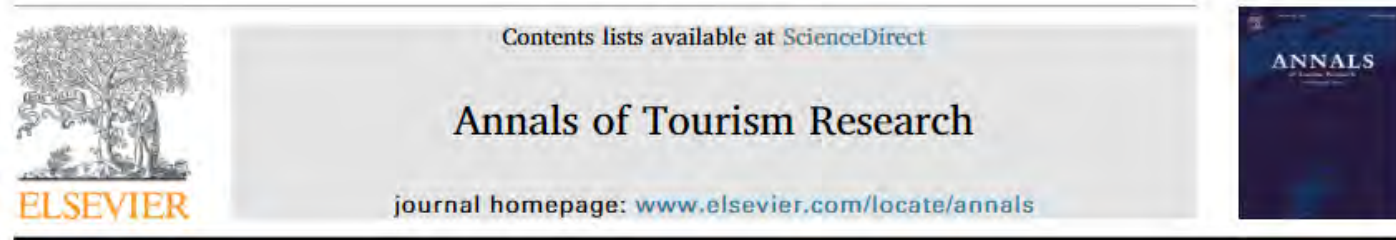
2. Text-mining: sleep quality

S_mattress, S_temperature, S_noise and S_lighting—are positive and significant, suggesting that more positive evaluations of mattresses, temperature, noise and lighting are associated with a higher levels of reported sleep quality.

S_mattress has the largest coefficient, which again highlights the important role mattresses play in determining guests' sleep quality.

	Model 6	Model 7	Model 8	Model 9
I_pillow	0.161 (0.134)	-0.550*** (0.192)	0.137 (0.133)	-0.558*** (0.191)
I_mattress	-0.991*** (0.234)	-1.604*** (0.267)	-0.993*** (0.237)	-1.590*** (0.273)
I_bedding	0.00114 (0.192)	-0.233 (0.247)	-0.0592 (0.198)	-0.272 (0.255)
I_temperature	-0.744*** (0.128)	-0.989*** (0.126)	-0.738*** (0.128)	-0.986*** (0.132)
I_noise	-0.750*** (0.071)	-0.898*** (0.081)	-0.764*** (0.073)	-0.903*** (0.082)
I_lighting	-0.211* (0.128)	-0.476*** (0.153)	-0.229* (0.134)	-0.495*** (0.160)
I_vent	-0.769*** (0.259)	-0.725** (0.285)	-0.795*** (0.266)	-0.749** (0.297)
S_pillow		2.473*** (0.461)		2.431*** (0.468)
S_mattress		2.974*** (0.731)		2.895*** (0.757)
S_bedding		0.800 (0.557)		0.678 (0.567)
S_temperature		2.131*** (0.483)		2.075*** (0.486)
S_noise		1.213*** (0.222)		1.154*** (0.222)
S_lighting		1.284*** (0.415)		1.304*** (0.436)
S_vent		-0.334 (0.993)		-0.391 (1.154)

3. CNN: air quality study



Research article

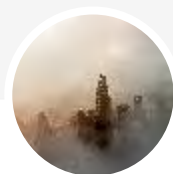
Designing tourist experiences amidst air pollution: A spatial analytical approach using social media

Xiaowei Zhang^a, Yang Yang^{b,*}, Yi Zhang^c, Zili Zhang^a

Impact of Air Pollution

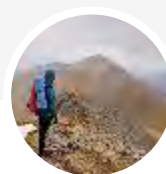
(Measured by PM 2.5 concentration level)

Smog Awareness



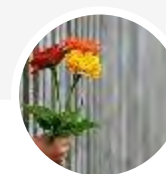
Whether the tourists are aware of the issue of air pollution, such as smog attacks

Behavioral Consequence



How tourists change their behavior in terms of location visited, travel scope, and duration.

Emotional Consequence



How tourists' emotion and sentiment change.

Health Consequence



The occurrence of health related issues such as illness and insomnia.

3. CNN: air quality study

Many tourist destinations are suffering from varying levels of air pollution. The situation is particularly worse in urban destinations.

POLLUTION

Foreign Tourists Skipping Delhi over Air Quality Fears

By Niharika Lal | TNN | 12 December 2018 | TWC India



Tourists at Chandni Chowk market in Delhi (RAJESH MEHTA/ BCLL Delhi)

South China Morning Post

SIGN IN/UP

China

Air pollution takes toll on China's tourism

Shocking levels of air pollution have cast a pall over China's burgeoning tourism industry



Published: 4:02pm, 13 Aug, 2013



3. CNN: air quality study

Sentiment analysis



Opinion Mining within Natural Language Processing (NLP) that builds systems that try to identify and extract opinions



Chinese text segmentation: built to be the best Python Chinese word segmentation module.



Provides solutions from automatic Chinese words segmentation to psychological analysis.

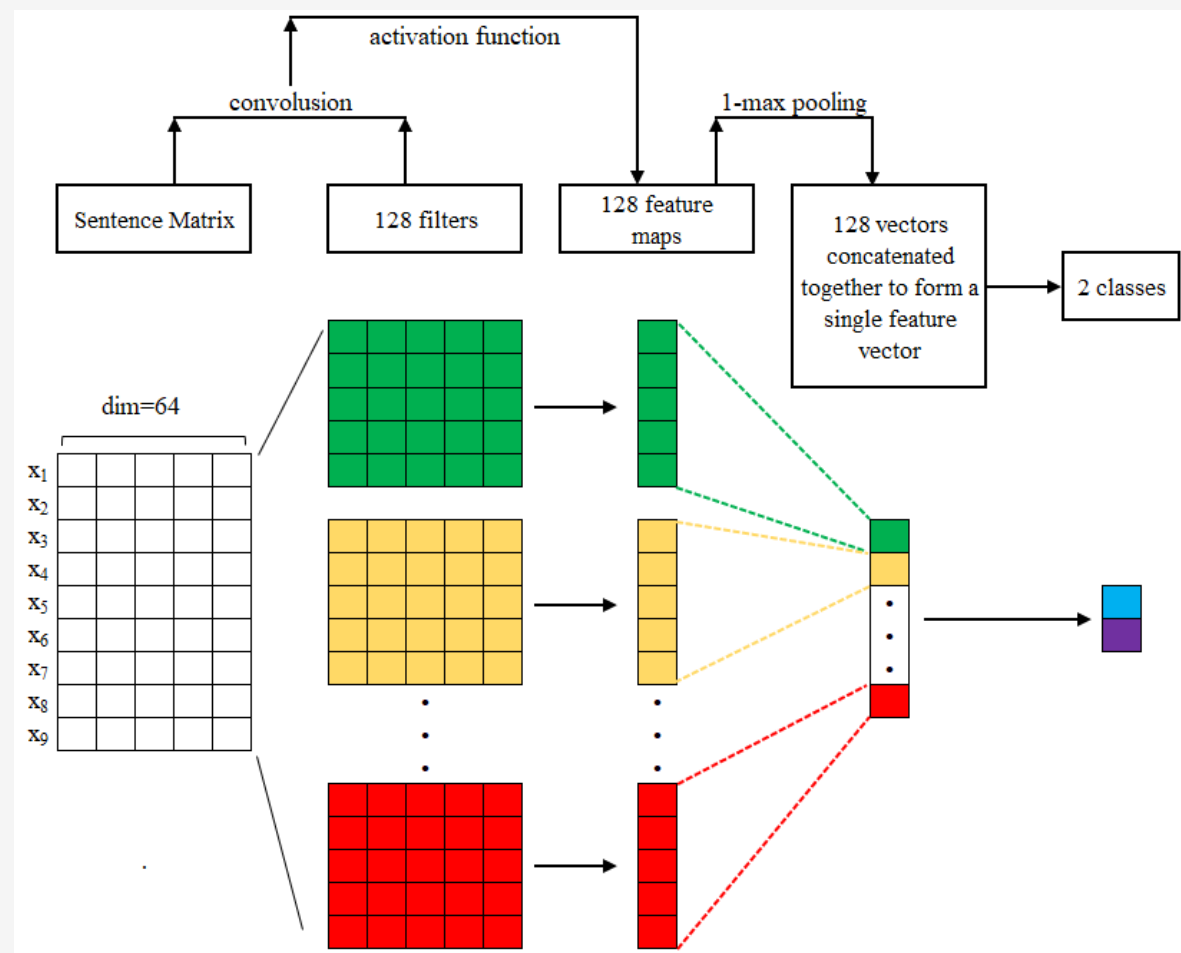
3. CNN: air quality study

Deep learning classification

manually check 14300 Weibo for health issues

health-related keywords include coughing“咳嗽”, dizzy“头晕”, sickness“生病”, headache“头痛”, and difficulty breathing “呼吸困难”

11000 texts as the training dataset, 2200 texts as the testing dataset, and the other 1100 as the validation dataset



3. CNN: air quality study

Variables	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	smog_tag	distance_TAM	radius	sentiment	health_issue	sleep_issue
pm2_5*(tourist=0)	0.00398***	-0.000508***	0.000146***	-0.000605***	0.000262***	0.000244***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
pm2_5*(tourist=1)	0.00397***	-0.000103**	0.000213**	-0.000685***	0.000391***	0.000120
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
tourist	0.505***	-0.182***	0.923***	-0.187***	-0.193***	-0.200***
	(0.020)	(0.007)	(0.012)	(0.011)	(0.006)	(0.017)
sunny	0.151***	0.0300***	-0.0680***	-0.0436***	0.0259***	-0.0385***
	(0.015)	(0.004)	(0.006)	(0.007)	(0.004)	(0.010)
precipitation	0.132***	0.0336***	0.0499***	-0.0950***	-0.00410	0.00614
	(0.017)	(0.004)	(0.006)	(0.007)	(0.004)	(0.010)
windy	-0.439***	-0.0317***	0.0593***	0.00185	-0.000743	-0.0120
	(0.016)	(0.004)	(0.006)	(0.007)	(0.003)	(0.010)
temp	0.0713***	0.0240***	-0.00936***	-0.0161***	0.00148***	-0.00955***
	(0.002)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)
temp2	-0.00480***	-0.000229***	0.000582***	0.000405***	-0.000159***	0.000243***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

3. CNN: air quality study

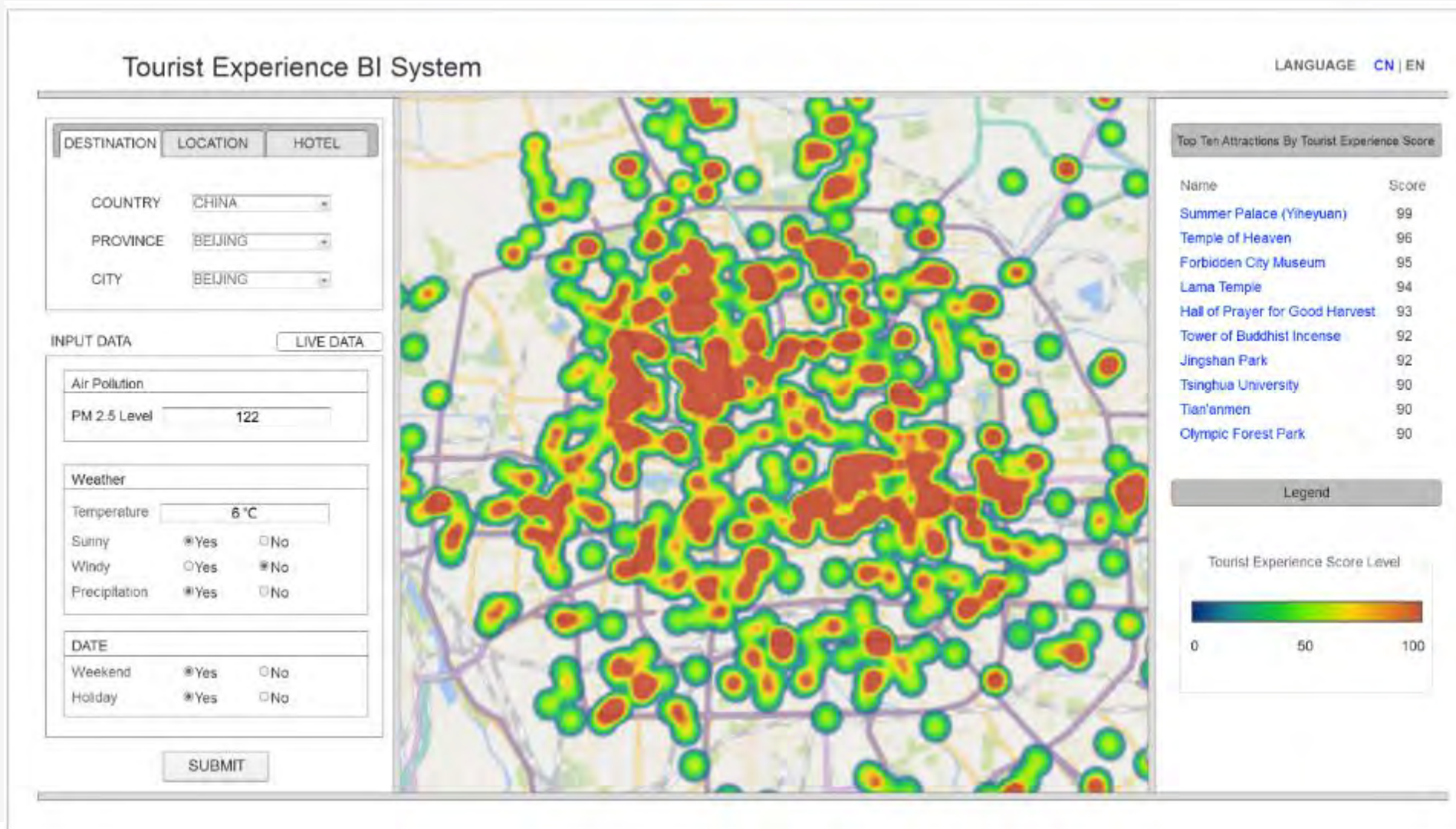
Tourists are more vulnerable than residents for health-related issues

- Tourists are less adaptive to an unfamiliar environment.

Tourists do not suffer from sleep issues due to air pollution.

- Many other factors shaping the sleep quality of tourists that may be more important than air pollution, such as travel fatigue, accommodation environment, jet lag, etc.

3. CNN: air quality study



4. Image analytics: avatar and review

Workshop on Tourism Design Analytics
Vienna, Austria. Nov 8, 2017

Do Avatars Matter?

Yang Yang Ph.D.
Simon Hu Ph.D. student
Temple University

Jingyin Tang Ph.D.
IBM

Real estate agent



Car dealer



4. Image analytics: avatar and review



Reviewers' profile photos (RPPs) were identified as an important indicator of reviewers' characteristics (Lee & Shin, 2014)

4. Image analytics: avatar and review

Based on supervised learning results on billions of photos

- Returns information about visual content found in an image.
- Uses tagging, descriptions, and domain-specific models to identify content and label it with confidence.
- Identifies image types and color schemes in pictures.
- Recognizes Celebrity and Landmark around the world.
- Detects faces and its gender and age.



4. Image analytics: avatar and review

	Model 2	Model 3	Model 4
	Category	Tag	Tag
green	-0.632* (0.343)	-0.0174 (0.041)	
red	-0.421 (0.287)	0.0314 (0.040)	
yellow	0.0222 (0.173)	-0.00994 (0.053)	
blue	0.0428 (0.228)	0.0409 (0.036)	
black	0.0926 (0.301)	-0.0389 (0.057)	
grey	1.890*** (0.529)	0.567*** (0.216)	
white	-0.366 (0.240)	-0.0211 (0.030)	
pink	-0.110 (0.569)	-0.200 (0.156)	
purple	1.176 (1.144)	0.509 (0.391)	
colors			0.0492*** (0.017)

	Model 5	Model 6	Model 7	Model 8	Model 9
	Category	Category	Category	Category	Tag
C_outdoor	0.0281** (0.011)				
C_food		-0.227** (0.113)			
C_people_portrait			0.0624** (0.029)		
C_people				-0.109 (0.068)	
T_smiling					-0.0429 (0.046)

4. Image analytics: avatar and review

Ask graduate students to rank credibility and social distance (scale 1-5) of 1,500 randomly selected RPPs in TripAdvisor.

Use machine learning tools to train a model based on 1,000 rated RPPs, and use the remaining 500 as a validation set.

	Model 10	Model 11	Model 12
Social dist	0.102*		0.088
	(0.060)		(0.059)
Trustworthiness		-0.017	-0.015
		(0.013)	(0.013)

5. Video analytics: Facebook videos

Video analytics of Facebook videos from DMOs

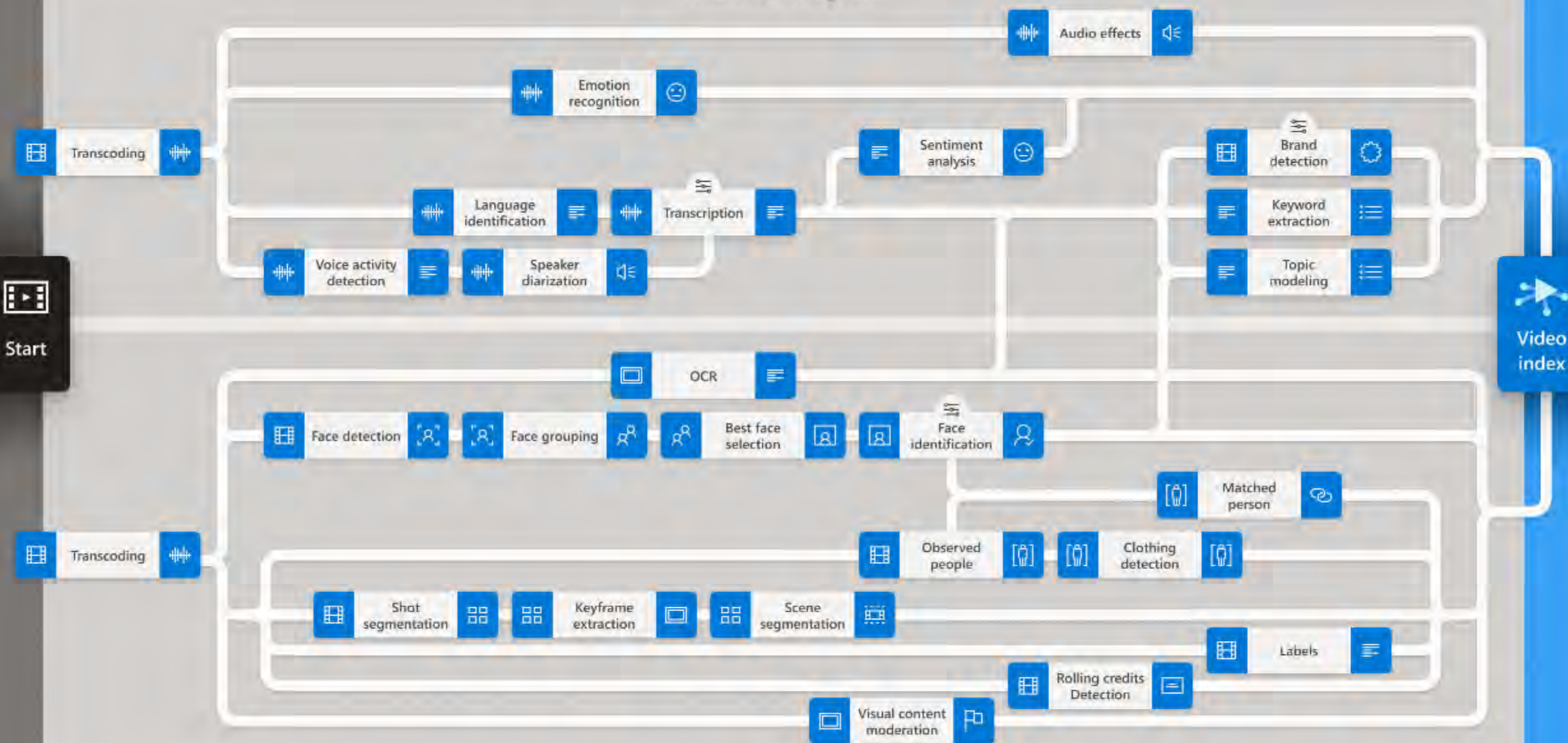


5. Video analytics: Facebook videos

We collected the video (<10 min) from state DMOs' Facebook accounts, and conduct the video analytics in Azure Video Analyzer APIs.

More than 3,000+ videos were analyzed, and regression analysis was used to understand what video factors are associated with social media post performance indicators (i.e., number of views, number of likes, number of comments, number of likes, number of shares....)

Audio analysis



Vision analysis

5. Video analytics: Facebook videos

	Variable definition	(1)	(2)				
Intitle_len	Lenth of title in words (in log)	Inviews -0.0286 (0.077)	Inlikes -0.109* (0.061)				
Inbody_len	Lenth of body content in words (in log)	0.217*** (0.036)	0.278*** (0.031)				
Invideo_len	Lenth of video in seconds (in log)	0.273*** (0.090)	0.0808 (0.073)				
is_hd	Whether the resolution is 720 or higher	0.418*** (0.130)	0.233** (0.097)	Inword_count	Word count spoken (in log)	-0.125*** (0.038)	-0.0982*** (0.030)
L_start	Luminance of first 30sec	-0.000252 (0.001)	-0.000308 (0.001)	Sentiment_Neg	(Based on audio/text in the video) negative sentiment score	1.119* (0.577)	0.193 (0.445)
L_mid	Luminance at midpoint	0.000634 (0.001)	0.000562 (0.001)	Sentiment_Pos	(Based on audio/text in the video) positive sentiment score	0.121 (0.139)	0.0861 (0.114)
L_end	Luminance of last 30sec	0.000357 (0.001)	0.0000918 (0.001)	smile	Indicators of smiles in video content	0.0690 (0.107)	-0.0408 (0.083)
C_start	Colorfulness of first 30sec	0.00187** (0.001)	0.00174** (0.001)	laugh	Indicators of laughs in video content	1.134*** (0.330)	1.137*** (0.259)
C_mid	Colorfulness at midpoint	-0.000829 (0.001)	-0.000574 (0.001)	historic	Indicators of historical in video content	0.566* (0.321)	0.211 (0.223)
C_end	Colurfulness of last 30sec	-0.000832 (0.001)	-0.000615 (0.001)	nature	Indicators of nature in video content	0.362*** (0.075)	0.419*** (0.061)
luminance_var	standardar deviation between L_start L_mid, and L_end to capture the varia	-0.00157 (0.002)	-0.00327** (0.001)	plant_tree	Indicators of trees and plants in video content	0.0480 (0.085)	0.00238 (0.068)
colorfulness_var	standardar deviation between C_start, C_mid, and C_end to capture the var	-0.00152 (0.002)	-0.000869 (0.001)	animal	Indicators of animals in video content	0.154** (0.072)	0.171*** (0.059)
Innumb_scenes_minute	Number of scenes per minute (in log)	0.120* (0.066)	-0.0143 (0.052)	people	Indicators of people in video content	-0.0117 (0.080)	0.0112 (0.066)
Innumb_faces_minute	Number of faces per minute (in log)	-0.0470 (0.052)	-0.148*** (0.042)	_cons		6.205*** (0.518)	3.508*** (0.424)
Innumb_locations_minute	Number of well-known locations per minute (in log)	0.115* (0.062)	0.148*** (0.048)	State-specific effects		Yes	Yes
Num_Speakers	How many people are speaking	0.0453*** (0.011)	0.0170* (0.009)	Year-month-specific effects		Yes	Yes
TalkToListenRatio	How many people are speaking / How many people are listening	0.567*** (0.218)	-0.123 (0.152)	Day-of-week-specific effects		Yes	Yes
				Posting-hour-specific effects		Yes	Yes
				N		2681	2676
				AIC		9625.4	8398.2
				BIC		9790.4	8563.1
				ll		-4784.7	-4171.1
				r2		0.585	0.643
				r2_a		0.553	0.616

6. Experiment study: robotic involvement



Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

International Journal of Hospitality Management

journal homepage: www.elsevier.com/locate/ijhman

Tech-touch balance in the service encounter: The impact of supplementary human service on consumer responses

Laurie Wu^a, Alei Fan^b, Yang Yang^c, Zeya He^{d,*}

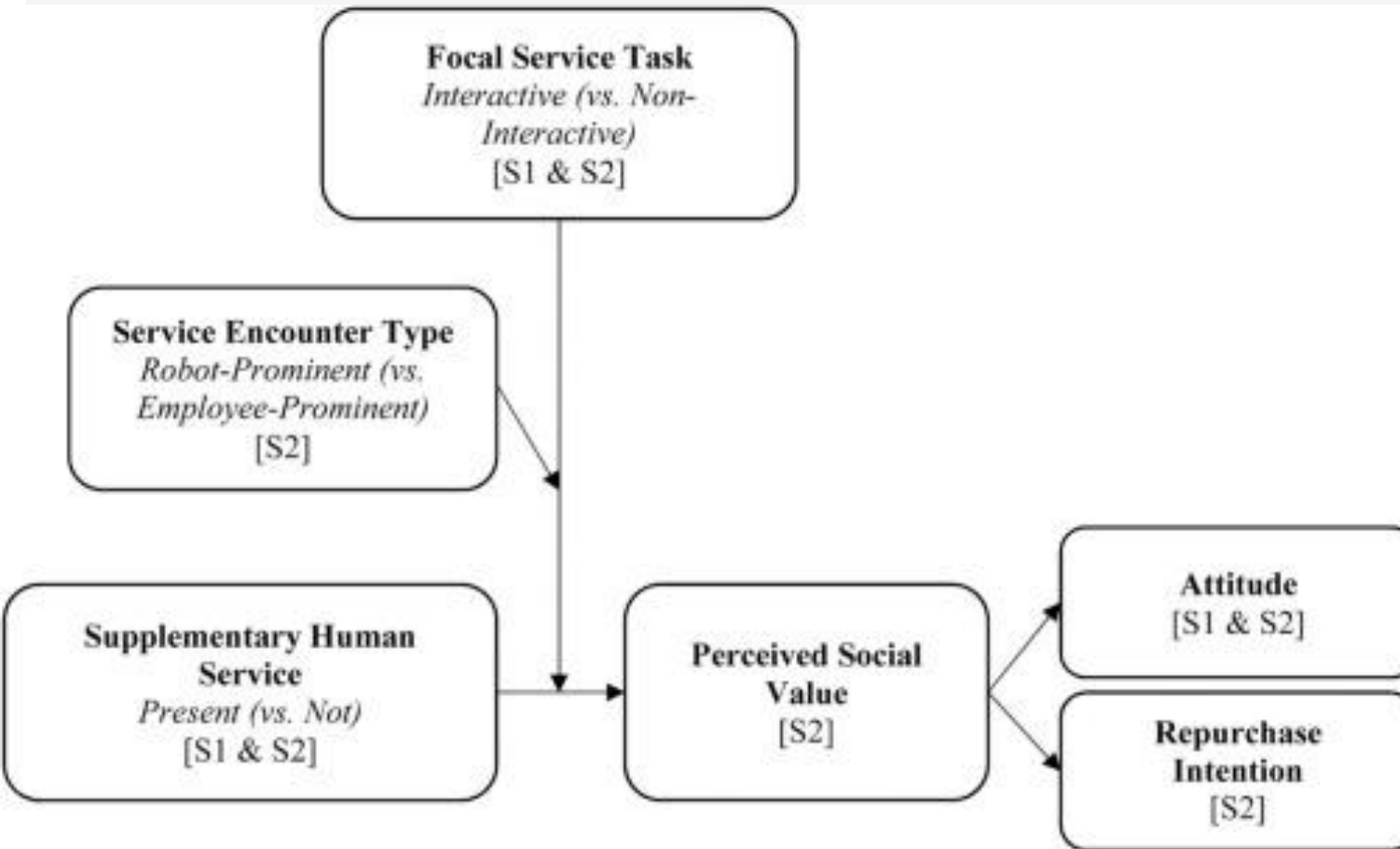


6. Experiment study: robotic involvement

Trends in service innovation continue to shift service encounters from human-prominent to robot-prominent (Larivière et al., 2017, Van Doorn et al., 2017).

To date, relatively little scholarly attention has been paid to designing robot-prominent service encounters to overcome obstacles imposed by an absence of socialization (e.g., [Fernandes and Oliveira, 2021](#); [Yoganathan et al., 2021](#)).

We draw upon the notion of “tech–touch” balance and the theoretical perspective of “Service Encounter 2.0” to address these issues (Larivière et al., 2017).



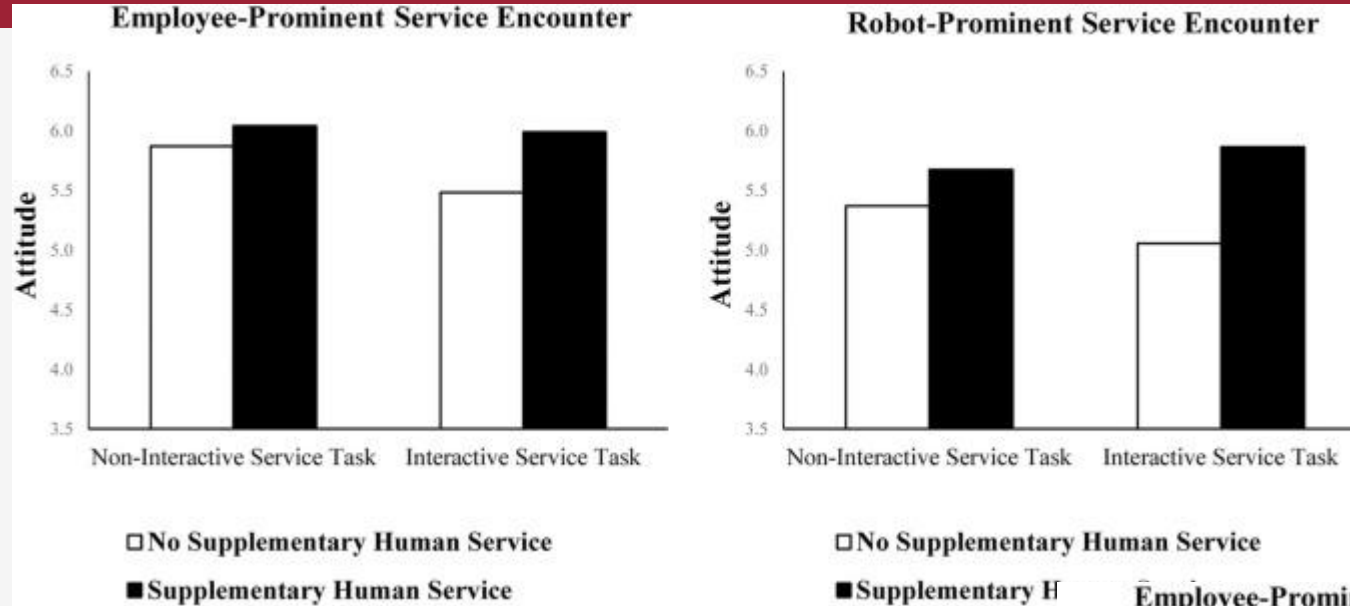
H1: In robot-prominent service encounters where robots are programmed to deliver interactive focal service tasks, supplementary human service will enhance consumers' (a) attitudes and (b) repurchase intentions towards the service business.

H2: In robot-prominent service encounters where robots are programmed to deliver non-interactive focal service tasks, supplementary human service will not influence consumers' (a) attitudes and (b) repurchase intentions towards the service business.

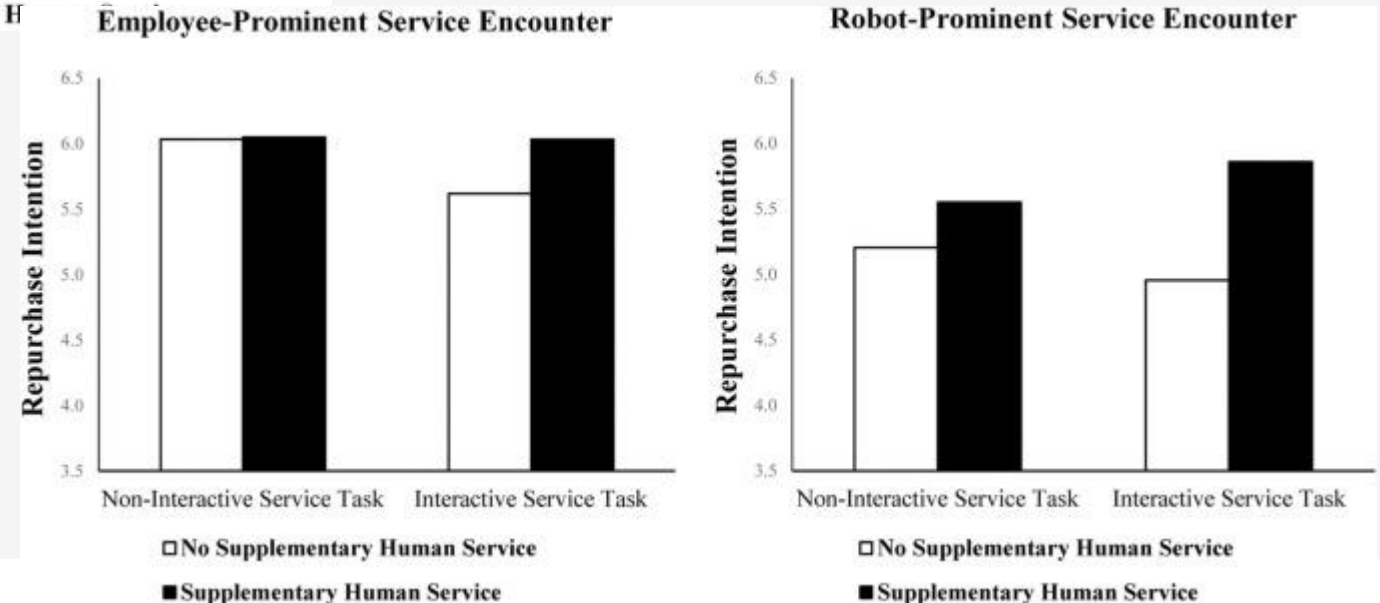
H3: In robot-prominent service encounters, perceived social value will mediate the focal service task–moderated effect of supplementary human service on consumers' (a) attitudes and (b) repurchase intentions towards the service business.

H4: In employee-prominent service encounters, the focal service task–moderated effect of supplementary human service on consumers' (a) attitudes and (b) repurchase intentions towards the service business will be attenuated.

6. Experiment study: robotic involvement



The presence (vs. not) of supplementary human service (in the form of managerial follow-up) could lead to stronger attitude and repurchase intention towards a service business in robot-prominent service encounters where robots were programmed to complete interactive service tasks but not when programmed to complete non-interactive service tasks (H1 and H2 supported).



Other popular areas of studies

- Tourism demand forecasting and revenue management
- Tourist spatial trajectory mining and itinerary recommendation
- Emotion and sentiment mining of service experience
- Service failure prediction and remedy

Future research

- Four research priorities are suggested: designing beneficial AI, facilitating adoption, assessing the impacts of intelligent automation, and creating a sustainable future with artificial intelligence (Iis, 2020).
- Dictionary specific to the tourism and hospitality context.
- Field experiment on AI applications.

Contact

Dr. Yang Yang

Associate Professor

Executive Director of The U.S.-Asia Center for Tourism and Hospitality Research

Email: yangy@temple.edu

https://scholar.google.com/citations?user=DeQT_UAAAAJ&hl=en