Transportation Research Part F 79 (2021) 35–48

Contents lists available at ScienceDirect

Transportation Research Part F

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

Lessons learned from pedestrian-driver communication and yielding patterns

Xiaoqiang "Jack" Kong [a,*], Subasish Das [b,*], Yunlong Zhang [a], Xiao Xiao [a]

[a] Texas A&M University, 3135 TAMU, College Station, TX 77843-3135, United States
[b] Texas A&M Transportation Institute, 3500 NW Loop 410, San Antonio, TX 78229, United States

ARTICLE INFO

Article history:
Received 23 November 2020
Received in revised form 2 March 2021
Accepted 16 March 2021

Keywords:
Pedestrian driver interaction
Naturalistic driving
Tacit communication
Pattern recognition
Correspondence analysis

ABSTRACT

Understanding the hidden patterns of tacit communication between drivers and pedestrians is crucial for improving pedestrian safety. However, this type of communication is a result of the psychological processes of both pedestrians and drivers, which are very difficult to understand thoroughly. This study utilizes a naturalistic field study dataset and explores the hidden patterns from successful and failed communication events using a pattern recognition method known as Taxicab Correspondence Analysis (TCA). The successful communication scenarios indicate the combinations of variable attributes such as eye contact, facial expression, the assertion of crossing, and effective traffic control devices are strongly associated with successful scenarios. The patterns for failed scenarios are most likely to be on the roadway with a relatively higher speed limit (e.g., 35 mph) and a relatively lower speed limit (e.g., 15 mph) under different conditions. On roadways with a higher speed limit, the failed scenarios are highly associated with passive and undecisive pedestrians, pedestrians far away from the crosswalk, regardless of pedestrian-driver eye contact and facial expression of the pedestrians. Instead of waiting for pedestrians to making a crossing decision, overspeeding drivers are more likely to speed up and pass the crosswalk. On roadways with a lower speed limit, the failed scenarios are often associated with distracted pedestrians, vehicles having the right of way, and the absence of effective traffic control devices. These findings could help transportation agencies identify appropriate countermeasures to reduce pedestrian crashes. The findings on driver-pedestrian communication patterns could provide scopes for improvement in computer vision-based algorithms designed for autonomous vehicle industries.

1. Introduction

Pedestrian safety has been a long-existing field studied by transportation researchers due to pedestrians' vulnerability in crashes. Based on the WHO Global Status Report on Road Safety 2018, out of 1,354,840 road users have died in 2017, 311,614 (about 23%) are pedestrians (World Health Organization, 2018). According to the National Highway Traffic Safety Administration (NHTSA, 2020), over 6,000 pedestrians were killed in traffic crashes in the U.S. Increasing pedestrians' safety by

Abbreviations: CA, Correspondence Analysis; TCA, Taxicab Correspondence Analysis; DAS, data acquisition system; VI, variable importance; VIP, variable importance plot; RF, random forest; GB, gradient boosting; ROW, right of way.

* Corresponding authors.
E-mail addresses: x-kong@tti.tamu.edu (X. "Jack" Kong), s-das@tti.tamu.edu (S. Das), yzhang@civil.tamu.edu (Y. Zhang), xx1991@tamu.edu (X. Xiao).

https://doi.org/10.1016/j.trf.2021.03.011
1369-8478/

decreasing the potential risk of crashes is an essential task in traffic safety research. It is also important to consider transportation efficiency and the environmental impacts of interactions between pedestrians and vehicles. Studies have shown that interruptions from pedestrians could increase traffic congestion and fuel consumption (Li & Sun, 2014). Therefore, to decrease potential crashes, improve transportation efficiency, and ensure pedestrian safety, especially for non-signalized intersections, successful communication between pedestrians and drivers is essential.

It is reported that nearly one-third of crashes occur at non-signalized crosswalks (Olszewski, Szagała, Wolański, & Zielińska, 2015). About 78% of pedestrian crashes occurred at a non-intersection crossing in the U.S. (Gómez et al., 2011). At non-signalized crosswalks, interactions between motor vehicles and pedestrians are inevitable. The interaction between the pedestrian and drivers is often called tacit communication, referring to the unspoken and inferred communication. The complex psychological process that involves humans understanding and reasoning the combination of facial expressions, eye contact, and other visible factors make this process difficult for transportation researchers and engineers to decipher. Moreover, as autonomous vehicles become more prevalent in the near future, the absence of a "driver" may create another barrier for successful pedestrian-driver communication, and a 100 percent yielding rate to any pedestrians (e.g., pedestrians just stand on sidewalks with no intention of crossing the crosswalk yet) may not be an efficient approach. Thus, understanding communication patterns between pedestrians and drivers could greatly benefit future autonomous vehicle design and allow them to present proper decisions.

Researchers have concluded that the failure to follow the law of crossing and yield behaviors are strongly associated with pedestrian-involved crashes (Bella & Silvestri, 2015; Mitman, Cooper, & DuBose, 2010). From another perspective, these failures are the failure of communications. The passengers chose to cross the road without confirmation from the drivers, or the drivers chose not to yield without the affirmation from pedestrians. The study points out that eye contact and hand gesture are the most common and effective approaches affecting the success of tacit communication. Other factors may also post impacts on the communication result, such as the presence of traffic control stop signs, driver's path impairment, pedestrian's distance to the crosswalk, and etc. (Roediger & Hickman, 2019). However, Existing studies mainly focus on studying factors separately that might affect the interactions and final decision-making process. Traditional statistical models are used by the majority of studies (e.g., logistic regression). As a complex phycological and non-verbal communication approach, the pedestrian's decision on crossing the road or not and drivers' decision on yielding or not are most likely affected by a combination of influential factors, rather than one or two factors. The traditional statistical method may good for identify marginal effects of individual variables or interaction effects among two or three variables. However, it is not suitable or optimal to use these traditional statistical methods to identify patterns that may involve more than three variables. Thus, the patterns associated with the failed and successful communication scenarios are not investigated. The purpose of this study is to bridge this gap. This study utilizes naturalistic field study data and performs a pattern recognition algorithm to mine the hidden patterns for successful communication scenarios and failed communication scenarios. By comparing the hidden patterns from the successful and failed communication scenarios, the results could provide insightful findings and help to develop countermeasures from various perspectives: law enforcement, user education or training, infrastructure, and vehicle design.

The following sections are presented in this order. In the next section, existing studies on this subject are reviewed and summarized. The following section is the methodology section that provides an overview of the dataset and states the approach of selecting variables and provides an introduction of the TCA method. The next section provides the results of the analysis conducted for successful and failed communication patterns. The conclusion section re-states the objectives, results, and potential future applications.

2. Literature review

At an intersection, a stop sign is placed if no signal is used. When no stop sign is used at a non-signalized intersection, the right hind right-hand priority rule is used (Elvik, Høye, Vaa, & Sørensen, 2009). However, usually, no stop sign is located at non-signalized crosswalks. The yielding has been an important behavior of drivers that can ensure traffic safety. According to previous studies, a significant share (about one-third) of drivers does not yield to pedestrians at crossings (Sucha, Dostal, & Risser, 2017). Investigating the factors associated with the yielding behavior at the non-signalized crosswalk has always been necessary and crucial in improving traffic safety. The research uncovering the factors associated with the yielding behavior of drivers when confronted with pedestrians helps better operation, design, and planning of traffic facilities. For example, the previous findings have been used for the designs for better communications between automounts vehicles and pedestrians that convey the awareness and intent for autonomous vehicles (Mahadevan, Somanath, & Sharlin, 2018).

An earlier study shows that the factors associated with yield behavior from three aspects: driver, pedestrian, and environment (Himanen & Kulmala, 1988). Characteristics of drivers are studied in previous research. For example, drivers of SUVs or pickups are less likely to yield compared to those who drive conventional cars (Figliozzi & Tipagornwong, 2016). Speed is an important factor that influences the yielding behavior. Studies show that with the increase in speed, the behaviors of yielding become less frequent (Bertulis & Dulaski, 2014). Except for speed, other vehicle dynamics such as deceleration rates and the platooning status decrease the frequency of yielding (Schroeder & Rouphail, 2011). When the divers just drive from a freeway or pass an upstream traffic light without stopping, the rate of yield is also lower (Figliozzi &

Tipagornwong, 2016). When the driver is driving at the nearside curb lane, the rate of yielding is low compare to when driving on other lanes (Stapleton, Kirsch, Gates, & Savolainen, 2017).

The characteristics of pedestrians are also very significant in the issue. For example, when the pedestrian stands in the street and behaves assertively, the rate of yielding is higher (Schneider, Sanatizadeh, Shaon, He, & Qin, 2018). The yielding is more likely to occur when the pedestrian is closer to the curb while waiting (Al-Kaisy, Miyake, Staszcuk, & Scharf, 2018). The approaching direction of pedestrians is also studied. Pedestrian from the opposite sidewalk gains more compliance from drivers than that from the near sidewalk (Gorrini, Crociani, Vizzari, & Bandini, 2018). Smile has been investigated as a characteristic of pedestrians that influencing the driver's behavior in a positive direction by slowing the speed and increasing the number of stops of drivers at intersections with or even without pedestrian crossing (Gueguen, Eyssartier, & Meineri, 2016). Surprisingly, in general, eye contact and gestures are not significant in influencing the decisions of drivers (Dey & Terken, 2017). Some gestures, such as L-bent-level, are found to be efficient in shaping drivers to yield (Zhuang & Wu, 2014). Besides the factors discussed above, other social demographical factors may also introduce impacts on pedestrians' behaviors, such as educational levels, marriage status, gender, age, and perception of their safety (Guo, Wang, Meng, Wang, & Liu, 2019; Puttawong & Chaturabong, 2020). Moayedi, Kheyroddin, and Shieh (2019) showed that the pedestrian-orientation of urban areas is associated with an increase in communication patterns by influencing social trust and solidarity. These demographic factors are not considered in this study due to limited data, but they are worth mentioning.

Finally, the characteristics of the environment, including driving conditions, are associated with yielding or compliance behaviors in many studies. When the roadway had a lower speed limit or less traffic, the rate of yielding is higher, according to empirical study. The density of the traffic, the pedestrian flows, and the speed of approaching vehicles have a negative relation with the rate of yielding (Sucha et al., 2017). The geometric design of the intersection is another environmental factor. When the intersection had a shorter crossing distance or a bus stop, drivers are more likely to yield (Schneider et al., 2018). Four design elements (i.e., the existence of stripes, color, texture, and a visually narrow or raised road) can increase the rate of yielding at courtesy crossings (Anciaes, Di Guardo, & Jones, 2020). The treatment of crosswalks can influence yielding behavior. The rate of yielding is positively related to the existence of crosswalk marking, additional devices such as rectangular rapid flashing beacon (RRFB), pedestrian hybrid beacon (PHB), or in-street R1–6 signs (Al-Kaisy et al., 2018; Stapleton et al., 2017). The rate of yielding is also found to be higher nearby a university campus (Al-Kaisy et al., 2018).

The studies analyzing the influential factors for yielding behavior are listed in Table 1. According to this table, majority of the previous studies use logistic regression or statistical analysis. Half of these datasets are obtained through observational field studies or simulations. With the development of data acquisition technology, more data are available from the field. Machine learning technology can help to uncover the hidden relations among many factors. Studies that involve comprehensive variables to yield behaviors are very few. Therefore, to fill in this gap, this paper investigates the problem comprehensively using machine learning tools considering all the factors from the side of drivers, pedestrians, and the environment at the same time.

3. Methodology

3.1. Dataset overview

The dataset is obtained through Safe-D Dataverse (Roediger & Hickman, 2019). The data were collected through a data acquisition system (DAS) installed on the experimental vehicles by the Virginia Tech Transportation Institute (VTTI). Raw data collected through the DAS system includes video, audio, and kinematic information. After processing the information and cleaned by VTTI. The final data consists of 1808 pedestrian and driver interactions and 97 different variables describing the details of the driving environment as well as the characteristics and reactions of drivers and pedestrians. Out of 97 variables, 45 variables were selected first for a feature selection training process using tree models based on existing studies.

Table 1
Data and Methods use in Past Studies.

Publication	Data type	Methodology
(Gómez et al., 2011)	Observational Field study data	Simulation
(Bertulis & Dulaski, 2014)	Naturalistic field study data	Statistical analysis
(Zhuang & Wu, 2014)	Observational Field study data	Statistical analysis
(Liu, Lu, Wang, & Zhang, 2014)	Naturalistic field study data	Classification tree method
(Figliozzi & Tipagornwong, 2016)	Naturalistic field study data	Logistic regression
(Gueguen et al., 2016)	Observational Field study data	Hypothesis testing
(Stapleton et al., 2017)	Naturalistic field study data	Logistic regression
(Dey & Terken, 2017)	Observational Field study data	Statistical analysis
(Sucha et al., 2017)	Naturalistic field study data	Logistic regression
(Al-Kaisy et al., 2018)	Observational Field study data	Hypothesis testing
(Schneider et al., 2018)	Naturalistic field study data	Logistic regression
(Gorrini et al., 2018)	Naturalistic field study data	Evidence-based approach
(Anciaes et al., 2020)	Naturalistic field study data	Statistical modeling, Before-after analysis

3.2. Research approach

Fig. 1 shows a flow chart to explain the methods used in this stusy. There are 95 variables available in the original dataset. Using all available variables for the analysis may compromise the efficiency of the training process and the clarity of the results. Three tress models are applied to select the most relevant variables using the feature importance for classifying events into failed and successful communication scenarios. The dataset is divided into two separate datasets in terms of the failed and successful scenarios to perform pattern recognition through dimension reduction. Therefore, the patterns for failed and successful communication scenarios are mined and discussed (see Fig. 2).

3.3. Variable selection

The 45 variables concerned in this paper are described in Table 2. These variables provide a comprehensive description of the driving environment. They include the variables reflecting driver characteristics, such as driver reaction; the variables reflecting pedestrian characteristics, such as pedestrian position and path; the variables showing the roadway conditions, such as traffic control and traffic flow; variables showing environment such as weather, lighting. Note that many of these variables are not included in conventional traffic crash databases.

For a dataset with a large number of exploratory variables, there is a need for quantifying variable importance (VI) to understand the global contribution of each of the exploratory variables has to a model prediction on the target variable, successfully communication, and failed communication in this study. Computing VI and explaining the significance with variable importance plots (VIPs) is an essential component of data-driven analysis. This study used open-source R software *'vip'* to determine the VIP measures (Greenwell & Boehmke, 2020). Decision trees are considered the most general tool to quantify VIP. In a binary decision tree, at each node t, one can use a single predictor to divide the data into two homogeneous groups. The chosen predictor is considered the one that maximizes improvement measure i^t. The relative importance measure of predictor X is the quantity of the whole node-level squared improvements for which X has been considered the partitioning

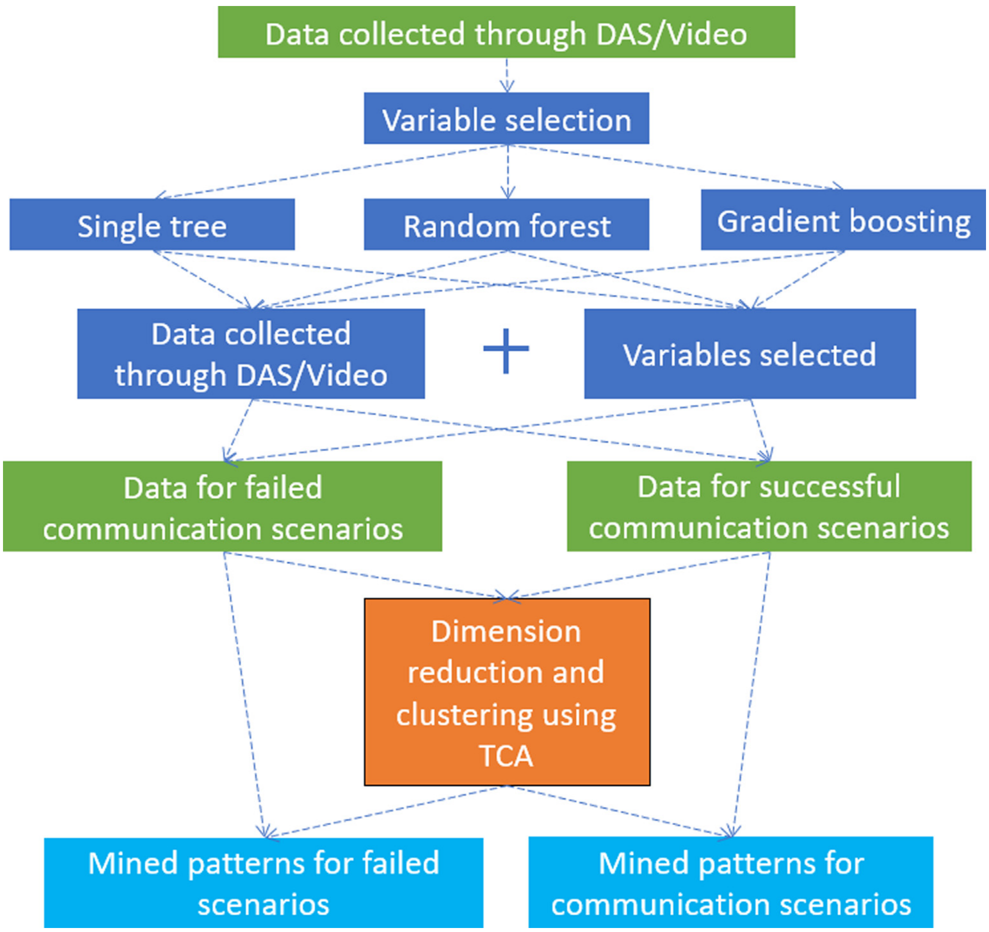

Fig. 1. Flow chart of the methodology.

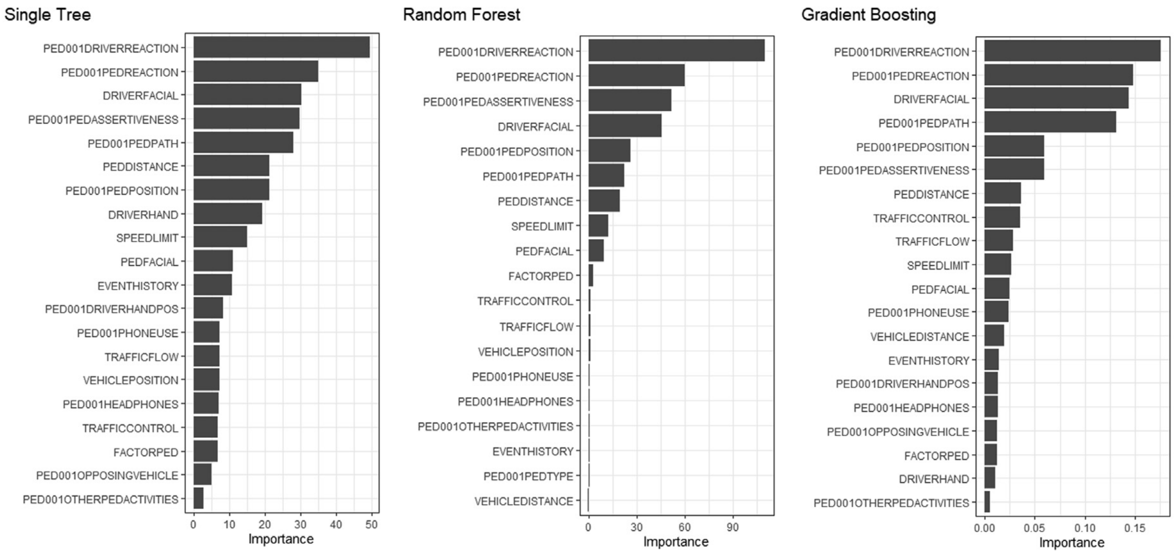

Fig. 2. Variable importance plots.

variable. Extension of ensembles of decision trees, such as random forest (R.F.) and gradient boosting (G.B.), improves the quantification measures. Fig. 1 shows the VIPs (limiting the maximum number of variables to 20) by using three modeling techniques: 1) single tree, 2) R.F., and 3) G.B. The variables common in all three plots are considered for the TCA analysis.

3.4. Descriptive statistics

After applying the tree ensemble models, sixteen features are selected for data mining. This dataset provides a total of 180 pedestrian and driver communication events. There are 1,603 successful communication events and 197 failed communication events. Successful events are defined as the vehicle successfully stopped while the pedestrian has or has no right of way (ROW), and the pedestrian successfully stopped while the vehicle has ROW. In 197 failed communication events, 167 events are pedestrian having ROW, and the vehicle failed to yield, and 30 events are vehicle having ROW and vehicle failed to yield. In this naturalistic field study dataset, the experimental vehicle only drove on the roadways with 15 mph, 25 mph, and 35 mph speed limits. Roadways with a 25 mph speed limit or less are where most of the events occurred. Table 3 lists the proportion distribution of the attributes of the selected variables for TCA analysis. To better present the study results, many variables' attribute names are renamed as shortened attribute names, as showing in Table 3.

3.5. Taxicab correspondence analysis

The method Taxicab Correspondence Analysis (TCA) was applied to perform dimension reduction and to explore patterns of variable attribute cooccurrences that are associated with yielding or not yielding behaviors.

TCA can provide interpretable results for a dataset (Choulakian, 2006; Choulakian, Allard, & Simonetti, 2013). It has been used to analyze the influencing factors for car accidents or crashes (Das, Avelar, Dixon, & Sun, 2018; Tsala, Onomo, Mvogo, & Ohandja, 2020) and pedestrian crashes (Baireddy, Zhou, & Jalayer, 2018; Sivasankaran & Balasubramanian, 2020). Compared to Correspondence Analysis (CA), TCA uses a Manhattan distance rather than a Euclidean distance. Therefore, the singular value decomposition in TCA is based on the taxicab norm, which is called taxicab singular value decomposition (TSVD). TCA is computationally efficient because it only assigns uniform weights.

TCA is applied in this paper to conduct dimensional reduction and to clarify the patterns behind the variables that are associated with yielding or not yielding behaviors. Firstly, a contingency table T with I rows and J columns is built based on the dataset with a sample size of n. Denote $X = T/n$ as the associated correspondence matrix. The element x_{ij} of X are indexed by i and j that have:

$$x_i = \sum_{j=1}^{J} x_{ij} \tag{1}$$

$$x_j = \sum_{i=1}^{I} x_{ij} \tag{2}$$

Table 2
Description of Variables.

Variable Name	Explanation	Variable Name	Explanation
EventHistory	Did any of the following occur before the current event	**Notes**	Any unusual circumstances
SpeedLimit	What is the speed limit on the current road	**PedNum**	Number each pedestrian starting at 001 restart the count within each event
TrafficFlow	How is travel divided	**Yield**	Who yielded during the interaction
ContiguousTravelLanes	How many lanes for same direction travel were available?	**PedType**	What category does the target pedestrian best fit
ThroughTravelLanes	How many lanes for same direction travel were available	**PedPosition**	What is the pedestrian's position at start of possible interaction for pedestrian interaction
Bicyclelane	Was a lane available for bicycle travel	**PedPath**	What is the pedestrian's intended path at start of possible interaction
VisualObstructions	What environmental factors impaired the driver's ability to see a possible pedestrian interaction	**PedReaction**	What is the type of reaction (task change) that occurred by the pedestrian in response to the experimental vehicle
Weather	What type of weather is applicable to the experimental vehicle at event start	**DriverReaction**	What is the type of reaction (task change) that occurred by the driver in response to the pedestrian
Lighting	What type of ambient lighting is applicable to the experimental vehicle at event start	**PedClothing**	What color of clothing was the pedestrian wearing
TrafficControl	What type of traffic control is applicable to the experimental vehicle at event start	**PedGender**	What is the pedestrian's gender
PedestrianSignal	What type of signal is applicable to the pedestrian at event start	**PedAge**	What age group does the pedestrian best fit
RightofWay	Who has the right-of-way for the impending encounter	**Dependency**	What is the pedestrian's situation regarding someone else's dependency on them or their own vulnerability
VehiclePosition	What type of scenario is the current or impending experimental vehicle pedestrian encounter	**PhoneUse**	What type of phone use is occurring if any by the pedestrian
VehicleSpeed	How fast is the vehicle traveling at the start of the event	**Headphones**	Is the pedestrian using headphones of any type
VehicleDistance	How far away is the vehicle from the crosswalk at event start	**Carrying**	Is the pedestrian carrying anything
PedinCrosswalk	How many pedestrians were in the crosswalk at event start	**EatDrink**	Is the pedestrian eating or drinking
PedNearCrosswalk	How many pedestrians were within 0–10 feet of the crosswalk at event start	**Jogging**	Is the pedestrian running for exercise
PedDistance	How far away is the nearest pedestrian from the crosswalk at event start	**OtherPedActivities**	Is the target pedestrian doing any of the following actions
UtteranceCoding	Transcribe any verbal comments or utterance made by the driver	**PedFacial**	What facial gesture(s) if any did the pedestrian make toward the driver
Coached	Did the participant receive any coaching for the think-aloud protocol	**PedHand**	What hand gesture(s) if any did the pedestrian make toward the driver
DriverFacial	What facial gesture(s) if any did the driver make toward the pedestrian	**PedAssertiveness**	Rate your perception of the pedestrian's assertiveness on a scale from 1
FactorPedPath	What are the factors affecting the pedestrian's path and behavior	**DriverHand**	What gesture(s) if any did the driver make toward the pedestrian
FactorDrivePath	What are the factors affecting the driver's path and behavior	**OpposingVehicle**	Was a vehicle traveling in the opposing lane at any point during the event which could have interacted with the target pedestrian
PedCross	In relation to the experimental vehicle when does the pedestrian cross	**DriverHandPos**	Where were the driver's hands during the pedestrian encounter
		AnotherPed	Is there another ped to code

Define the vector $r = (x_i)$ and$c = (x_j)$ and their diagonal matrix $D_r = Diag(r)$ and $D_c = Diag(c)$, respectively. Define X_0 as the origin X and X_k shows the residual correspondence matrix at the kth iteration. The number of iterations k is $rank(X) - 1$. Let $R_0 = {D_r}^{-1}X_0$ be the row profiles, in which elements are denoted as r_{ij} and $C_0 = {D_c}^{-1}{X_0}'$ be the column profiles, in which elements are denoted as c_{ij}.The optimization problem in TCA is to find the principle axis u of r row points considering the weighted L_1 dispersion of the projection of the row profiles r_{0i} on the principal axis u. The optimization problem is similar to find the principle axis v of c column points:

$$max\|X_0u\|_1 = max\|D_rR_0u\| \tag{3}$$

$$\text{s.j.}\|u\|_\infty = 1$$

$$max\|{X_0}'v\|_1 = max\|D_cC_0v\|_1 \tag{4}$$

Table 3
Proportions of Attributes by Communication Patterns.

Variable and Attributes	Shortened Attribute Name	Ped ROW Veh Fail to Yield (N = 167)	Successful Commu. (N = 1603)	Veh ROW Veh Fail to Yield (N = 30)
EVENTHISTORY (Event)				
Vehicle stopped/slowed for infrastructure	Stopped_for_infrastructure	13 (7.78%)	167 (10.4%)	4 (13.3%)
Vehicle stopped/slowed for non-experimental vehicle	Stopped_for_vehicle	16 (9.58%)	323 (20.1%)	3 (10.0%)
Other	Others	8 (4.79%)	178 (11.1%)	1 (3.33%)
None	None	130 (77.8%)	935 (58.3%)	22 (73.3%)
SPEEDLIMIT (PSL)				
15 mph (Drillfield)	15mph	24 (14.4%)	387 (24.1%)	5 (16.7%)
25 mph (other town roads)	25mph	143 (85.6%)	1208 (75.4%)	14 (46.7%)
35 mph (Patrick Henry Drive)	35mph	0 (0.00%)	8 (0.50%)	11 (36.7%)
TRAFFICFLOW (Flow)				
One-way traffic	1_waytraffic	21 (12.6%)	387 (24.1%)	7 (23.3%)
Not divided - simple 2-way trafficway	2_waytraffic	118 (70.7%)	944 (58.9%)	7 (23.3%)
Divided (median strip or barrier)	Divided	24 (14.4%)	227 (14.2%)	13 (43.3%)
Others	Others	4 (2.40%)	45 (2.81%)	3 (10.0%)
TRAFFICCONTROL (Control)				
Yield to pedestrian sign	Ped_sign_yield	64 (38.3%)	582 (36.3%)	0 (0.00%)
Stop sign	Stopsign	2 (1.20%)	92 (5.74%)	0 (0.00%)
Yield to traffic sign	Traffic_sign_yield	0 (0.00%)	7 (0.44%)	0 (0.00%)
None	None	101 (60.5%)	918 (57.3%)	30 (100%)
Other	Others	0 (0.00%)	4 (0.25%)	0 (0.00%)
PEDDISTANCE (PedDist)				
0 feet	0ft	31 (18.6%)	186 (11.6%)	16 (53.3%)
0–5 feet	0_5ft	68 (40.7%)	600 (37.4%)	11 (36.7%)
10 + feet	10 + ft	1 (0.60%)	12 (0.75%)	0 (0.00%)
5–10 feet	5_10ft	60 (35.9%)	162 (10.1%)	3 (10.0%)
In Crosswalk	In_crosswalk	7 (4.19%)	643 (40.1%)	0 (0.00%)
PED001PEDPOSITION (PedPos)				
Approaching curb to cross	Approaching_curb_cross	128 (76.6%)	772 (48.2%)	9 (30.0%)
In roadway, crossing (before vehicle trajectory)	Crossing_before_vehi_trajec	0 (0.00%)	113 (7.05%)	0 (0.00%)
In roadway, crossing (in vehicle trajectory)	Crossing_in_vehi_trajec	4 (2.40%)	583 (36.4%)	0 (0.00%)
Waiting on curb to cross	Waiting_curb_to_cross	32 (19.2%)	124 (7.74%)	21 (70.0%)
Other	Others	3 (1.80%)	11 (0.69%)	0 (0.00%)
PED001PEDPATH (PedPath)				
Crosswalk	Crosswalk	163 (97.6%)	1508 (94.1%)	4 (13.3%)
Roadway (no crosswalk)	No_crosswalk	4 (2.40%)	95 (5.93%)	26 (86.7%)
PED001PEDREACTION (PedReac)				
Continue, but accelerate	Continue_accel	1 (0.60%)	111 (6.92%)	0 (0.00%)
Continue, but decelerate	Continue_decel	40 (24.0%)	4 (0.25%)	5 (16.7%)
Interrupted walking and abort	Interrupted_abort	29 (17.4%)	10 (0.62%)	4 (13.3%)
Interrupted walking, then continued	Interrupted_continue	10 (5.99%)	183 (11.4%)	0 (0.00%)
Other	Others	0 (0.00%)	22 (1.37%)	0 (0.00%)
Begin walking from stationary	Walk_from_stationary	0 (0.00%)	106 (6.61%)	0 (0.00%)
None	None	87 (52.1%)	1167 (72.8%)	21 (70.0%)
PED001DRIVERREACTION (DrvReac)				
Comes to a complete stop	Complete_stop	0 (0.00%)	423 (26.4%)	0 (0.00%)
Continue, but accelerate	Continue_accel	51 (30.5%)	32 (2.00%)	1 (3.33%)
Continue, but decelerate	Continue_decel	31 (18.6%)	714 (44.5%)	5 (16.7%)
Interrupted driving, then continued	Interrupted_continue	1 (0.60%)	322 (20.1%)	0 (0.00%)
None	None	84 (50.3%)	112 (6.99%)	24 (80.0%)
PED001PHONEUSE (PedPh)				
Listening or talking	Listen_talk	4 (2.40%)	39 (2.43%)	2 (6.67%)
Looking and/or manipulating controls	Looking_manipulating	9 (5.39%)	130 (8.11%)	2 (6.67%)
Unknown	Unknown	20 (12.0%)	83 (5.18%)	5 (16.7%)
None	None	134 (80.2%)	1351 (84.3%)	21 (70.0%)
PED001HEADPHONES (PedHeadPh)				
Earbuds	Earbuds	4 (2.40%)	40 (2.50%)	1 (3.33%)
Over ear headphones	Over_ear_headph	0 (0.00%)	15 (0.94%)	0 (0.00%)
Unknown	Unknown	62 (37.1%)	474 (29.6%)	15 (50.0%)
None	None	101 (60.5%)	1074 (67.0%)	14 (46.7%)
PED001OTHERPEDACTIVITIES (PedActy)				
Looking at personal item	Looking	0 (0.00%)	12 (0.75%)	1 (3.33%)
Talking to someone in person	Talking	10 (5.99%)	138 (8.61%)	2 (6.67%)
Unknown	Unknown	1 (0.60%)	23 (1.43%)	0 (0.00%)
None	None	156 (93.4%)	1430 (89.2%)	27 (90.0%)

(*continued on next page*)

Table 3 (*continued*)

Variable and Attributes	Shortened Attribute Name	Ped ROW Veh Fail to Yield (N = 167)	Successful Commu. (N = 1603)	Veh ROW Veh Fail to Yield (N = 30)
PED001PEDASSERTIVENESS (PedAssert)				
Extremely assertive (7)	Extreme_assert	0 (0.00%)	153 (9.54%)	0 (0.00%)
Assertive (6)	Assert	2 (1.20%)	357 (22.3%)	0 (0.00%)
Somewhat assertive (5)	Some_assert	7 (4.19%)	439 (27.4%)	1 (3.33%)
Neutral (4)	Neutral	20 (12.0%)	238 (14.8%)	1 (3.33%)
Somewhat passive (3)	Some_passive	28 (16.8%)	179 (11.2%)	4 (13.3%)
Passive (2)	Passive	52 (31.1%)	142 (8.86%)	9 (30.0%)
Extremely passive (1)	Extreme_passive	58 (34.7%)	95 (5.93%)	15 (50.0%)
PEDFACIAL (PedFac)				
Eye contact	Eye_contact	123 (73.7%)	868 (54.1%)	25 (83.3%)
No facial gesture	No_facial_gesture	36 (21.6%)	589 (36.7%)	4 (13.3%)
Some facial expression	Some_facial_express	8 (4.79%)	146 (9.11%)	1 (3.33%)
DRIVERFACIAL (DrvFac)				
Eye contact	Eye_contact	74 (44.3%)	1492 (93.1%)	14 (46.7%)
No facial gesture	No_facial_gesture	91 (54.5%)	74 (4.62%)	16 (53.3%)
Smile or Head nod	Smile_or_head_nod	2 (1.20%)	37 (2.31%)	0 (0.00%)
FACTORPED (PedFac)				
No path obstruction	No_obstruction	125 (74.9%)	1225 (76.4%)	16 (53.3%)
Other pedestrians	Other_ped	3 (1.80%)	173 (10.8%)	0 (0.00%)
Other vehicles	Other_veh	35 (21.0%)	188 (11.7%)	13 (43.3%)
Other	Others	4 (2.40%)	17 (1.06%)	1 (3.33%)

s.j. $\|v\|_\infty = 1$

For the 0th principal, the u and v from the singular value decomposition is:

$$u_0 = 1_J \tag{5}$$

$$v_0 = 1_I \tag{6}$$

The representation of X in the principal component space is called principal factor scores. The 0th principal factor scores are calculated as

$$f_0 = D_r{}^{-1} X_0 u_0 = R_0 u_0 = 1_I \tag{7}$$

$$g_0 = D_c{}^{-1} X_0{}' v_0 = C_0 u_{0I} = 1_J \tag{8}$$

For v_0 and u_0, they are derived as:

$$v_0 = sgn(f_0) \tag{9}$$

$$u_0 = sgn(g_0) \tag{10}$$

The taxicab dispersion measure is calculated as:

$$\lambda_0 = \|Xu_0\| = v_k{}' D_r f_0 \tag{11}$$

For the 0th iteration, the taxicab dispersion measure is $\lambda_0 = \|X_0{}' v_0\| = 1$.

The first residual correspondence matrix is then calculated as:

$$X_1 = X_0 - \frac{X_0 u_0 v_o{}' X_0}{\lambda_0} = X_0 - \frac{D_r f_0 g_o{}' D_c}{\lambda_0} \tag{12}$$

In the next step, the $(v_1, u_1, \lambda_1, f_1, g_1)$ are calculated by repeating previous steps. These steps repeat until $(v_k, u_k, \lambda_k, f_k, g_k)$ is generated. The residual correspondence matrix in each step is written as:

$$\boldsymbol{X_k} = \boldsymbol{X}_0 - \sum_{\boldsymbol{k}=1}^{\boldsymbol{k}-1} \frac{\boldsymbol{D_r f_k g_k{}' D_c}}{\boldsymbol{\lambda_k}} \tag{13}$$

Note that the principal factor scores are centered:

$$\sum_{i=1}^{I} f_k(i) x_i = 0 \tag{14}$$

$$\sum_{j=1}^{J} g_k(j)x_j = 0 \tag{15}$$

Having developed principal factors, the visual maps in this paper are obtained by plotting the points $(f_1(i), f_2(i)), i \in I$or $(g_1(i), g_2(i)), j \in J$to show the patterns of behavior variables in successful or failed communication between pedestrians and drivers.

4. Results

This study performs the dimension reduction and clustering using the TCA method to mine the hidden patterns from this high-dimensional categorical dataset. This unsupervised algorithm shows the first two dimensions can explain about 53 percent of the dataset's variation. The following four plots are four quadrants (see Fig. 3 and Fig. 4) of one TCA plot. One dilemma with categorical data analysis is that the plot could be crowded with all category values and impede the process of revealing hidden patterns. Thus, this study visualizes the TCA plot by quadrant. The red dots in the plots are plotted with that category's coordinates on Axis 1 and Axis 2 – dimension 1 and dimension 2. The distance between the dots suggests the possible associations among them. In 7 clusters identified in the TCA plots, 4 clusters show the pattern for failed communication scenarios (Fig. 3) and 3 clusters for successful communication scenarios (Fig. 4).

4.1. Patterns for failed communication scenarios

Quadrant 1

There are two clusters in the first quadrant (see Fig. 3a). Even mining patterns through the unsupervised algorithm, cluster 1 shows two failed commutation types, the vehicle has ROW, but the vehicle failed to yield, and the pedestrian has ROW, but the vehicle failed to yield, are very close to each other. The closeness between these two failed communication types indicates that two failed communication scenarios occur during similar situations and associate with similar patterns. In cluster 1, two failed communication types are highly associated with three pedestrian reaction categories, two categories of pedestrian assertion, and a 35 mph speed limit. The combination of passive or extreme passive pedestrian and undecisive walking behaviors like being interrupted, abort, and continue may confuse both drivers and pedestrians themselves. The confusion leads to failed communications. This cluster also associates the occurrence of failed communication scenarios with a relatively higher speed limit – 35 mph. The combination of passive pedestrians, undecisive walking behaviors, and a relatively high-speed limit roadway segment intrigue frequently failed communication scenario. Drivers who drive at a relatively higher speed need more time and longer travel distance to successfully yield. Without enough time to be successfully stopped before the crosswalk, combining with the undecisive walking behaviors, drivers are likely to abort the yield behavior and choose the drive through the crosswalk as soon as possible. Moreover, many existing studies have confirmed the likelihood of fatal pedestrian-involved crashes and driving speed. In an early study conducted by Pasanen and Salmivaara (1993), the results concluded that a pedestrian fatality was eight times higher when a vehicle was traveling at 30 mph compared with 20 mph. Kröyer, Jonsson, and Várhelyi (2014) also concluded that a high risk of fatality exists at a relatively low vehicle speed – 25 mph and speed over 40 mph becomes extremely dangerous. Thus, a roadway segment with a relatively high-speed limit (e.g., 35 mph) could be extremely dangerous for passive passengers. Proper countermeasures, such as traffic control devices like a stop sign, may be introduced to protect pedestrians.

Cluster 2 associates drivers' and pedestrians' facial expressions, pedestrians' distance from crosswalks, and pedestrians' assertion of crossing the street with two failed communication types. This cluster is also close to a road characteristic – a 35 mph speed limit from cluster 1. This cluster shows the pedestrians' position and distance from the crosswalk are critical for drivers' decisions to yield or not yield for pedestrians while driving on the roadway with a higher speed limit – 35 mph. The drivers are more likely to speed up and cross the crosswalk when the passive pedestrian is not presented in the crosswalk. This pattern may happen with or without facial or eye interactions between the driver and pedestrian. This pattern again denotes the importance of pedestrians' assertation on successful communication. Schneider et al. (2018) showed that the yielding rate is higher due to pedestrians' and drivers' assertation. Another study from Al-Kaisy (2018) proved that the likelihood of a successful yielding behavior is higher if the pedestrian is closer to the curb while waiting. Cluster 2 demonstrates the dominance of the position and distance of the pedestrian to the crosswalk and the assertion of the pedestrian to cross the crosswalk among all factors that might be associated with a failed communication scenario. In other words, when the pedestrian is not close to the crosswalk and walks without strong assertion, regardless of the facial expressions from drivers and passengers, drivers are likely to choose to pass the crosswalk rather than yielding and waiting for the pedestrians. This finding may support a surprising finding in Dey and Terken (2017) research, which is that eye contact and gestures are not significant in influencing the decisions of drivers. With common knowledge, it is reasonable to believe that the eye contact and facial expression between the pedestrian and driver are important and essential to ensure successful communication. The eye contact and facial expression may not as dominant as other factors for securing successful communication, such as the pedestrian's position and distance from the crosswalk.

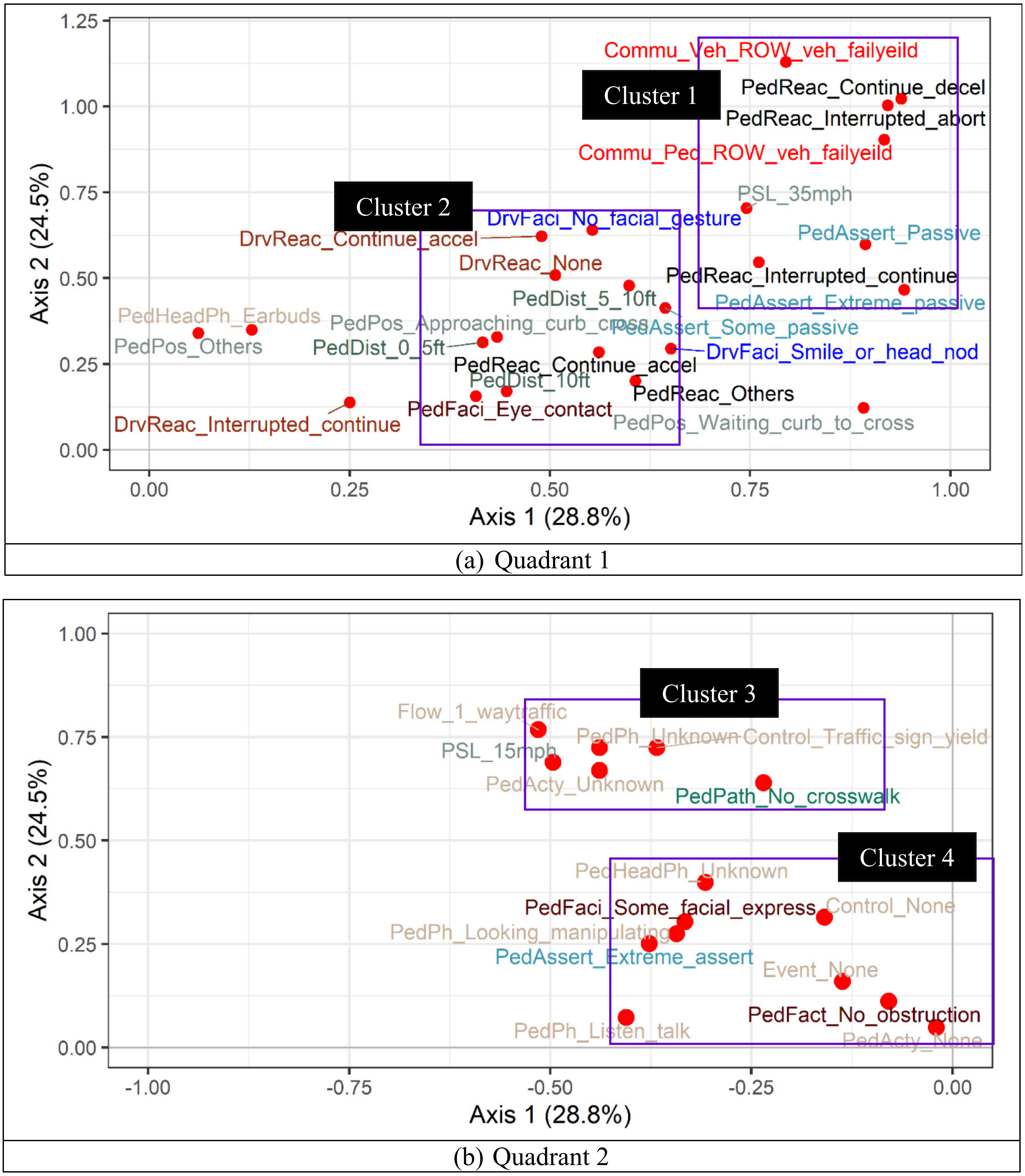

Fig. 3. Clusters of variable attributes (quadrant 1 and quadrant 2).

<u>Quadrant 2</u>

In general, quadrant 2 is associated with two failed communication scenarios in quadrant 1 since both quadrants are at the upper side of Axis 1. Two clusters, cluster 3 and cluster 4 are identified in this quadrant (see Fig. 3b). Cluster 3 associates the one-way traffic flow, yield to a traffic sign, relatively low-speed limit (15 mph), and no existence of crosswalk on pedestrians’ path with failed communication scenarios. One-way traffic flow indicates a relatively simplistic driving environment. Driving at lower speeds would boost drivers’ confidence in stopping in time when needed, especially when the driving environment is not complex. Drivers are more comfortable about not yielding for pedestrians while there is no sign to yield for pedestrians and no crosswalk presence. The pattern with the combination of a relatively lower speed limit, not-complex driving environment, no presence of traffic sign of yield to the pedestrian, and no existence of crosswalk is highly associated

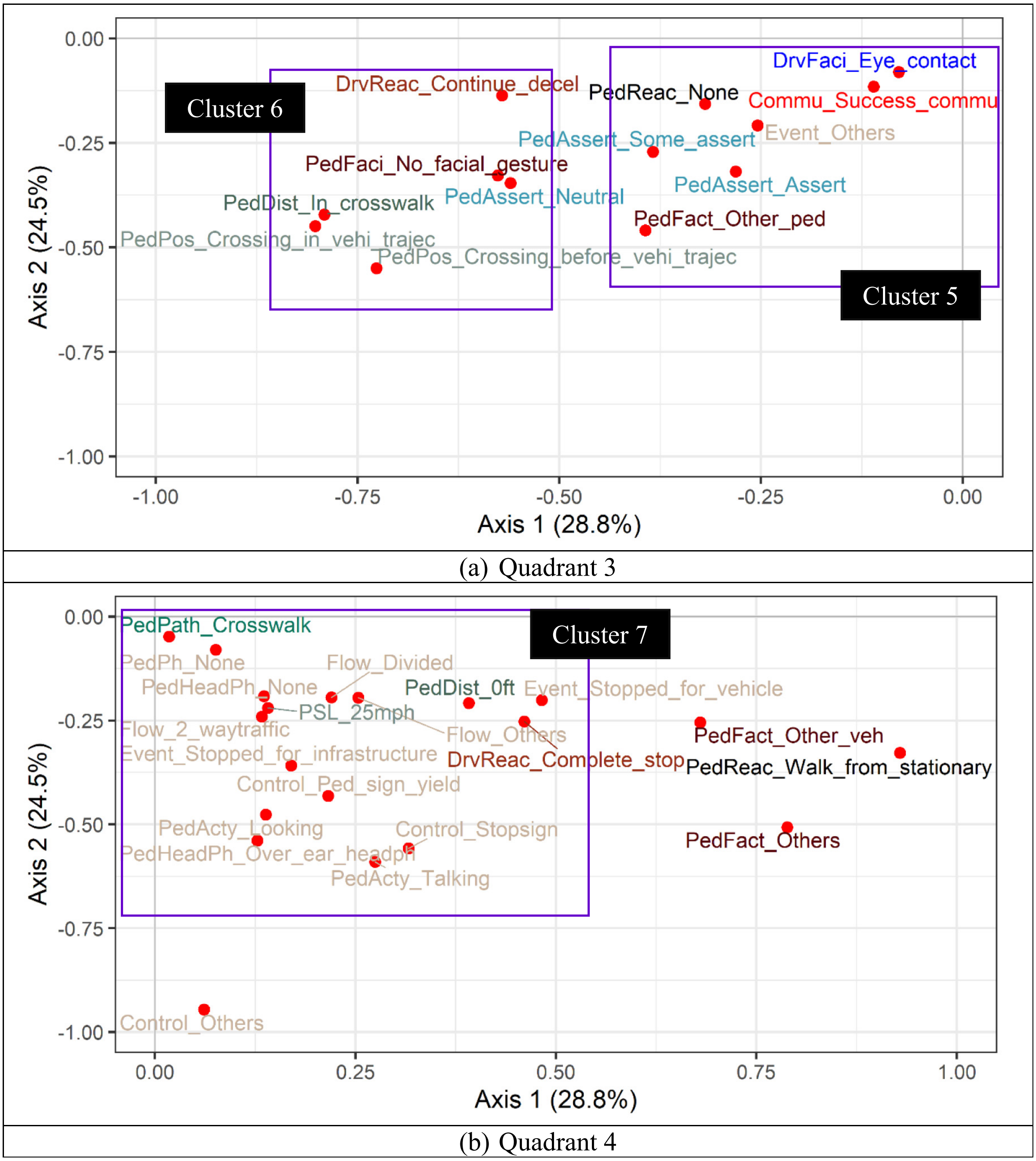

Fig. 4. Clusters of variable attributes (quadrant 3 and quadrant 4).

with failed communications. With the relatively low driving speed, the risk of crashes is low, and the probability of fatal crashes is even lower. However, the existence of this pattern suggests the high frequency of this type of failed communication occurrence in the dataset. Even the consequences are not as severe as fatal crashes, the possibility of severe consequences still nonnegligible due to the volubility of pedestrians against vehicles. Proper adjustments to this type of crosswalk should be made if the existing crash data provide sufficient evidence (e.g., pedestrian-involved crashes happened at this type of crosswalk at certain locations).

Cluster 4 is close to cluster 3. Cluster 4 also associates the failed communication scenarios with the absence of traffic control devices, pedestrians' extreme assertion of crossing the roads, and pedestrians being distracted or talking on the phone. As discussed in cluster 3, the drivers are less likely to stop for pedestrians if there are no traffic control devices and no crosswalk, especially at a relatively lower driving speed. This failed communication would often occur during these scenarios, especially when pedestrians are with the extreme assertion and distracted by their phones. The little communication between drivers and pedestrians in this pattern could introduce a high likelihood of a crash occurrence. Multiple studies have associated distracted pedestrians due to phone-using behaviors and the high risk of conflicts with vehicles (Simmons, Caird, Ta, Sterzer, & Hagel, 2020; Zhang, Zhang, Chen, & Wei, 2019). To reduce possible conflict due to distracted pedestrians, safety education for the targeted population (e.g., younger generations that are more likely to play with phones while walking) may be needed.

4.2. Patterns for successful communication scenario

Quadrant 3 and 4 introduce the patterns for successful communication scenarios. Two clusters are identified in quadrant 3 (see Fig. 4a), and one cluster is identified in cluster 4 (see Fig. 4b). The patterns and findings in these two quadrants could provide valuable suggestions for improving the failed patterns in the previous section.

Quadrant 3

In cluster 5, the closest category to the successful communication scenario is eye contact with the driver and pedestrian. This assures the importance of eye contact in the successful communication between the pedestrian and the driver. This may be contradicted with the findings in cluster 2 – eye contact and facial expression do not prevent failed communication events at first glance. However, the above discussion of cluster 2 illustrates that eye contact and facial expression may not be as dominant as other factors in some specific conditions. Except for the eye contact category, pedestrians' assertion of crossing the road and walk with other pedestrians are all highly associated with the successful communication scenario. This finding aligns with an existing study (Schneider et al., 2018). Walking with other pedestrians could attract more attention from the driver, and eye contact could further facilitate the communication process. Cluster 6 illustrates the pattern that the vehicles are more likely to completely stop for pedestrians already in the crosswalk or in the vehicle trajectory. In this scenario, a pedestrian's physical presence in the crosswalk or in the vehicle trajectory is a strong signal for the vehicle to yield. The vehicles are most likely to yield even without facial and eye contact with pedestrians. The findings from these two clusters support the patterns for failed communication scenarios from another angle. The presence of the pedestrian in the crosswalk or in the vehicle trajectory is a dominant factor associated with yield compliance.

Quadrant 4

One cluster is identified in this quadrant. This cluster's pattern associates successful communication with divided and two-way traffic flow, pedestrians' distance with the crosswalk, traffic control devices, crosswalk presence, and walking with other pedestrians (see Fig. 4b). The length of crosswalks on two-way and divided traffic roadways is relatively longer than in one-way traffic roadways. It requires a longer time to cross. Pedestrians are more cautious about stepping in this relatively longer crosswalk. The more complicated traffic conditions and longer crosswalks introduce more caution and responsibilities to pedestrians. These locations with divided two-way traffic and long crosswalks often have crosswalks and position the traffic control devices, such as yield to pedestrians and stop signs, to facilitate pedestrians to cross the roads safely. With crosswalk and traffic control devices, the drivers are more likely to be more cautious and more likely to yield for pedestrians. This cluster also associates the successful communication with pedestrians talking with other people, which echoes cluster 5 that multiple pedestrian presences at the crosswalk facilitate successful communications between the drivers and pedestrians.

In general, the hidden patterns are explored using the TCA method. Quadrants 1 and 2, including Clusters 1–4, show the factors associated with failed communication, while Quadrants 3 and 4, including Clusters 5–7, demonstrate that with successful communication. The factors associated with failed communication in Cluster 1 are related to ROW. In Cluster 2, they are drivers' and pedestrians' facial expressions, pedestrians' distance from crosswalks, and pedestrians' assertion of crossing the street. In Cluster 3, the one-way traffic flow, yield to a traffic sign, relatively low-speed limit, and no existence of crosswalk on pedestrians' paths are presented. Cluster 4 appears to be a combination of the following items: the absence of traffic control devices, pedestrians' extreme assertion of crossing the roads, and the pedestrian being distracted or talking on the phones. The factor associated with successful communication in Cluster 5 is eye contact with the driver, while in Cluster 6 it is the pedestrian's physical presence on the crosswalk or the vehicle trajectory. In Cluster 7, divided and two-way traffic flow, pedestrians' distance with the crosswalk, traffic control devices, crosswalk presence, and walking with other pedestrians are shown.

TCA, like other CA variants, provide the benefit of dimension reduction to explain datasets with large number of variables and variable attributes. Condensational statistical methods often usually supervised, and hypothesis based. Data mining method like TCA can identify hidden patterns in the complex dataset. These methods focus mostly on the co-occurrence of the variable attributes based on their relative distance in a two-dimensional plane. Statistical methods provide more weightage on the influence of a variable attribute without considering any sub-group or cluster effect, which is often biased if not conducted in a robust fashion. TCA has benefits of overcoming these issues and provide more contexts towards the co-occurrence of a relatively large dataset with presence of many variables. The performance comparisons of the CA variants can be determined the inertia or variance explanation capability. MCA was applied by using the same dataset and it explained only 33% of the inertia or variance of the dataset in the first plane. TCA explained over 53% of the variance in the first plane. This comparison indicates that TCA performs better than MCA in this study. Additionally, TCA uses a uniform weight on all observation points, which is more suitable and robust compared to other CA variants (e.g., MCA) in solving the current research problem.

5. Conclusions

Understanding the hidden patterns contributing to successful and failed driver-pedestrian communication is crucial for improving pedestrian safety and facilitating autonomous vehicles' future deployment on roadways. This study utilizes a naturalistic field study dataset to recognize the patterns for successful and failed pedestrian-driver communication patterns. An unsupervised machine learning algorithm TCA is adopted. TCA uncovers the hidden patterns in the high-dimensional categorical dataset through dimension reduction. The method uses the closeness of the categories in clusters, indicating the possible associations among categories.

This study revealed four patterns for failed communication scenarios and three patterns for successful communication scenarios. Traditional studies approached this issue by identifying contributing factors through statistical modeling. This study innovatively tackling this issue by recognizing patterns among all possible categories rather than locating individual factors. As complex psychological decision-making progress, the drivers and pedestrians communicate through non-verbal expressions and make the decision based on them. The decisions could be affected by multiple factors and resulted in successful or failed scenarios. Traditional statistical methods are not suitable or optimal for identifying complex patterns associating multiple categories. The results indicate the eye contact, facial expression, assertion, traffic control devices, pedestrians present in the crosswalk are strong combinations associated with successful communication scenarios. The findings also found the failed communications could occur on roadways with a relatively high-speed limit – 35 mph and a relatively low-speed limit – 15 mph with a different combination of categories. On roadways with a higher speed limit, the failed scenarios could happen if combining with passive and undecisive pedestrians, pedestrians far away from the crosswalk, regardless of the existence of the eye contacts and facial expressions. Instead of waiting for pedestrians to decide whether to cross or not, drivers of vehicles with higher driving speed are more likely to speed up and pass the crosswalk. On roadways with a lower speed limit, the failed scenarios are often associated with distracted pedestrians, vehicles having ROW, and the absence of the crosswalk and traffic control devices.

For the purpose of improving pedestrian safety, several countermeasures could be implemented. First, for failed communications on roadways with a high-speed limit (e.g., 35 mph), a signalized crosswalk sign might encourage drivers to yield to pedestrians and boost the assertation of the pedestrian to cross the crosswalk on a relatively high-speed limit roadway. Second, for failed communications on roadways with a low-speed limit (e.g., 15 mph), adding crosswalks and pedestrian signs could alert drivers and pay more attention to pedestrians. Meanwhile, more public education on transportation safety may be needed for certain populations to reduce phone use while crossing the roadways. More investigations are needed for this suggestion to be employed at a local level. Moreover, this study provides a comprehensive understanding of pedestrian-driver communication from successful and failed scenarios. The findings from this study could provide knowledge for future autonomous vehicle development and allow the machines to better imitate, understand, and interact with pedestrians. The insightful patterns found through data mining could also benefit many other transportation-related fields. For example, future studies could further explore the correlations between the failed communication pattern and pedestrian-involved crashes. Understanding the reasons for failed commutation patterns could help transportation agencies to mitigate pedestrian-involved crashes.

There are multiple limitations that exist in this study. One limitation is that the dataset does not contain social demographic information about the drivers. Existing studies have shown that age, marital status, and other demographic factors can also affect walking behaviors. Driver gender and distraction type should be considered in future research to mitigate this limitation. For example, Kipkorir, Ngeno, and Serem (2019) examined the perspective of different genders on smoking as a distracted driving behavior. Further research may identify more interesting patterns by including these factors. Another limitation is that the data do not contain geometric features of crosswalks, which may have an impact on crossing behavior. For example, the existence of a median and a safe island could increase pedestrians' assertion when crossing and further impact the drivers' yielding behavior. Without knowing this information, the impacts of these factors are not considered in this study.

Acknowledgment

We would like to thank two anonymous reviewers for their excellent comments and suggestions.

References

Al-Kaisy, A., Miyake, G. T., Staszcuk, J., & Scharf, D. (2018). Motorists' voluntary yielding of right of way at uncontrolled midblock crosswalks with rectangular rapid flashing beacons. *Journal of Transportation Safety & Security, 10*(4), 303–317. https://doi.org/10.1080/19439962.2016.1267827.

Anciaes, P., Di Guardo, G., & Jones, P. (2020). Factors explaining driver yielding behaviour towards pedestrians at courtesy crossings. *Transportation Research part F: Traffic Psychology and Behaviour, 73*, 453–469. https://doi.org/10.1016/j.trf.2020.07.006.

Baireddy, R., Zhou, H., & Jalayer, M. (2018). Multiple correspondence analysis of pedestrian crashes in rural Illinois. *Transportation Research Record, 2672*(38), 116–127. https://doi.org/10.1177/0361198118777088.

Bella, F., & Silvestri, M. (2015). Effects of safety measures on driver's speed behavior at pedestrian crossings. *Accident Analysis & Prevention, 83*, 111–124. https://doi.org/10.1016/j.aap.2015.07.016.

Bertulis, T., & Dulaski, D. M. (2014). Driver approach speed and its impact on driver yielding to pedestrian behavior at unsignalized crosswalks. *Transportation Research Record, 2464*(1), 46–51. https://doi.org/10.3141/2464-06.

Choulakian, V. (2006). Taxicab Correspondence Analysis. *Psychometrika, 71*(2), 333–345. https://doi.org/10.1007/s11336-004-1231-4.

Choulakian, V., Allard, J., & Simonetti, B. (2013). Multiple taxicab correspondence analysis of a survey related to health services. *Journal of Data Science, 11*(2), 205–229.
Das, S., Avelar, R., Dixon, K., & Sun, X. (2018). Investigation on the wrong way driving crash patterns using multiple correspondence analysis. *Accident Analysis & Prevention, 111*, 43–55. https://doi.org/10.1016/j.aap.2017.11.016.
Dey, D., & Terken, J. (2017). Pedestrian interaction with vehicles: Roles of explicit and implicit communication, in. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 109–113). https://doi.org/10.1145/3122986.3123009.
Elvik, R., Høye, A., Vaa, T., & Sørensen, M. (2009). *The Contribution of Research to Road Safety Policy-Making*'. The Handbook of Road Safety Measures: Emerald Group Publishing Limited.
Figliozzi, M. A., & Tipagornwong, C. (2016). Pedestrian Crosswalk Law: A study of traffic and trajectory factors that affect non-compliance and stopping distance. *Accident Analysis & Prevention, 96*, 169–179. https://doi.org/10.1016/j.aap.2016.08.011.
Gómez, R. A., Samuel, S., Gerardino, L. R., Romoser, M. R., Collura, J., Knodler, M., & Fisher, D. L. (2011). Do Advance Yield Markings Increase Safe Driver Behaviors at Unsignalized, Marked Midblock Crosswalks?: Driving Simulator Study. *Transportation research record, 2264*(1), 27–33. https://doi.org/10.3141/2264-04.
Gorrini, A., Crociani, L., Vizzari, G., & Bandini, S. (2018). Observation results on pedestrian-vehicle interactions at non-signalized intersections towards simulation. *Transportation research part F: traffic psychology and behaviour, 59*, 269–285. https://doi.org/10.1016/j.trf.2018.09.016.
Greenwell, B. M., & Boehmke, B. C. (2020). Variable Importance Plots–An Introduction to the vip Package. *R Journal, 12*, 1. https://doi.org/10.32614/RJ-2020-013.
Gueguen, N., Eyssartier, C., & Meineri, S. (2016). A pedestrian's smile and drivers' behavior: When a smile increases careful driving. *Journal of safety research, 56*, 83–88. https://doi.org/10.1016/j.jsr.2015.12.005.
Guo, Y., Wang, X., Meng, X., Wang, J., & Liu, Y. (2019). Pedestrians' Speed Analysis for Two-Stage Crossing at a Signalized Intersection. *Civil Engineering Journal, 5*(3), 505–514. https://doi.org/10.28991/cej-2019-03091263.
Himanen, V., & Kulmala, R. (1988). An application of logit models in analysing the behaviour of pedestrians and car drivers on pedestrian crossings. *Accident Analysis & Prevention, 20*(3), 187–197. https://doi.org/10.1016/0001-4575(88)90003-6.
Kipkorir, P., Ngeno, V., & Serem, A. (2019). Gender Perspective on Drivers of Cigarette Smoking: Two Part Model Approach. *SciMedicine Journal, 1*(1), 12–19. https://doi.org/10.28991/SciMedJ-2019-0101-2.
Kröyer, H. R., Jonsson, T., & Várhelyi, A. (2014). Relative fatality risk curve to describe the effect of change in the impact speed on fatality risk of pedestrians struck by a motor vehicle. *Accident Analysis & Prevention, 62*, 143–152. https://doi.org/10.1016/j.aap.2013.09.007.
Li, X., & Sun, J.-Q. (2014). Effect of interactions between vehicles and pedestrians on fuel consumption and emissions. *Physica A: Statistical Mechanics and its Applications, 416*, 661–675. https://doi.org/10.1016/j.physa.2014.09.028.
Liu, M., Lu, G., Wang, Y., & Zhang, Z. (2014). Analyzing drivers' crossing decisions at unsignalized intersections in China. *Transportation research part F: traffic psychology and behaviour, 24*, 244–255. https://doi.org/10.1016/j.trf.2014.04.017.
Mahadevan, K., Somanath, S., & Sharlin, E. (2018). Communicating awareness and intent in autonomous vehicle-pedestrian interaction, in. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–12). https://doi.org/10.1145/3173574.3174003.
Mitman, M. F., Cooper, D., & DuBose, B. (2010). Driver and pedestrian behavior at uncontrolled crosswalks in tahoe basin recreation area of California. *Transportation Research Record, 2198*(1), 23–31. https://doi.org/10.3141/2198-04.
Moayedi, M., Kheyroddin, R., & Shieh, I. (2019). Determining the Role of Pedestrian-Orientation, Concerning the Public Places: Improvement of Urban Social Capital Quality. *Civil Engineering Journal, 5*(4), 901–912. https://doi.org/10.28991/cej-2019-03091298.
NHTSA, 2020. Overview of Motor Vehicle Crashes in 2019 (Research Note No. DOT HS 813 060), Traffic Safety Facts.
Olszewski, P., Szagała, P., Wolański, M., & Zielińska, A. (2015). Pedestrian fatality risk in accidents at unsignalized zebra crosswalks in Poland. *Accident Analysis & Prevention, 84*, 83–91. https://doi.org/10.1016/j.aap.2015.08.008.
Pasanen, E., & Salmivaara, H. (1993). Driving speeds and pedestrian safety in the City of Helsinki. *Traffic Engineering and Control, 34*(6), 308–310.
Puttawong, C., & Chaturabong, P. (2020). Willingness-To-Pay for Estimation the Risk Pedestrian Group Accident Cost. *Civil Engineering Journal, 6*(6), 1064–1073. https://doi.org/10.28991/cej-2020-03091529.
Roediger, M., Hickman, J., 2019. Exploring Human-Vehicle Communication to Balance Transportation Safety and Efficiency: A Naturalistic Field Study of Pedestrian-Vehicle Interactions (02-014). http://doi.org/10.15787/VTT1/IC4KCQ.
Schneider, R. J., Sanatizadeh, A., Shaon, M. R. R., He, Z., & Qin, X. (2018). Exploratory analysis of driver yielding at low-speed, uncontrolled crosswalks in Milwaukee. *Wisconsin. Transportation Research Record, 2672*(35), 21–32. https://doi.org/10.1177/0361198118782251.
Schroeder, B. J., & Rouphail, N. M. (2011). Event-based modeling of driver yielding behavior at unsignalized crosswalks. *Journal of Transportation Engineering, 137*(7), 455–465. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000225.
Simmons, S. M., Caird, J. K., Ta, A., Sterzer, F., & Hagel, B. E. (2020). Plight of the distracted pedestrian: A research synthesis and meta-analysis of mobile phone use on crossing behaviour. *Injury Prevention, 26*(2), 170–176. https://doi.org/10.1136/injuryprev-2019-043426.
Sivasankaran, S. K., & Balasubramanian, V. (2020). Investigation of pedestrian crashes using multiple correspondence analysis in India. *International Journal of Injury Control and Safety Promotion, 27*(2), 144–155. https://doi.org/10.1080/17457300.2019.1681005.
Stapleton, S., Kirsch, T., Gates, T. J., & Savolainen, P. T. (2017). Factors affecting driver yielding compliance at uncontrolled midblock crosswalks on low-speed roadways. *Transportation Research Record*. https://doi.org/10.3141/2661-11.
Sucha, M., Dostal, D., & Risser, R. (2017). Pedestrian-driver communication and decision strategies at marked crossings. *Accident Analysis & Prevention, 102*, 41–50. https://doi.org/10.1016/j.aap.2017.02.018.
Tsala, S. A. Z., Onomo, C., Mvogo, G., & Ohandja, L. M. A. (2020). Elaboration of Explanatory Factors of Accidents in Cameroon by Factorial Correspondence Analysis. *Journal of Transportation Technologies, 10*(03), 280. https://doi.org/10.4236/jtts.2020.103018.
World Health Organization, 2018. Global status report on road safety 2018: Summary. World Health Organization.
Zhang, H., Zhang, C., Chen, F., & Wei, Y. (2019). Effects of mobile phone use on pedestrian crossing behavior and safety at unsignalized intersections. *Canadian Journal of Civil Engineering, 46*(5), 381–388. https://doi.org/10.1016/j.aap.2006.07.001.
Zhuang, X., & Wu, C. (2014). Pedestrian gestures increase driver yielding at uncontrolled mid-block road crossings. *Accident Analysis & Prevention, 70*, 235–244. https://doi.org/10.1016/j.aap.2013.12.015.