He and Ma
European Transport Research Review (2022) 14:47
https://doi.org/10.1186/s12544-022-00572-z

European Transport Research Review

ORIGINAL PAPER

Open Access

Examining the factors influencing microtransit users' next ride decisions using Bayesian networks

Jiajing He[1] and Tai-Yu Ma[2*]

Abstract

The progress of microtransit services across the world has been slower than expected due to institutional, operational, and financial barriers. However, how users' ride experiences and system attributes affects their future ride decisions remain an important issue for successful deployment. A Bayesian network approach is proposed to infer users' next ride decisions on a microtransit service based on historical ride data from Kussbus, a pilot microtransit system operating in the Belgium–Luxembourg cross-border areas in 2018. The results indicate that the proposed Bayesian network approach could reveal a plausible causal relationship between different dependent factors compared to the classical multinomial logit modeling approach. By examining public transport coverage in the study area, we find that Kussbus complements the existing public transport and provides an effective alternative to personal car use.

Keywords: Microtransit, Demand-responsive transport, On-demand mobility, Bayesian network, Next ride occurrence

1 Introduction

In past decades, an increasing number of transport network companies and public transport authorities have offered a spectrum of on-demand mobility services [1]. Despite numerous pilot studies, recent years have seen mixed results. Some operations such as Kutsusplus in Helsinki and Bridj in Boston failed to achieve financial sustainability [2, 3]. There is a need to learn from past experiences to improve business models [3]. Existing studies mainly focus on the evaluation of the impacts of microtransit services on transport systems, based on either simulation [4, 5] or post-evaluation [6, 7]. However, building a decision support tool to infer users' future ride decisions based on system attributes and users' ride experiences (e.g., delays, walking distance, in-vehicle riding time) could be more useful to the operator for a successful deployment. This study fills this gap, investigating how customers' ride experiences influence their future use of a microtransit service. An empirical study is conducted based on a recent microtransit service, Kussbus, operating in the cross-border areas of Luxembourg in 2018. The aim is to develop a tool to infer users' next ride decisions and draw insight from this pilot implementation for the future successful deployment of on-demand mobility services.

The contributions of this study are threefold. First, we present the system characteristics of Kussbus and analyze its system performance. To understand how competitive the Kussbus service is, we compare it with alternative transport modes and illustrate the inconvenience of using public transport for commuters. While Kussbus ridership increased during the study period, the service was discontinued in early 2019. To understand the reasons for this, we propose a Bayesian network (BN) approach to identify the factors affecting users' next ride decisions and their causal/correlational structure. The considered features include spatial-effect factors

*Correspondence: tai-yu.ma@liser.lu

[2] Luxembourg Institute of Socio-Economic Research (LISER), 11 Porte des Sciences, 4366 Esch-sur-Alzette, Luxembourg
Full list of author information is available at the end of the article

(i.e. origin–destination zones), customer ride experiences (e.g., fare, walking distance, in-vehicle travel time), weekly/seasonal factors, and relative travel time gain/loss compared to customers' habitual commuting modes. We compare the result obtained by the BN approach with that obtained from the multinomial logit (MNL) model and test the proposed methodology on an independent dataset based on a fivefold cross-validation scheme for users' next ride occurrence inference. Finally, we draw insights and discuss the findings of this study.

The remainder of this study is organized as follows. Section 2 reviews the literature related to factors that influence customers' willingness to use microtransit services, as well as the barriers and successful determinants of microtransit services. Section 3 presents the system characteristics of Kussbus, key performance indicators, and use cases of public transport in the studied area. In Sect. 4, a BN approach is proposed to model Kussbus users' next ride decisions and compared with the MNL model. Finally, we discuss our findings, policy insights, and methodological limitations, and offer some concluding remarks.

2 Related work

Previous studies on microtransit services have mainly focused on operation policy design [8], performance assessment [3, 6, 7, 9], and success/failure determinants [10, 11]. Due to the limited availability of data from microtransit companies, only limited studies examine service performance based on empirical trip data. For example, [6] propose an evaluation framework to analyze the performance of "Breng flex", a microtransit operating in the Arnhem-Nijmegen region of the Netherlands. The authors compare passengers' perceived trip journey times between the microtransit and fixed-route transit. They find that significant mobility improvements were observed thanks to the microtransit service. Haglund et al. [3] propose an evaluation framework to analyze the spatio-temporal distribution of Kutsuplus's rides in Helsinki. However, the relationships between users' experienced journey attributes and their future ride decisions were not investigated. Ma et al. [12] propose a stable matching approach to assess the impact of different operational policies on the ridership of a microtransit service in Luxembourg. They find that reducing in-vehicle travel time and operational costs are two key factors in improving ridership and making the service sustainable.

In the past, several studies have focused on the lessons learned from past experiences [13, 14]. A notable example is Helsinki's Kutsuplus, which was ceased due to an operational cost overrun [7]. Several authors point out that insufficient fare revenue due to low prices or ridership has led to the end of many microtransit pilots [3]. However, higher fares may lead to an unexpected decrease in ridership [15]. Westervelt et al. [14] analyze the experiences of the public–private partnership of three microtransit pilots in the United States and find that these pilots placed an emphasis on technological innovation but did not equally focus on customer needs. Volinski [1] points out that many-to-many services (i.e., door-to-door-like services) are often seen as more complex designs as they try to cover many requested origins and destinations with one route. Applying many-to-one services, which fixes one destination as the trip end, could reduce operating costs significantly [16]. Westervelt et al. [14] argue that operators should maintain users' needs as a priority when designing and implementing their services. In regards to this, Avermann et al. [10] analyze user satisfaction of demand responsive transport (DRT) systems based on an ordered logit model and survey data. They find waiting times and the perceived effort to catch the buses are two key determinants of DRT user satisfaction. Yu and Peng [17] apply a weighted Poisson regression model to analyze the relationship between built environment factors and ridesourcing demand in Austin, Texas. They find that ridesourcing demand is positively associated with land-use mix and population density. Deka and Fei [18] model ridesourcing trip frequency based on a zero-inflated negative binomial model to analyze the influence of individual socio-demographic attributes and the neighborhood effect. However, fewer studies have focused on how users' ride experience impact their willingness to continue the service due to limited empirical data availability. In addition, most studies focus on the aspect of system attributes [13, 19], the effect of users' socio-demographic attributes is mainly studied using survey questionnaires. Table 1 highlights the influence of individuals' sociodemographic attributes on DRT system usage based on the literature review. It shows that the socio-demographic factors of individuals (gender, car ownership, household income, reduced mobility, attitude towards the service and lifestyle) can affect the propensity to use DRT services. The reader is referred to a more comprehensive review on factors influencing user acceptance and use of DRT services [20].

3 Kussbus service characteristics and performance analysis

3.1 Kussbus service characteristics

Kussbus is a microtransit pilot using a fleet composed of a variety of shuttles to provide a commuting service in Luxembourg and its cross-border areas (Belgium and France). Due to the low coverage of public transport in these areas and job concentration in the city of Luxembourg, more than 70% of cross-border workers (more than 200,000 individuals in 2019 according to [25]

Table 1 Influence of individuals' socio-demographic attributes on the use of DRT system

Factor	System	Measuring methodology	Reported effect on user's ride decision
Car ownership	Ride-hailing	Survey on urban and suburban populations (USA)	Ride-hailing users have higher personal car ownership compared to transit-only users. Users substitute their personal car driving by ride-hailing [19]
Lifestyle, attitude towards public transport, perception of public transport, attachment to the car, individual characteristics (income, gender, environmental concerns)	Public transport	Interviews on the motivation and intention of individuals' mode choice (Portugal)	Individual characteristics, lifestyle, and perception of the level of service affect the use of public transport [21]
Gender, age, car ownership, household income	Rural DRT	System user survey (UK)	Female/elderly, lack of access to a car, lower household income are the restrictive motivations for their use of DRT services [22, 23]
Car access, age, mobility-related disability	Rural DRT	System user survey (Germany)	Car ownership is negatively associated with the intention to use DRT, while the elderly and people with reduced mobility are positively associated [24]

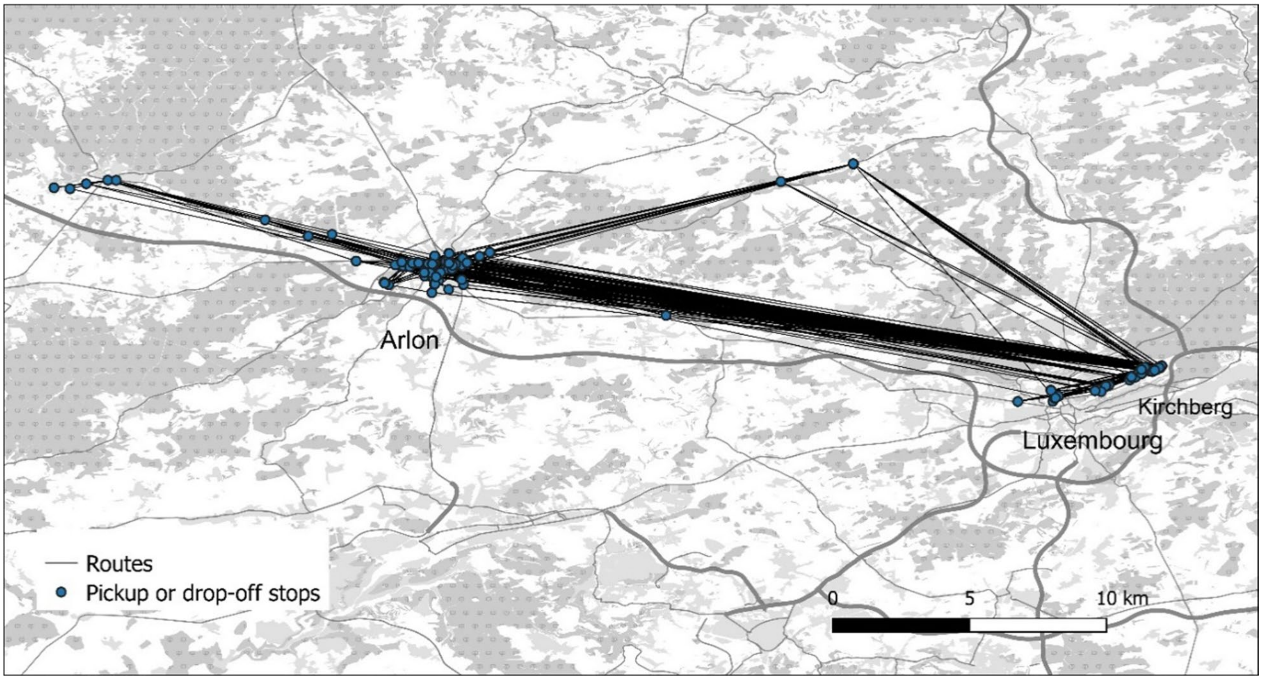

Fig. 1 Kussbus operating routes from Arlon region to Luxembourg City [12]

commute by car [26]. Consequently, 163 extra hours per driver were spent driving in rush hour in 2019 [27]. Kussbus aims to provide an alternative and flexible transit solution for these cross-border workers and reduce their personal car use.

The first Kussbus line was launched on April 25, 2018, connecting Luxembourg City (specifically, the Kirchberg district) and the Arlon region in Belgium. A second line linking the Kirchberg district to the Thionville region in France was launched later in September 2018. As there is very limited data available for the second line, our case study focuses on the first line. We summarize the features of the Kussbus service as follows.

- *Operating policy* Kussbus utilizes so-called "virtual stops" (i.e., optional bus stops within walking distance, e.g., less than 1 km.) to pool passengers into these stops near their origin/destination locations. The virtual stops are optimized based on historical ride-request data. To increase user convenience, the maximum journey time and maximum detour time are used for vehicle route planning. The service operates from 5:30 to 9:30 a.m. (Arlon → Luxembourg) and from 4:00 to 7:00 p.m. (Luxembourg → Arlon) on weekdays excluding public holidays.
- *Booking* A reservation can be made in advance or at short notice via the smartphone application of Kussbus. Users input their origin, destination, and desired pick-up time. The app will display the nearest Kussbus stop on the app, and users can track the locations of Kussbus vehicles in real-time. Notifications are sent via the app to inform users of a bus approaching, as well as delays or changes in the vehicles' routes.
- *Vehicle* A mixed fleet of shuttles with 7, 16, and 19 seats are used. Based on the user's booking information, historical ride data, and operational constraints, the operator decides which type of vehicles to use to minimize the daily operational costs.
- *Pricing* Kussbus offers 6 free rides for new users to experience the service. Afterward, each ride costs 4.95 euros and a monthly subscription is also available. Note that the Kussbus ticket price is about twice the regular Luxembourg bus ticket fare in 2018.

3.2 Kussbus ride statistics and system performance

The ride data was provided by Utopian Future Technologies S.A. for the period of April 25, 2018, to October 17, 2018. In total, 2,846 rides (trips) were realized by 134 users during the study period. Figure 1 shows Kussbus operation routes during the studied period. The evolution of weekly ridership is shown in Fig. 2. It shows several quite distinctive phases. After an initial slow period during the first two months, an initial steep increase in rides can be observed. Then there is a period of stagnation until the end of August, then the rides increase again

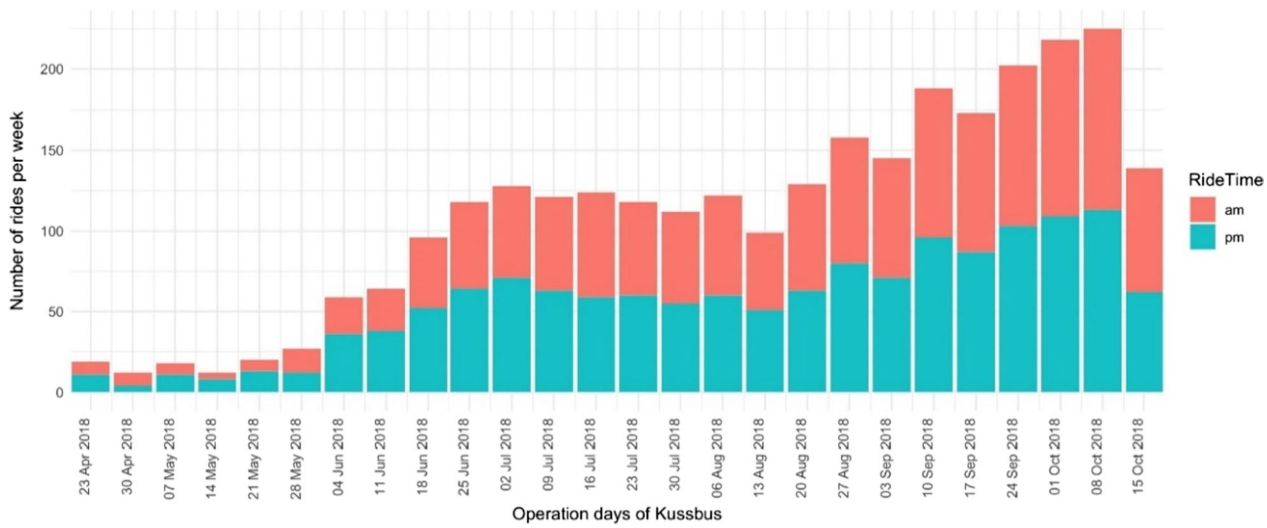

Fig. 2 Weekly ridership of Kussbus between Arlon and Luxembourg City

Table 2 Kussbus service attributes and users' journey time of other transport alternatives

Category	Indicator	Mean	S.D	Min	Max
Kussbus ridership	Number of rides per week	109.46	64.8	12.0	225.0
	Number of rides per weekday	23.7	13.7	2.0	59.0
Service attributes	In-vehicle travel time (minute)	49.2	11.4	10.0	116.1
	Total walking distance (km)	0.5	0.5	0.0	3.5
	Door-to-door travel time (minute)	54.7	12.7	11.5	122.0
	Fare (euro)	2.8	2.5	0.0	5.0
	Starting time of trips (morning)	7h11	44.8	5h48	8h33
	Starting time of trips (afternoon)	17h21	41.9	15h48	19h33
Journey time of users' alternatives	Car (minute)	42.8	6.6	25.6	67.6
	Public transport (minute)	75.2	18.0	42.7	393.5

from September due to a new advertising campaign. Each ride observation data point contains detailed information about users' trips, including latitude and longitude coordinates of users' origins and destinations, shuttle pick-up and drop-off locations (allowing for accurate measurement of walking distance), users' reservation time and pick-up and drop-off times (accurate in seconds), zip codes, country and street names, travel time, vehicle ID, passenger capacity, user ID, and fees paid. The service attributes obtained in this study are quite reliable given the accurate trip data obtained from the service application and GPS tracking of vehicles. User socio-demographic characteristics are not available for our analysis. Table 2 reports the system performance in terms of ridership, user experience, and the competitiveness of the Kussbus service compared to its alternatives (car and public transport). The average number of rides per week and weekday is 109 (=2846 rides/26 weeks, see Fig. 2) and 24 (=2846 rides / number of weekdays excluding public holidays during the studied period), respectively. The average in-vehicle travel time of Kussbus users is 48.7 min. The average walking distance to pick-up stops and drop-off stops is 0.21 and 0.25 km, respectively. Regarding the journey time of Kussbus users (54.7 min on average), it is higher than for cars (42.81 min on average). Users' travel times by car and public transport are calculated from Google's Distance Matrix API[1] by considering traffic congestion and the departure times of trips. Note that one can compute the generalized journey time as a weighted sum of different travel-time legs (i.e., walking time, waiting time, in-vehicle travel time, and the number of transfers [6] to compare their performance. In terms

[1] https://developers.google.com/maps/documentation/distance-matrix/overview.

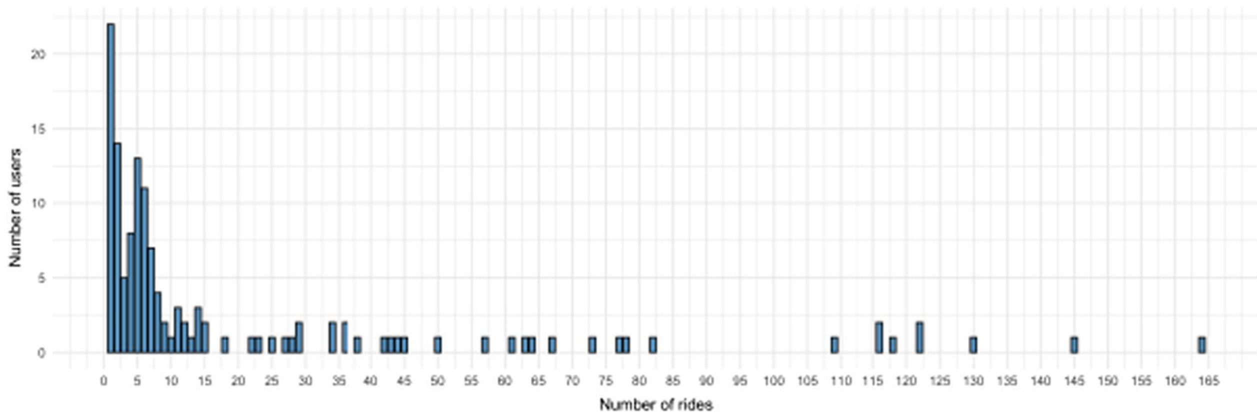

Fig. 3 Distribution of the number of rides taken by Kussbus users

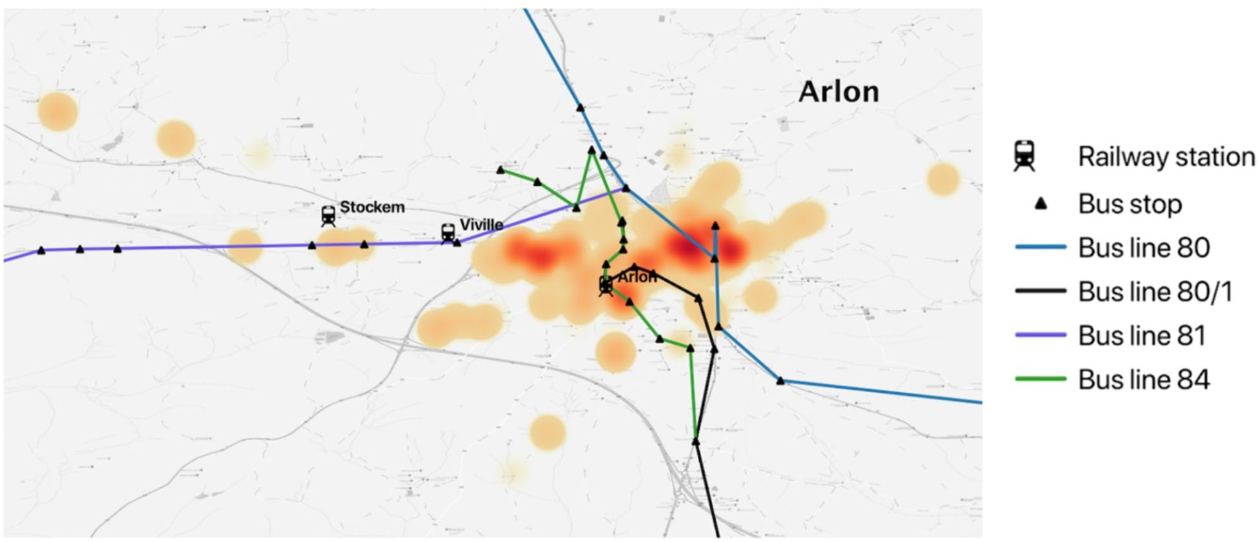

Fig. 4 Available cross-border bus lines for Kussbus users in the Arlon region (the hot spots are Kussbus ride request origins)

of the number of rides per user, Fig. 3 shows a highly skewed distribution. It shows that there are quite a few people who only used the service once, while some people used it as it was free. Apart from that, there are some regular users. A large share of customers used Kussbus for fewer than 15 rides (based on user's ID information in the dataset). Users' monthly subscription information is not available to help explain users' ride patterns.

3.3 Public transport coverage in the study area

We analyze the coverage of public transport connecting Arlon and Luxembourg City. There is one railway line and four cross-border bus lines. The train operates from Monday to Friday with a frequency of 10–20 min, with the first departure at 06:05 from Arlon station. The travel time from Arlon to the major train station in Luxembourg City takes around 30 min. Figure 4 illustrates the itineraries of these cross-border bus lines and Kussbus users' pickup locations in Arlon in the morning. We find that Kussbus users are poorly covered by these train and bus lines. Regarding bus line frequency, there is only one bus operation for lines 80, 81, and 84 every morning departing at 6:40 and 7:00. Line 80/1 departs from Arlon terminal at 05:54 and 07:40. This shows that Kussbus complements the existing public transport network in the Arlon–Luxembourg corridor. Regarding the competitiveness of Kussbus compared with public transport, Kussbus users' riding times (49.2 min on average) are lower than those of public transport (75.2 min on average) (see Table 2). Figure 5 illustrates the transit network

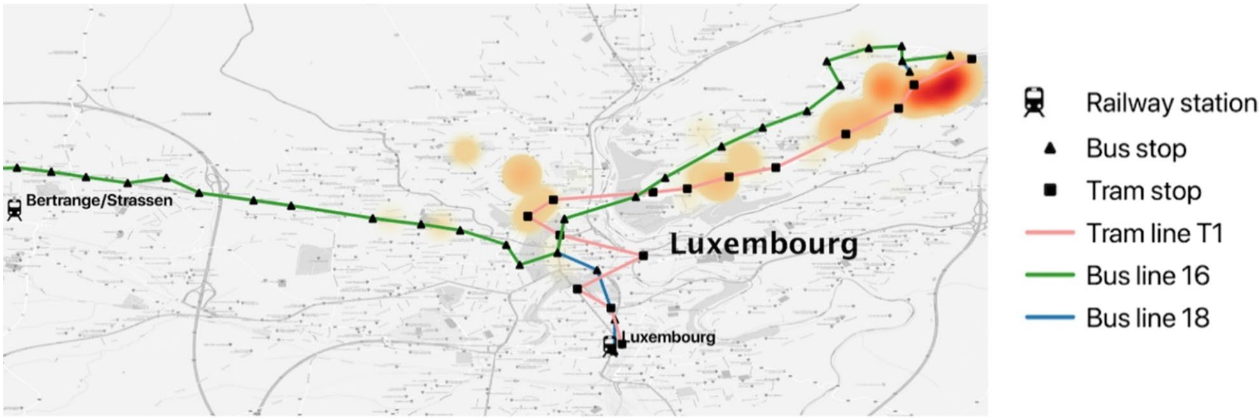

Fig. 5 Example of possible public transport options to commute from Arlon to Luxembourg City (the hot spots are Kussbus ride request destinations)

coverage for commuters arriving in Luxembourg City on public transport. For bus users, another bus transfer is required at Bertange in Luxembourg. For train users, they need to take additional bus transfers at the Luxembourg central railway station as the tram network did not cover the central station in 2018. Consequently, commuting from Arlon to Luxembourg City using public transport requires at least one transfer, with a higher travel time compared to driving or Kussbus.

4 User's next ride occurrence modeling and prediction

4.1 Factors affecting users' next ride occurrence

To understand how Kussbus users' ride experiences influence their future ride decisions, we present a BN approach to predict the next-ride occurrence of users. In Sect. 3.2, we observed that users' ride patterns are quite heterogeneous: some rarely use the service again after the first trial or after a couple of rides. As there is no socio-demographic information about the users in the data, using traditional count-based regression models performs very poorly (i.e. individual-based regression model needs the individual's socio-demographic attributes and other individual-related indicators as regressors to explain the variation in the response variable). Given the failure of Kussbus after around one year of operation, we are particularly interested in understanding how and to what extent users' ride experiences influence their ride patterns. For this reason, we apply the BN approach to model users' next ride decisions, i.e., to predict whether a user will continue to use the service or not, and their next ride occurrence if they continue. The model allows the operator to determine the factors affecting users' ride decisions and provides useful insights for improving their service. Note that it would be particularly interesting to model and compare the behavior of users who made more than 6 trips versus another group. Since the dataset used for the analysis is made by 132 users during the study period with 46.21% (61 users) made more than 6 trips. The sample size is not sufficient to fit two separate models. This could be the future extension of this work when more data is available.

A user's next ride occurrence is measured as the days elapsed between their current ride and their next ride. As the distribution of users' next ride days is highly skewed and no individual socio-demographic variables are available, directly modeling this decision variable based on the regression models results in a poor fit. Moreover, from the point of view of the operator, it is relevant to predict a user's next ride occurrence within one day, one week, or more than one week. For this purpose, we classify the next ride occurrence of users into *four* categories: within 1 day, 2–7 days, $\geq$ 7 days, and no use of the service after that ride within the studied period. The reason for setting the maximum observation period as 7 days for each ride is motivated by prior surveys on commute mode choice that over 70% of people show a habitual behavior within 7 days [28]. On the other hand, as a commuter service operator, it is important to understand why some customers use the ride service daily, weekly or lower frequency. In this case, the operator could identify the determinants of ride characteristics to improve their service over different planning horizons.

Given the available data fields, the influential factors include users' ride characteristics: pickup and drop-off locations, departure time category (peak hour or not), trip journey time, trip journey time difference with cars, walking distance, free ride or not, fare of the next ride, and whether it is their first ride or not. We also control for calendar and seasonal effects by adding determinants

Table 3 Description of the key variables of the discrete BN (N = 2,783)

Variable	Definition	Levels	%
Occurrence_nr	Category of user's next ride occurrence	≤ 1day	71.8
		2 − 7 days	19.4
		≥ 7 days	5.6
		Not utilize anymore	3.2
OD_pair	Class of user's origin–destination pair	od_pair 1 (Arlon-Limpersberg)	7.1
		od_pair 2 (Habay to Luxembourg)	2.2
		od_pair 3 (Arlon-Kirchberg)	24.5
		od_pair 4 (Arlon-Luxembourg)	61.4
		od_pair 5 (within Luxembourg)	4.8
Is_morning	1 if the ride is in the morning and 0 otherwise		49.0
Is_peak	1 if the vehicle departure time is during peak hours and 0 otherwise		70.0
Is_august	1 if the ride occurred in August and 0 otherwise		20.5
Is_holiday	1 if the day after the ride is a public holiday and 0 otherwise		19.4
Is_first	1 if the ride is a user's first ride and 0 otherwise		4.7
T_Kussbus	User's journey time using Kussbus	≤ 41	23.9
		> 41 and ≤ 52.1	43.2
		> 52.1 and ≤ 64	22.8
		> 64	10.1
T_diff	Journey time difference between Kussbus and a car (minutes)	≤ 2.5	39.8
		> 2.5 and ≤ 14.7	39.4
		> 14.7 and ≤ 25.9	15.8
		> 25.9	5.0
Walk_dist	Total walking distance between user's origin/destination and Kussbus stops (km)	< 0.4	50.8
		> 0.4 and ≤ 0.8	39.5
		> 0.8	9.7
Is_free	1 if this ride is free and 0 otherwise		42.8
Isfree_next_ride	1 if the next ride is free and 0 otherwise		41.2

related to whether the next weekday of the current ride is a holiday or not and whether it falls in August. Table 3 shows the list of key variables and their descriptive statistics. Note that the user origin and destination (OD) pair is classified into 5 categories based on geographical coordinates using the K-means clustering approach. The number of clusters is determined by inspecting whether the within-cluster sum of squares error stabilizes when increasing the number of clusters. Note that for the continuous variables (journey time using Kussbus (T_Kussbus) and the journey time difference between Kussbus and a car (T_diff)), their histograms suggest that they follow normal distributions. We perform the Shapiro–Wilk normality tests, the associated p-values are less than 0.001 suggesting the rejection of this hypothesis. The Pearson correlation coefficient between these two variables is 0.8364. As we are interested in studying the effects of in-vehicle travel time and relative delay to the user's usual mode of transportation, both variables are included in the BN model. More sophisticated independent variables could be used by considering the interaction effect between these two variables. In this study, we use Hartemink's discretization algorithm [29] to preserve the correlation structure and mutual information between these two variables to learn discrete BNs.

4.2 BN for users' next ride occurrence inference

BNs have been widely applied in different fields to uncover the casual or dependency relationships between domain variables under uncertainty. A BN is a probabilistic graphical model represented by a directed acyclic graph $G = G(X, A)$, where X is a set of random variables and A is a set of directed arcs representing the probabilistic correlations. A directed arc A_{ij} from node X_i to X_j means that X_i has a direct causal/dependent effect on X_j. We call X_i the parent of X_j and X_j the child of X_i. Based on the chain rule, the joint distribution of X over G can be expressed as Eq. (1).

$$P(X_1, X_2, \ldots, X_n) = P(X_1) \times P(X_2|X_1) \times \ldots \times P(X_n|X_1, X_2, \ldots, X_{n-1}) \tag{1}$$

Table 4 Hybrid BN structure-learning algorithm

1	**Input a set of variables** X **and the empirical ride data** D
2	Identify an expert $BN_{expert} = (X, A_{expert})$ based on expert knowledge
3	Given BN_{expert} as the structural restrictions, utilize a bootstrap sampling approach from D to learn n plausible data-driven BN structures according to the score-based learning algorithms
4	Utilize a model averaging approach to select the robust arcs with a probability greater than a statistical significance threshold (0.5 or more). This significance threshold reflects the probability that the selected arcs belong to the true (unknown) structure
5	Refine the newly added arcs from Step 4 based on the domain knowledge to obtain a final BN. This step involves checking and adjusting the directions of these newly added arcs so that they are consistent with the causal/dependent effect between the connected nodes
6	Perform parameter learning to fit the data with maximum likelihood and obtain the local distributions associated with the nodes of the final BN

In the context of BNs, the variables are conditionally independent given their parents. The conditional probability distribution of X_i can be expressed as $P(X_i|pa(X_i))$, where $pa(X_i)$ denotes the parents of X_i. The joint distribution over all variables in $G(X, A)$ can then be factorized into a product of local distributions over X as Eq. (2).

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|pa(X_i)) \quad (2)$$

The BN approach consists of learning the graphical structure and then estimating the local distributions associated with each node, given the learned network structure. The problem of finding the exact BN structure is an NP-hard problem [30]. The structure-learning algorithms can be classified into three categories: constraint-based, score-based, and hybrid learning algorithms [31]. The constraint-based algorithms utilize conditional independence tests to identify the dependency relationships between the variables and construct the graph. The score-based algorithms try to maximize a fitness score using some heuristics. The hybrid algorithms utilize the constraint-based algorithms or expert/domain knowledge to identify the partial graphical structure and then apply the score-based algorithms to maximize the fitness score, given the restricted graphical structure [31, 32]. The advantage of the hybrid algorithms is that they allow combining domain knowledge as a structural skeleton and then learning plausible structures to fit the data. Parameter learning consists of estimating the local distributions over the variables based on the maximum likelihood estimator. In this study, we apply the hybrid structure-learning algorithm using structural restrictions based on domain knowledge and model averaging to learn the BN structure so as to infer a user's next ride occurrence decision. The reader is referred to [31, 32] for a more detailed description. The hybrid algorithm is described in Table 4.

4.3 Results

4.3.1 BN structure and parameter learning

We apply the hybrid structure-learning algorithm to learn the BN structure. First, the potential dependency relationships between the variables are identified based on domain knowledge, a literature review, and a Pearson correlation matrix to study the potential dependence between these variables. The identified dependency relations between the variables are as follows.

- The pickup and drop-off locations have a direct effect on users' trip journey times.
- The departure time influences users' trip journey times, in particular during peak hours.
- If a user utilizes the Kussbus service for their morning commute, they will likely use the service for their returning trip.
- The fare of the next ride depends on the fare of the current ride, as Kussbus provides 6 free rides for each user.
- The journey time difference between Kussbus and a car for the conducted trip is correlated to the Kussbus journey time.

Based on the identified dependency relations, we build the structural skeleton of the BN and apply the score-based algorithms and model averaging for BN structure learning. The parameters and results of the learned BN is shown in Table 5. We use the k-fold cross-validation scheme, which randomly divides the full data set into k subsets (k = 5), then we use one subset to test the prediction accuracy based on the model fitted by the remaining k–1 subsets. The implementation is based on the bnlearn package in R for learning the BN structure learning and then using Netica BN software for visualization and the sensitivity analysis.

Figure 6 shows the learned BN for users' next ride occurrence inference. We summarize the causal/dependency relations of the learned BN as follows.

Table 5 Parameters and results of the BN using the hybrid structure-learning algorithm

Attribute/Parameter	Value	Attribute/Parameter	Value
Number of nodes	12	Score-based algorithm	Hill climbing
Number of arcs	15	Log-likelihood	−21,184.91
Number of samples using bootstrap resampling	100	BIC	−26,288.7
Significance threshold	0.5	Parameter learning	Maximum likelihood estimates

- A user's next ride occurrence is directly influenced by the commute time dissonance between the actual commute time using Kussbus and a user's habitual commuting time by car. A user's next ride decision is indirectly affected by the Kussbus commute time. The latter is determined by the user's OD pair and departure time.
- The walking distance to Kussbus stops measures the inconvenience a user experiences in using the service and affects their tendency to continue to use the service.
- A user's next ride occurrence is influenced by the fare (whether the next ride is free or not), which is determined by the current fare as Kussbus provides 6 free rides to its users.
- The 'Is_first', 'Is_morning', 'Is_holiday', and 'Is_august' variables have a direct influence on whether a user's next ride occurrence is within one day or over a longer horizon.

To evaluate the proposed structure-learning approach, we compare it with the benchmark BN, i.e., naive Bayes. The naive Bayes assumes dependency between the determinants and the target variable, and independence between the determinants. We use the fivefold cross-validation scheme to evaluate the prediction accuracy. The result shows that the BN learned from the hybrid structure-learning algorithm significantly improves the performance of the naive Bayes with an average prediction accuracy of 0.79 (vs. naive Bayes of 0.66). Table 6 reports the MNL model estimation results using the same variables with "within 1 day" as the reference class. The pseudo R^2 value of the MNL model is 0.2893. The coefficients allow us to analyze the related positive (or negative) influence of the covariates on the class (category) of the user's next ride occurrence. As the sample size of different classes is unbalanced, the interpretation of the estimated coefficients need be cautious. For the class of 'Not utilize anymore', the variable T_Kussbus is statistically significant, suggesting that higher user's ride time is, users tend to continue to use the service, which seems counter-intuitive. However, higher users' journey time difference between Kussbus and a car (T_diff) tends to discourage users to continue to use the service. A similar effect is observed for the total walking distance between user's origin/destination and Kussbus stops (Walk_dist). Users' OD pairs between Habay/Arlon and Luxembourg City or Kirchberg district tend to use the service frequently (coefficients are negative for '≥ 7 days' and 'Not utilize anymore'). 'Isfree_next_ride' and 'Is_morning' have a negative effect on not continuing to use the service, while 'Is_holiday' and 'Is_first' have a positive effect

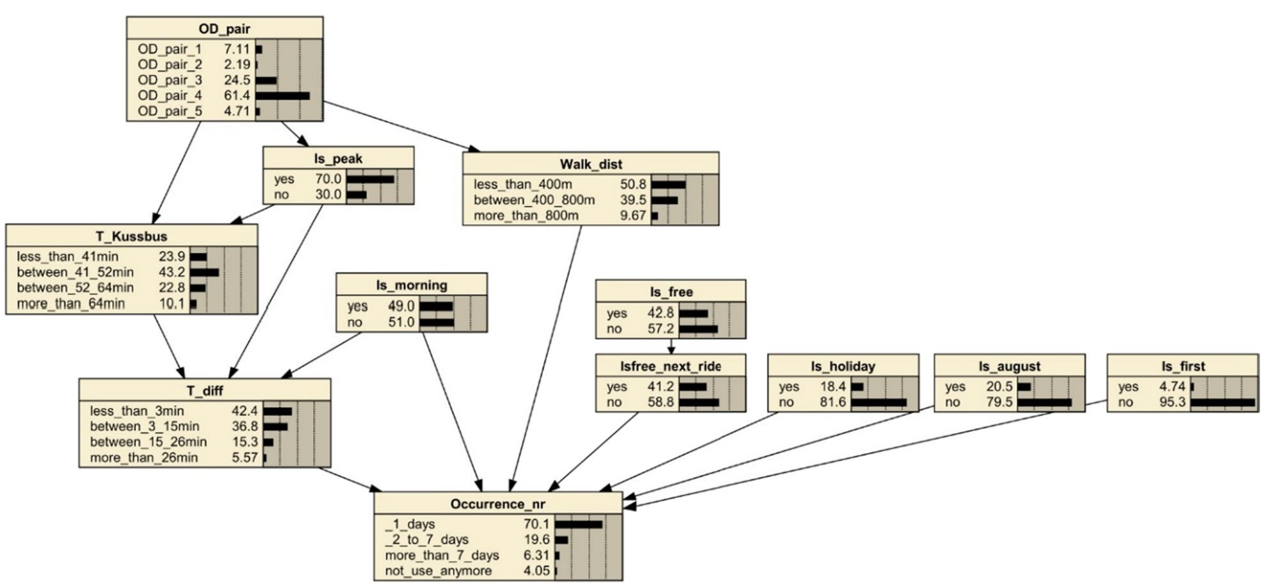

Fig. 6 Learned BN structure for Kussbus user next ride occurrence inference

Table 6 Multinomial logit model results

Category of user's next ride occurrence	2 to 7 days		≥ 7 days		Not utilize anymore	
Variable	Coef	z-value	Coef	z-value	Coef	z-value
T_Kussbus	0.03	1.63	− 0.02	− 0.75	− 0.05*	− 1.76
T_diff	− 0.04**	− 2.48	0.01	0.34	0.06**	2.09
Walk_dist	0.29**	2.11	0.72***	4.43	0.72***	3.28
OD_pair						
Habay–Luxembourg	− 0.33	− 0.67	− 2.28**	− 2.14	− 1.72**	− 2.06
Arlon–Kirchberg	0.10	0.36	− 1.32***	− 3.87	− 2.40***	− 5.05
Arlon–Luxembourg	0.01	0.04	− 0.95***	− 3.17	− 1.69***	− 4.73
within Luxembourg	0.71	1.59	− 19.12	− 0.01	0.41	0.36
Is_peak	− 0.42***	− 3.02	− 0.63***	− 3.21	0.004	− 0.02
Isfree_next_ride	0.04	0.19	− 0.25	− 0.93	− 1.91***	− 6.56
Subsidy	0.36*	1.94	0.63**	2.28	19.89	0.02
Is_morning	− 3.49***	− 17.19	− 2.68***	− 10.09	− 1.69***	− 4.89
Is_holiday	2.82***	17.16	2.40***	10.8	1.38***	4.04
Is_august	− 0.04	− 0.27	− 0.19	− 0.8	− 0.08	− 0.22
Is_first	0.56	1.61	1.69***	4.99	2.14***	5.96
Constant	− 1.94**	− 2.52	− 0.55	− 0.52	− 17.44	− 0.02
N			2783			
DF			42			
Log-Likelihood			− 1634.79			
McFadden's Pseudo R2			0.2893			
Likelihood-ratio test (Prob > χ2)			< 0.0001			

Remark The reference class is more within 1 day.

Statistical significance levels: *0.05 < *p*-value ≤ 0.1; **0.01 < *p*-value ≤ 0.05; ****p*-value ≤ 0.01

on users' not continuing to use the service. A similar interpretation could be given for the other two categories '2–7 days' and '≥ 7 days'. In terms of prediction accuracy, the MNL model provides a similar performance (0.79) compared with the BN approach. As an alternative to the MNL approach, the BN approach learns a rule-based decision model providing an intuitive way to reveal the dependency relationship between different influencing factors from the learned BN structure.

4.3.2 Sensitivity analysis

We further analyze the changes in the conditional probability distribution of users' next ride occurrence when new evidence is provided (see Table 7). For example, the operator might be interested in knowing such a probability if users' journey times using Kussbus are (1) similar to (T_diff ≤ 2.5 min.) or (2) much longer (T-diff > 25.9 min.) than using cars. For the first case, the probability of users' next ride occurrence being within one day would increase by 5.1%, with slightly decreasing probabilities for within one week/weeks and no further use of the service. For the second case, we observe that users' next ride occurrence would be negatively affected: − 5.5% and − 4% probabilities for the next ride occurrence being within one day and within one week, + 3.69% for more than one week, and + 5.85% for no further use of the service. If the ride is the first ride, this increases the probability of no longer utilizing the service by 13.85%. Similarly, when a user's next trip has to be paid for, the probability of not using the service increases from 4.05 to 5.9%. However, other factors may influence this result, such as the socio-demographic characteristics of users (income, attitude and perception of the service, mobility needs, etc.). Further research is needed to investigate this aspect, which could provide useful information to the operator for their system design and service improvement. The operator can further quantify the probability changes of the target variable by inspecting the interaction effect of several variables. Based on the learned BN model in Figs. 6 and 7 illustrates an example of such an interaction effect with a journey time greater than 25.9 min for a user's first ride and when the next ride is not free. In this case, the probability of not using the service again increases from 4.05 to 25.1% (+ 20.05%).

From this example, we see how the operator could apply this tool to infer the next ride occurrence of users under uncertainty.

Table 7 Probability changes in users' next ride occurrence given new evidence from other nodes

Node	New evidence	Conditional probability distribution of the target node (Occurrence_nr), measured in %			
		< = 1 day	Between 2 and 7 days	more than 7 days	No further use
T-diff (in minutes)	≤ 2.5	75.1(5.1)	16.9(− 2.7)	5.3(− 1.01)	2.7(− 1.35)
	> 2.5 and ≤ 14.7	67.8(− 2.2)	21.9(2.3)	6.8(0.49)	3.5(− 0.55)
	> 14.7 and ≤ 25.9	63.4(− 6.6)	23(3.4)	6.5(0.19)	7.1(3.05)
	> 25.9	64.5(− 5.5)	15.6(− 4)	10(3.69)	9.9(5.85)
Is_first	Yes	45.2(− 24.8)	19.4(− 0.2)	17.5(11.19)	17.9(13.85)
	No	71.3(1.3)	19.6(0)	5.8(− 0.51)	3.4(− 0.65)
Is_free_next_ride	Yes	73.6(3.6)	19.2(− 0.4)	5.8(− 0.51)	1.4(− 2.65)
	No	67.6(− 2.4)	19.8(0.2)	6.7(0.39)	5.9(1.85)
Is_morning	Yes	88(18)	6(− 13.6)	3.3(− 3.01)	2.7(− 1.35)
	No	52.8(− 17.2)	32.6(13)	9.2(2.89)	5.4(1.35)

Remark The numbers in parentheses are the probability changes with respect to the case without new evidence from other nodes of the BN, i.e., 0.7 (< = 1 day), 0.196 (2–7 days), 0.0631 (≥ 7 days), and 0.0405 (no further use)

5 Discussion and conclusions

5.1 Discussion

In this section, we discuss the main findings, policy recommendations, and methodological limitations of this study.

(a) It has been demonstrated that many microtransit services play the role of complementing public transport in a rural area [15]. From our analysis of public transport coverage and Kussbus ride demand, we see that the current transit network in the study area could not meet users' mobility needs, with higher trip journey times and inconvenient transit transfers. Kussbus provides a user-centered, flexible shuttle service with advanced booking, meeting points, and the latest real-time vehicle location tracking technologies. The progress of Kussbus ridership over time shows the potential of promoting flexible transit to change commuters' mode-choice behavior from their habitual car use.

(b) Despite the success of Kussbus in attracting car users by providing 6 free trials and setting a ticket price in between the cost of using a car and public transport fares, Kussbus discontinued its service after one year due to insufficient revenue and cost overrun [12]. Lessons learned from Kussbus operations suggest that financial viability remains a barrier to successful deployment of such a service. In

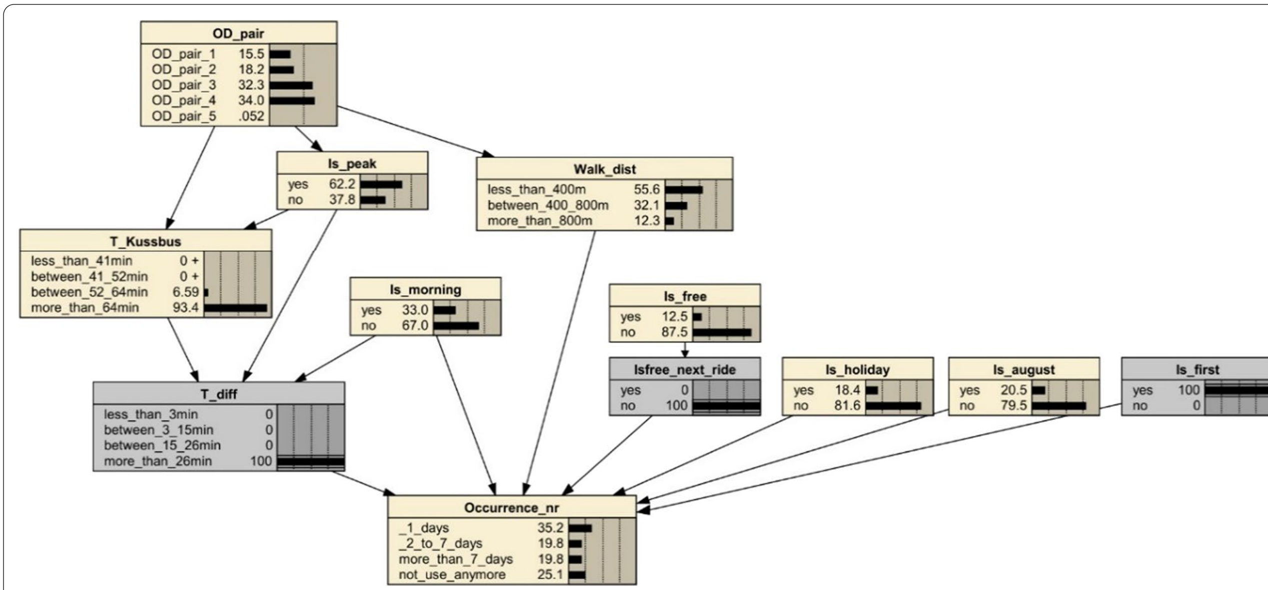

Fig. 7 Example of the probability changes in users' next ride occurrence given new evidence from "T_diff", "tIsfree_next_ride", and "Is_first"

terms of operation policy, the operator could consider providing feeder services to connect train stations as a part of seamless multimodal transit solutions to increase their ridership and reduce their operating costs.

(c) To understand the factors affecting users' next ride decisions, we were able to model the causal/dependent relationships between users' ride experiences and their next ride decisions with a discrete BN and compare it with an MNL model. Overall, our findings suggest that the results obtained by the BN approach provide similar prediction power compared to the MNL model, while the former presents an advantage of revealing the dependency structure of different factors and easy to understand. We find that the commuting time difference between Kussbus and cars plays a key role in their willingness to continue to use the service. Moreover, when users experience longer commute times in their initial trials, they tend to not continue to use the service. This is not surprising as this commute time dissonance with respect to travelers' ideal/actual commute time would negatively impact users' travel satisfaction and thus their mode-choice behavior [33, 34]. It is then necessary to improve this issue by examining operation policies or changing current transport policy in this area to favor the use of public transport. Another interesting research line is to compare the factors affecting users' ride decisions for the free trial and paid user groups. We were unable to conduct reliable analysis due to restricted sample size.

(d) In terms of methodological limitations, future research could consider the hybrid BN with both discrete and continuous variables [31]. In our empirical data, the continuous variables do not follow continuous probability distributions (e.g., normal distribution), so we adopt a discrete BN approach. Collecting data over a longer period with additional fields regarding users' socio-demographic attributes is expected to improve the model fitness and prediction performance. Another possible extension is to apply under-/over-sampling techniques to address the issue of imbalanced class (i.e., the classes of "$\geq$ 7 days" and "no further use") so as to increase the prediction accuracy for the class of interest [35].

5.2 Conclusions

On-demand microtransit services have been considered an efficient alternative to reduce personal car use in rural areas as they provide a user-centered service and have the potential to complement traditional fixed-route transit. While many studies have focused on the ex-post evaluation of microtransit services based on empirical ride data, few studies have tried to understand the relationships between users' ride experiences and their next ride decisions. In this study, we aim to analyze these relationships and propose a BN approach to analyze the factors affecting users' next ride decisions (i.e., next ride within the same day, within one week/weeks, or no further use). Using the historical ride data provided by a recent microtransit pilot, "Kussbus", in the Arlon–Luxembourg cross-border area, we were able to identify key factors and the relationships between them for predicting the next ride occurrence decisions of users. Furthermore, we find that the Kussbus service plays a role in complementing the existing bus and train network for Belgium cross-border commuters, who have largely been relying on personal car use.

The results of the proposed BN model allow the operator to forecast future ride demand for a short horizon and manage their resource allocation in advance. Given that the public transport supply in the study area does not currently provide a convenient option for cross-border commuters (multiple transit transfers are required), commuting by private car has been the preferred option, causing serious traffic congestion and raising public health concerns in the study area. Our findings suggest that new operational strategies and a thorough analysis of financial feasibility are needed to improve service viability and promote public transportation in the study area.

Acknowledgements
We thank the Utopian Future Technologies S.A. for providing Kussbus ride data for this research.

Author contributions
Conceptualization, JH and TM; methodology, JH and TM; formal analysis, JH and TM; investigation, JH and TM; writing—original draft preparation, JH and TM. All authors have read and agreed to the published version of the manuscript.

Funding
The work was supported by the Luxembourg National Research Fund (C20/SC/14703944).

Availability of data and materials
Not applicable.

Declarations

Conflict of interests
The author declare that they have no conflict of interests.

Author details
[1]Katholieke Universiteit Leuven, Oude Markt 13, 3000 Leuven, Belgium. [2]Luxembourg Institute of Socio-Economic Research (LISER), 11 Porte des Sciences, 4366 Esch-sur-Alzette, Luxembourg.

Received: 16 May 2022 Accepted: 6 October 2022
Published online: 27 October 2022

References

1. Volinski, J. (2019). Microtransit or general public demand–response transit services: state of the practice (No. Project J-7, Topic SB-30).
2. Bliss, L. (2017). Bridj is dead, but microtransit Isn't. Citylab. https://www.urbanismnext.org/resources/bridj-is-dead-but-microtransit-isnt. Accessed 20 Oct 2022.
3. Haglund, N., Mladenović, M. N., Kujala, R., Weckström, C., & Saramäki, J. (2019). Where did Kutsuplus drive us? Ex post evaluation of on-demand micro-transit pilot in the Helsinki capital region. *Research in Transportation Business & Management, 32*, 100390.
4. Ma, T.-Y., Rasulkhani, S., Chow, J. Y. J., & Klein, S. (2019). A dynamic ridesharing dispatch and idle vehicle repositioning strategy with integrated transit transfers. *Transportation Research Part E: Logistics and Transportation Review, 128*, 417–442. https://doi.org/10.1016/j.tre.2019.07.002
5. Martinez, L. M., & Viegas, J. M. (2017). Assessing the impacts of deploying a shared self-driving urban mobility system: An agent-based model applied to the city of Lisbon, Portugal. *International Journal of Transportation Science and Technology*. https://doi.org/10.1016/j.ijtst.2017.05.005
6. Alonso-González, M. J., Liu, T., Cats, O., Van Oort, N., & Hoogendoorn, S. (2018). The potential of demand-responsive transport as a complement to public transport: An assessment framework and an empirical evaluation. *Transportation Research Record*. https://doi.org/10.1177/0361198118790842
7. Jokinen, J. P., Sihvola, T., & Mladenovic, M. N. (2019). Policy lessons from the flexible transport service pilot Kutsuplus in the Helsinki capital region. *Transport Policy, 76*, 123–133.
8. Ho, S. C., Szeto, W. Y., Kuo, Y. H., Leung, J. M., Petering, M., & Tou, T. W. (2018). A survey of dial-a-ride problems: Literature review and recent developments. *Transportation Research Part B: Methodological, 111*, 395–421.
9. Sandlin, A. B., & Anderson, M. D. (2004). Serviceability index to evaluate rural demand-responsive transit system operations. *Transportation Research Record, 1887*(1), 205–212.
10. Avermann, N., & Schlüter, J. (2019). Determinants of customer satisfaction with a true door-to-door DRT service in rural Germany. *Research in Transportation Business and Management*. https://doi.org/10.1016/j.rtbm.2019.100420
11. Ferreira, L., Charles, P., & Tether, C. (2007). Evaluating flexible transport solutions. *Transportation Planning and Technology, 30*(2–3), 249–269.
12. Ma, T. Y., Chow, J. Y. J., Klein, S., & Ma, Z. (2021). A user-operator assignment game with heterogeneous user groups for empirical evaluation of a microtransit service in Luxembourg. *Transportmetrica A: Transport Science, 17*(4), 946–973. https://doi.org/10.1080/23249935.2020.1820625
13. Brake, J., Mulley, C., Nelson, J. D., & Wright, S. (2007). Key lessons learned from recent experience with flexible transport services. *Transport Policy, 14*(6), 458–466.
14. Westervelt, M., Huang, E., Schank, J., Borgman, N., Fuhrer, T., Peppard, C., & Narula-Woods, R. (2018). UpRouted: Exploring microtransit in the United States. https://www.enotrans.org/eno-resources/uprouted-exploring-microtransit-united-states/. Accessed 20 Oct 2022.
15. Perera, S., Ho, C., & Hensher, D. (2020). Resurgence of demand responsive transit services–Insights from BRIDJ trials in inner west of Sydney. *Australia. Research in Transportation Economics, 83*, 100904.
16. Currie, G., & Fournier, N. (2020). Why most DRT/Micro-transits fail–what the survivors tell us about progress. *Research in Transportation Economics, 83*, 100895.
17. Yu, H., & Peng, Z. R. (2019). Exploring the spatial variation of ridesourcing demand and its relationship to built environment and socioeconomic factors with the geographically weighted Poisson regression. *Journal of Transport Geography, 75*, 147–163.
18. Deka, D., & Fei, D. A. (2019). A comparison of the personal and neighborhood characteristics associated with ridesourcing, transit use, and driving with NHTS data. *Journal of Transport Geography, 76*, 24–33.
19. Clewlow, R. R., & Mishra, G. S. (2017). *Disruptive transportation: The adoption, utilization, and impacts of ride-hailing in the united states*. University of California Institute of Transportation Studies.
20. Schasché, S. E., Sposato, R. G., & Hampl, N. (2022). The dilemma of demand-responsive transport services in rural areas: Conflicting expectations and weak user acceptance. *Transport Policy, 126*, 43–54.
21. Beirão, G., & Cabral, J. S. (2007). Understanding attitudes towards public transport and private car: A qualitative study. *Transport policy, 14*(6), 478–489.
22. Nelson, J. D., & Phonphitakchai, T. (2012). An evaluation of the user characteristics of an open access DRT service. *Research in Transportation Economics, 34*(1), 54–65.
23. Wang, C., Quddus, M., Enoch, M., Ryley, T., & Davison, L. (2015). Exploring the propensity to travel by demand responsive transport in the rural area of Lincolnshire in England. *Case Studies on Transport Policy, 3*(2), 129–136.
24. Knierim, L., & Schlüter, J. C. (2021). The attitude of potentially less mobile people towards demand responsive transport in a rural area in central Germany. *Journal of Transport Geography, 96*, 103202.
25. Statec, 2020. "Domestic employment by place of residence and nationality 1995–2021". https://statistiques.public.lu/stat/TableViewer/tableView.aspx?ReportId=12916&IF_Language=fra&MainTheme=2&FldrName=3&RFPath=92.
26. Luxmobil, 2017. "Enquête Luxmobil 2017 Premiers résultats. " https://transports.public.lu/content/dam/transport/publications/contexte/situation-actuelle/20171207-enquete-mobilite-luxmobil-2017-premiers-resultats-presse-v2.pdf
27. TomTom, 2021. Luxembourg traffic. https://www.tomtom.com/en_gb/traffic-index/luxembourg-traffic/. Online; Retrieved July 29, 2021.
28. Kuhnimhof, T., Chlond, B., & Von Der Ruhren, S. (2006). Users of transport modes and multimodal travel behavior: Steps toward understanding travelers' options and choices. *Transportation research record, 1985*(1), 40–48.
29. Hartemink, A., & Gifford, D. K. (2001). Principled computational methods for the validation and discovery of genetic regulatory networks. PhD diss., Massachusetts Institute of Technology.
30. Cooper, G. F. (1990). The computational complexity of probabilistic inference using bayesian belief networks. *Artificial Intelligence*. https://doi.org/10.1016/0004-3702(90)90060-D
31. Scutari, M., & Denis, J. B. (2021). *Bayesian networks: with examples in R*. Chapman and Hall/CRC.
32. Ma, T. Y., Chow, J. Y. J., & Xu, J. (2017). Causal structure learning for travel mode choice using structural restrictions and model averaging algorithm. *Transportmetrica A: Transport Science, 13*(4), 299–325. https://doi.org/10.1080/23249935.2016.1265019
33. Humagain, P., & Singleton, P. A. (2020). Investigating travel time satisfaction and actual versus ideal commute times: A path analysis approach. *Journal of Transport & Health, 16*, 100829.
34. Ma, T. Y., Van Acker, V., Lord, S., & Gerber, P. (2021). Dissonance and commute satisfaction: Which reference point to use? *Transportation Research Part D: Transport and Environment*. https://doi.org/10.1016/j.trd.2021.103046
35. Fernández, A., García, S., Galar, M., Prati, R. C., Krawczyk, B., & Herrera, F. (2018). *Learning from imbalanced data sets* (Vol. 10, pp. 978–3). Springer.

Publisher's Note