A RULE-BASED APPROACH TO ROUTE CHOICE SET GENERATION

Sascha Hoogendoorn-Lanser

Transport & Planning, Technische Universiteit Delft

lanser@xs4all.nl of s.hoogendoorn@avv.rws.minvenw.nl

Bijdrage aan het Colloquium Vervoersplanologisch Speurwerk 2006,

23 en 24 november 2006, Amsterdam

Inhoudsopgave

Samenvatting

Een beslisregelgebaseerde aanpak voor het genereren van multi-modale routekeuzesets

Beschikbare data over routekeuzen bevat vaak alleen informatie over de gekozen route. Om inzicht te krijgen in het keuzegedrag, is kennis nodig over de set van routes waaruit de reiziger heeft gekozen. Deze bijdrage beschrijft een nieuwe aanpak voor het genereren van routekeuzesets in multi-modale netwerken. Deze methode gaat uit van dienstregelingen en genereert met behulp van branch&bound technieken een realistische set van multi-modale routes binnen een gegeven tijdvenster. Kernpunt is dat voor de branch&bound methode gebruik wordt gemaakt van randvoorwaarden gebaseerd op waargenomen reisgedrag. Het algoritme kan worden gebruikt voor een compleet multi-modaal netwerk, maar bij grote netwerken kan ook gebruik worden gemaakt van de structuur van multi-modale verplaatsingen, zoals bijvoorbeeld treinverplaatsingen, door eerst voor de afzonderlijke onderdelen alternatieven te genereren, gevolgd door een koppelprocedure. Op deze manier kunnen de gedragsrandvoorwaarden preciezer worden toegepast en blijft de rekentijd beperkt. Het algoritme is gekalibreerd en toegepast op een realistisch multi-modaal netwerk in Nederland. Vergelijking van de gegenereerde keuzesets met waargenomen gekozen en bekende routes, laat zien dat het algoritme zeer veel van de waargenomen routes automatisch genereert. Dit resultaat toont dat deze nieuwe methode voldoet aan een primaire eis voor keuzesetgeneratietechnieken. Het opsplitsen in aparte routedelen die daarna worden gekoppeld, leidt niet tot verlies aan kwaliteit van de keuzesets.

Summary

A rule-based approach to route choice set generation

Collected data often only includes information on chosen routes. To gain insights into travellers' route choice behaviour or to predict route shares, we need to know from which set of alternatives travellers have chosen their routes. This paper presents an alternative approach to choice set generation in mixed multi-modal networks. This new algorithm is a run-based, constrained enumeration method using branch&bound techniques and is suitable for both estimation and prediction purposes. Key characteristic of the algorithm is a set of constraints that reflects observed travel behaviour. The proposed choice set generation algorithm can be applied to a complete multi-modal network at once. However, by exploiting knowledge about the structure of multi-modal trips, separate application of the algorithm to partial networks and consecutive concatenation of subroutes into complete door-to-door routes substantially reduces computation times without resulting in incomplete choice sets. The presented choice set generation algorithm has been calibrated for and successfully applied to a real-size, mixed multi-modal transport network in The Netherlands. Comparing generated choice sets with reported chosen and known alternatives showed that the algorithm is able to generate these alternatives, thus resulting in high coverage levels. This result clearly shows that this constrained enumeration approach meets the requirements for choice set generation and thus offers very interesting perspectives for route choice analysis and prediction of route shares. Furthermore, separate application of the algorithm to partial networks and consecutive concatenation of subroutes into complete door-to-door trips substantially does not result in incomplete choice sets.

1 Introduction

It is envisioned that in the near future a huge amount of data on actual route choice behaviour will become available, e.g. using GPS technology. However, such collected data only includes information on chosen routes. To gain insights into travellers' route choice behaviour or to predict route shares, we need to know from which set of available alternatives travellers have chosen their routes. Several authors (14, 21, 25) have shown that the composition and quality of choice sets strongly influence parameter estimates and predicted route shares. Choice set generation thus is an important topic.

In this paper, we focus on so-called explicit choice set generation approaches instead of implicit ones, because explicit approaches offer (1) good opportunities for researchers to control the composition of generated choice sets, (2) more flexibility in applying models for generation and subsequent choice analysis, and (3) computational advantages in performing demand analysis in large networks. Generated choice sets should meet certain basic requirements; alternatives should (1) be logical themselves and compared to others, (2) have a sensible variation in characteristics, and (3) include the chosen alternative.

Literature on choice set generation shows a large variety of methods for generating alternative routes in networks such as the shortest path approach, k-shortest path approaches (18, 23, 26), link elimination approaches (4), link penalty approaches (11, 20), labelling approach (5), simulation methods (13) or enumeration methods combined with branch&bound techniques (12, 14, 15). Current techniques for generating choice sets have shown to have limitations in properly reproducing observed routes (see (21, 22) for a comparison of explicit choice set generation techniques). For one, generated choice sets do often not include the chosen route, which has to be added afterwards. Secondly, it is well known that many methods usually lead to routes that are simply variations of the shortest path with very much mutual overlap.

Most of these methods have been applied to road networks only. For mixed multi-modal networks - including private and public transport networks - the choice set problem is more complicated, because apart from spatial variety also modal variety is introduced as a dimension. In literature, examples of choice set generation for this type of network are rare (1, 2, 3, 6, 13, 14, 15). Typical approaches are to apply path composition rules or to use strict selection criteria to avoid irrelevant alternatives.

Hoogendoorn-Lanser (15) presented an alternative approach to choice set generation in mixed multi-modal networks, including private as well as public transport modes. This new algorithm is a deterministic, run-based, selective enumeration method using branch&bound techniques and is applicable for generating choice sets for individual travellers taking into account traveller and trip characteristics and is suitable for both estimation and prediction purposes. Key characteristic of the algorithm is a set of constraints that reflect observed travel behaviour. All feasible route alternatives between an origin and destination that satisfy such a set of constraints are generated. A subset of these constraints is applied to limit the generation process within feasibility bounds of space and time, while another subset is applied to restrict the generation to plausible routes. An advantage of this approach is its comprehensiveness within the given bounds; it generates all feasible paths. Another advantage is its deterministic property; the generation is perfectly reproducible in contrast to so-called stochastic generation approaches, where the outcomes such as size and composition of choice sets are random variables (see (17)). Since the generated set is exhaustive, the full choice set is known which forms a solid basis for selecting subsets of routes for particular purposes. One such purpose might be the estimation of parameters of discrete choice models that requires choice sets to satisfy special requirements with respect to size and composition (e.g. (24)).

Our newly developed branch&bound algorithm has been calibrated for and applied to generate choice sets to a real-size, mixed multi-modal transport network in an urbanized corridor in The Netherlands, including urban public transport (UPT) networks in Dutch cities, such as Rotterdam and The Hague, and the Dutch train network. Although the algorithm can be applied to the full multi-modal network at once, the algorithm can be applied more efficiently by exploiting knowledge about the structure of multi-modal trips. This paper shows the successful calibration and application of the branch&bound algorithm to partial networks and the subsequent concatenation of generated train, access and egress alternatives into full door-to-door trips. It also addresses issues regarding the definition of trip parts, selection of search time frames, selection of boarding and alighting stops, and formulation of concatenation criteria that need to be solved in order to apply this stepwise approach. This paper does however not include a mathematical formulation of the branch&bound algorithm (see (15) instead).

The paper is structured as follows. First, the separate generation of routes through partial networks and the subsequent concatenation of sub routes through these networks into full door-to-door trips are motivated. Then, an outline of the choice set generation algorithm distinguishing partial networks is given, followed by a description of the basic elements of the approach, i.e. the extended diachronic super-network, the search time frame, the selection of boarding/alighting stops, the tree search, the constraints and the re-evaluation step. Subsequently, the generation of access, egress and train alternatives is discussed as well as the concatenation into door-to-door trip alternatives. Next, the data used in the application of the choice set generation approach is described, followed by a discussion on the calibration of the choice generation algorithm and the quality of generated choice sets. The paper finishes with summarizing the main findings.

2 Motivation for choice set generation in partial networks

Application of the branch&bound algorithm to realistic, large-scale transport networks, including all individual PT stops and runs, might cause computational problems, i.e. large computation times and high memory requirements. The reason for this is that the width of the search tree grows exponentially with the number of nodes in the network, even when strict limiting constraints are used.

The algorithm can be applied more efficiently, by using knowledge about the structure of multi-modal, inter-urban train trips. Empirical analysis (27) showed that train is the only main transport mode in multi-modal trip making in the considered corridor (Dordrecht-Leiden). Therefore, a multi-modal inter-urban train trip can be split up in three trip parts, namely a train, an access and an egress part, where the access and egress parts each may consist of various uni-modal networks (car, bicycle, walk, local public transport) and combinations thereof. The choice set generation algorithm can then be applied separately to the three partial networks, while generated multi-modal alternatives for the different trip parts are finally combined into complete door-to-door trip alternatives. Due to the specific structure of the trips and the ability to control the strictness of the constraints, the risk of missing important trip alternatives is negligible.

Another advantage of separate application of the choice set generation algorithm to partial networks is that constraints can be calibrated for each network part separately, or even different types of constraints can be adopted. In this way, it is possible to better account for network-specific characteristics, such as network level, transport service types, service frequencies, and network-specific travel behaviour (e.g. maximum detours).

3 Outline of choice set generation algorithm distinguishing partial networks

The choice set generation algorithm distinguishing partial networks consists of the following seven steps (see Figure 1):

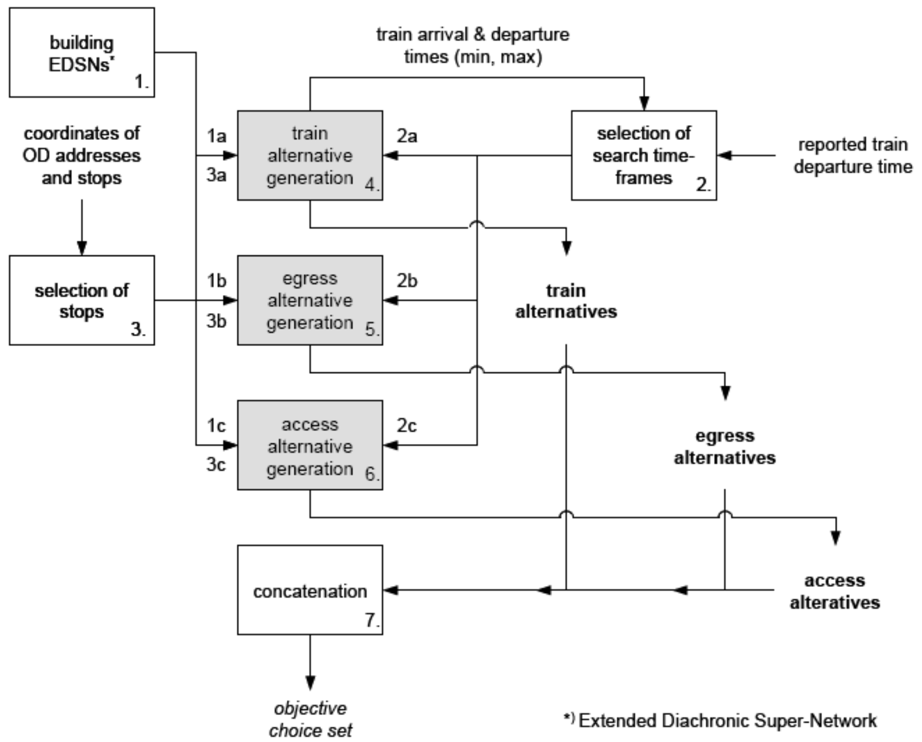

Figure 1 Overview of the structure of the multi-modal choice set generation algorithm for inter-urban train trips.

1. Generation of extended diachronic super-networks (for access, egress and train part);
2. Determination of search time frames (for access, egress and train part);
3. Selection of candidate boarding and alighting stops (for access, egress and train part);
4. Generation of train alternatives from all selected boarding stations to all selected alighting stations within the predefined train search time frame (tree search & re-evaluation step);
5. Generation of egress alternatives from all selected alighting stations to the destination address using UPT as well as private transport modes within the predefined egress search time frame (tree search & re-evaluation step);
6. Generation of access alternatives from the origin address to all selected boarding stations using UPT as well as private transport modes within the predefined access search time frame (tree search & re-evaluation step);
7. Concatenation of the generated train, access and egress alternatives and evaluation of resulting door-to-door alternatives.

In the remainder, the different steps are discussed in more detail.

4 Construction of extended diachronic super-networks

Our multi-modal branch&bound choice set generation approach is a so-called timetable-based or run-based approach. This means that each run of a public transport service is explicitly taken into account in both space and time. Each transport service is represented by a separate space-time graph or diachronic graph, consisting of a service sub-graph, a demand sub-graph and an access-egress sub-graph (see (9, 19) for details). These diachronic graphs are combined into a diachronic super-network by adding transfer-walking legs connecting the various public transport services. Depending on the type of partial network these public transport services are train, metro, tram and/or bus services. To restrict the complexity of the network only 'reasonable' transfer-walking legs are added (within a maximum transfer-walking distance).

To facilitate quick computing and prevent generation of illogical routes, the diachronic super-network is extended by adding transfer-free in-vehicle route segments (so-called legs) from every stop of a public transport line to all downstream stops of that line (see (10, 12)), resulting in an extended diachronic super-network. These legs are defined for each run of the public transport line using timetables. The generated trip alternatives are multi-modal routes in extended diachronic super-networks for the separate trip parts.

For the train network, demand is defined at railway stations, while for the local networks demand is defined at origin and destination addresses. The access/egress sub-graph for the train network consists of artificial legs assigning demand to railway stations for each time period. For the local networks, origin and destination addresses are connected to the network via private mode legs, namely walking, cycling and car legs. To restrict the complexity of the network only 'reasonable' access/egress legs are added (e.g. within minimum and maximum walking, cycling and car distances).

5 Specification of search time frames

Since multi-modal trips are split up in train, egress and access parts and the choice set generation approach is applied to these parts separately, search time frames have to be determined for each of them. In order to produce complete door-to-door trips, suitable egress alternatives from alighting station α to destination D have to be identified for each train alternative that arrives at α, and suitable access alternative from origin O to boarding station β

have to be identified for each train alterative that departs from β. Thus, search time frames for the train, access and egress parts are dependent.

The reference time for the specification of time frames can be based on an observed, reported, preferred or estimated departure time at the origin (1), departure time from the railway station (2), arrival time at the railway station (3) or arrival time at the destination (4). In this paper, the observed departure time τ^* from the chosen boarding station is used as a reference. The search time frame for all candidate boarding stations β of the train part is then given by $[\tau^*-\Delta_1, \tau^*+\Delta_2]$, where Δ_1 and Δ_2 reflect the maximum acceptable *earliness* and *lateness* (reported by travellers) at boarding station β, which reflect the strictness of traveller's *time-pressure constraints*. Within this time frame, train alternatives may depart every minute.

The time frame for the egress part of the trip is based on the arrival times of generated train alternatives (within the train time frame) at alighting stations. For each candidate alighting station α, earliest and latest arrival times are known. Let φ_α and Φ_α be the earliest and the latest arrival times at the alighting station α determined over all relevant train alternatives arriving at this railway station. Let Δ_3 be the acceptable maximum waiting time at an UPT boarding stop. The egress search time frame is then defined by $[\phi_\alpha, \Phi_\alpha+\Delta_3]$.

The time frame for the access part of the trip is based on the time interval at the candidate boarding stations. Let Δ_4 be the maximum (estimated) UPT access travel time. Then, the search time frame for UPT access to boarding station β is given by $[\tau^*-\Delta_1-\Delta_4, \tau^*+\Delta_2]$.

6 Selection of boarding and alighting stops

Separate application of the choice set generation algorithm requires the identification of origins and destinations for the train, access and egress parts. For the train part, this means selecting suitable boarding and alighting stations. To this end, maximum access and egress distances to railway stations have been deduced from the Dutch National Travel Survey[1] (DNTS). It appears that travellers accept longer access distances if railway stations offer high-speed trains, and higher train and UPT frequencies. Furthermore, feeder distances increase with total travel distance. Therefore, different maximum distances are applied to identify

[1] In the Dutch National Travel Survey (OVG/MON) travel data has been collected for more than 70.000 households annually, resulting in 600.000 trip records.

relevant Intercity, express and local railway stations. If no railway station is present within the maximum distance range, the railway station closest to the origin or destination address is selected.

For the generation of access and egress alternatives, admissible boarding and alighting UPT stops in the vicinity of boarding and alighting stations and near the travellers' origin and destination addresses have to be determined. DNTS trip data are again used to derive maximum walking distances used by train travellers to/from UPT stops. If the origin or destination address is closer to a railway station than a given minimum distance, UPT is disregarded as a feeder mode to the railway station and only private feeder modes (walk, bicycle, car) are considered.

7 Tree search step

In the tree search step (so-called *constrained path enumeration*) all valid routes through an extended diachronic super-network from origin node O to one or more destination nodes D_k with $k \in K$ (*one-to-all search algorithm*) are determined simultaneously. 'Valid' means that the alternatives satisfy predefined constraints and leave origins O during time interval $[\tau_O - \Delta T_{early,O}, \tau_O + \Delta T_{late,O}]$. In this paper, origin O is an origin address (generation of access alternatives), a boarding station (generation of train alternatives) or an alighting station (generation of egress alternatives).

Figure 2 shows how the search tree is structured. At level 1, all legs - leaving origin O during time interval $[\tau_O - \Delta T_{early,O}, \tau_O + \Delta T_{late,O}]$ - are selected (set Q_1). Constraints in this step refer mainly to admissible departure times (i.e. departures within the time frame) and availability of legs at O within this time interval. At level 2, the trip segments in set Q_1 (branches) are extended with legs starting from the alighting points of these trip segments if and only if the extended trip segments satisfy a predefined set of constraints. This results in a set of trip segments (set Q_2) from origin O having one transfer. Similarly, Q_i (index i indicates level) is the set of trip segments (branches) from origin O consisting of i consecutive transfer-free legs (thus having i-1 transfers) satisfying all constraints. At level i, constraints relate to unnecessary transfers, acceptable transfer times, acceptable number of transfers, acceptable walking, cycling and driving distances, etc. All trip segments in Q_i are processed consecutively. Each candidate trip segment from origin O to node x using mode combination m that satisfies all constraints is added to the set of all admissible trip segments using mode combination *m* from origin *O* to node *x* generated so far (thus irrespective of the number of transfers).

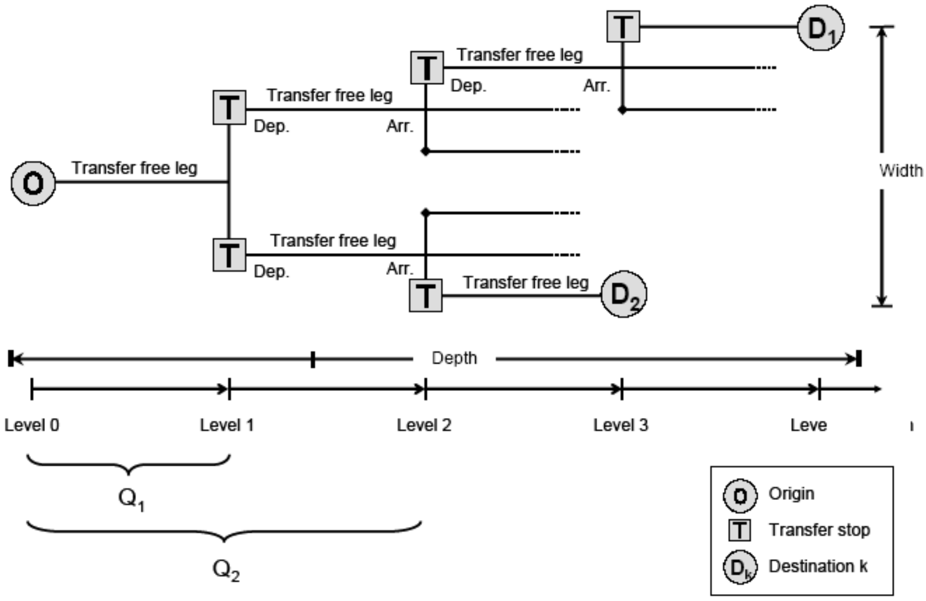

Figure 2 Structure of a search tree.

Each level is traversed completely before trip segments of the next tree level are considered. If the destination node of a trip segment is equal to one of the given destination nodes, a feasible route from origin node *O* to that destination node is found, and thus the search process via that branch is stopped. The algorithm terminates after all the trip segments of the current level *i* have been processed, if 1) the set of trip segments of the next level, Q_{i+1}, is empty, or 2) if *i* is equal to one plus the maximum number of transfers.

8 Constraints

In our branch&bound approach, constraints play a decisive role in bounding the problem. Each candidate trip segment - which can either be a complete route or part of a route - is evaluated (at each tree level). If a new trip segment satisfies all constraints, it is accepted as being valid. The constraints included in the choice set generation algorithm do not compensate one another. More specifically, when one of the constraints is not satisfied for an alternative, that alternative cannot be included in the choice set even if it satisfies all the other constraints.

Depending on network and trip characteristics, and on the purpose for which generated choice sets will be used, different types of constraints can be included in the branch&bound

approach. The constraints discussed in this paper apply to multi-modal trips and are based on logical, feasibility and behavioural criteria (each of which is described below).

Logical criteria, among other things, sort out the possibility of generating routes that:

- depart from a transfer node before arriving there;
- contain circles.

Feasibility criteria reflect the suitable of alternatives given personal, trip, and network characteristics and account for, among other matters, for feasibility in terms of:

- time (time pressure at origin and destination, travel speed, availability of transport modes in time);
- space (ownership and readiness to use of private modes, availability of public transport services);
- money (monetary budget, income);
- physical disability restricting the use of private transport modes and public transport services.

Behavioural criteria reflect observed travel behaviour and account, for example, for:

- bounds on trip attribute values, like transfer-waiting times, costs and walking / cycling / driving distances, irrespective of attributes values of other routes in the choice set as well as compared to the 'best' alternative in the choice set;
- travellers' tendency to travel in the direction of the destination and the uncomfortable feeling if the trip diverts too far in directions that lead them further away from the destination;
- fact that if alternatives are much alike, travellers are not likely to distinguish between them and consider them to be equal.

All criteria involved in the evaluation can be separated in two groups:

- criteria that apply to single trip segments only (so-called single route constraints);
- criteria that apply to sets of trip segments instead (so-called route set constraints).

Local and feasibility criteria are mainly single route constraints, while behavioural criteria can be either single route or route set constraints.

In order to be incorporated in the branch&bound algorithm, criteria need to be translated into constraints - formulated as inequalities. The branch&bound algorithm can be applied with different parameter settings (for the same set of inequalities) or with different sets of constraints.

9 Re-evaluation step

The previous sections describe how a set of trip alternatives from an origin to a destination can be generated in an extended diachronic super-network following a constrained tree search. During the tree search, candidate trip segments can only be compared to all valid trip segments generated so far. Trip segments that satisfy all constraints (route set as well as single route constraints) in the very beginning of the generation process may not satisfy certain constraints anymore when compared to trip segments that are generated in a later stage of the tree search. Later in the search process a trip segment - that is 'better' than all trip segments accepted so far - might be found. Therefore, the generated set of alternatives needs to be re-evaluated (so-called *re-evaluation step*) after completion of the tree search step to identify and delete alternatives that do not (anymore) satisfy the route set constraints. Note that constraints that apply to single routes do not need to be checked again.

To avoid that trip segments are disregarded too soon in the tree search, parameters should not be set too tight during the tree search. In the re-evaluation step, however, constraints are checked for entire trips and parameters can be adapted to specify stricter constraints. Besides the route set constraints that were used during the tree building it is possible to incorporate additional route set constraints during re-evaluation.

10 Generation of door-to-door alternatives

10.1 Generation of train alternatives

To generate train alternatives the branch&bound algorithm is applied to the extended diachronic train super-network. Since the algorithm is a one-to-many algorithm, it has to be applied $|B|$ times, where $|B|$ is the number of admissible boarding stations. This means that $|B|$ different search trees are build (one for every boarding station $\beta \in B$ to all alighting stations $\alpha \in A$). This results in $|B|$ sets $C_{\beta,A}$ of paths from a boarding station $\beta \in B$ to all alighting stations $\alpha \in A$, leaving β during time interval $[\tau^*-\Delta_1,\tau^*+\Delta_2]$.

10.2 Generation of egress alternatives

Once a set of train alternatives has been generated, the egress alternatives for the different alighting stations can be determined. To generate egress alternatives the branch&bound approach is applied to an extended diachronic super-network that corresponds to the area in which the alighting stations $\alpha \in A$ as well as destination address[2] δ are located. It includes both UPT and private modes. The algorithm needs to be applied $|A|$ times, where $|A|$ is the number of admissible alighting stations - once for each alighting station $\alpha \in A$ to the destination address δ. This thus results in $|A|$ sets $C_{\alpha,\delta}$ of paths from $\alpha \in A$ to δ, leaving α during time interval $[\varphi_\alpha, \Phi_\alpha + \Delta_3]$. Each of the UPT alternatives runs through one of the selected boarding stops near alighting station α and through one of the selected alighting stops near destination δ. A private mode alternative runs directly from the alighting station to the destination.

10.3 Generation of access alternatives

The approach, applied to generate access alternatives, is similar to that described for egress alternatives. However, the algorithm only has to be applied only once from the origin address o to all boarding stations $\beta \in B$, resulting in a set $C_{o,B}$ of paths from o to all boarding stations $\beta \in B$, leaving o in time interval $[\tau^* - \Delta_1 - \Delta_4, \tau^* + \Delta_2]$.

10.4 Concatenation

The generated train, access and egress alternatives have to be combined into complete door-to-door alternatives from origin address o to destination address δ via one of boarding station $\beta \in B$ and one of the alighting stations $\alpha \in A$. The concatenation points are thus the boarding and alighting stations of the train alternatives. In order to result in valid door-to-door alternatives a set of concatenation constraints has to be satisfied. More specifically, an access alternative can be concatenated to a train alternative if the waiting time at boarding station is within a specified range. If more than one generated access alternative with a specific mode combination results in a reasonable transfer-waiting time and can thus in principal be concatenated with the train alternatives, the one with the shortest transfer-waiting time is selected. The same constraints also apply to the concatenation of train and egress alternatives.

Finally, potential door-to-door trip alternatives are constructed by combining each train alternative with access and egress alternatives that satisfy the abovementioned constraints. Although access, train and egress parts of a door-to-door alternative separately satisfy all constraints, certain constraints may be violated when considering complete door-to-door trips.

---

[2] Note that *O* and *D* are the origin and destination nodes of the tree search step, while o and δ are the origin and destination addresses of a traveller's trip.

Therefore, for each complete trip certain constraints have to be checked (see (*15*) for details). The concatenation step can thus be considered as *constrained combinatorics*.

11 Data used to illustrate the choice set generation algorithm

11.1 Survey data

To illustrate the application of the separated branch& bound algorithm, multi-modal travel behaviour data are used from a survey conducted among train travellers in an urbanized corridor in The Netherlands, including the cities Dordrecht, Rotterdam, The Hague and Leiden (see (*15*)). The survey focused on the multi-modal trip itself (which modes used, which transfer nodes, which boarding and alighting nodes, etc) and on train-based trip alternatives known by the traveller, i.e. the *observed subjective choice set*. The first one will be referred to as the *reported chosen trip*, while the others are referred to as *reported subjective alternatives*. The sample contains 708 respondents, all corresponding to different OD-pairs - located throughout the study area.

11.2 Multi-modal network data

The choice set generation algorithm is calibrated for and applied to a mixed private and public transport network, including UPT networks and the Dutch train network. Timetables (years: 2000-2001) of all UPT services in The Hague, Rotterdam, Dordrecht, Leiden, Delft and Haarlem, and of all Dutch train services are used to build the extended diachronic super-network. Table 1 shows the number for stops, railway stations and links in the partial public transport networks (before inclusion of transfer-walking links, transfer-free in-vehicle legs and private mode access&egress legs to railway stations and public transport stops). In-vehicle times and transfer times have been derived from these timetables, implying that possible service irregularities are not taken into account. Transfer-walking times (from/to train, UPT and private mode parking) are deduced from transfer-walking distances measured on site. Public transport (passive) transfer-waiting times are defined as differences between total transfer times (according to timetables) and transfer-walking times. Total transfer times between private modes and trains consist of a transfer-walking time and a parking / retrieving time.

Table 1 Number for stops, railway stations and links in the partial public transport networks (before inclusion of transfer-walking links, transfer-free in-vehicle legs and private mode access&egress legs to railway stations and public transport stops).

Partial network	*Number of stops or railway stations*	*Number of legs in non-extended public transport network (run-based)*
Haarlem	192	24132
Leiden	234	28209
The Hague	468	127311
Rotterdam	1004	128688
Dordrecht	468	27659
Delft	100	10221
Train network	131	15664

12 Calibration of choice set generation algorithm

The initial parameter set is based on observations of actual choice behaviour (DNTS) and on expert judgment. For example, minimum and maximum walking, cycling and driving distances to railway stations are set equal to 10% and 90% percentiles of observed walking, cycling and driving distances in the DNTS data. In a similar way, estimates of maximum UPT access travel times to railway stations and maximum acceptable walking distances to bus, tram and metro stops have been derived.

This initial parameter set is used to generate objective choice sets for a substantial number (± 50) of travellers and compared with:

- choice sets generated manually by the researcher (using travel experiences, maps and timetables);
- objective choice sets provided by Dutch Travel Information Services (OVR);
- alternatives mentioned by respondents (reported chosen trips and reported subjective choice sets).

The set of parameters is adapted slightly according to the comparison results. The process of generating and comparing choice sets is repeated several times, until satisfactory comparison results are obtained.

13 Quality of generated choice sets

After model calibration, objective choice sets have been generated for 708 respondents and compared to reported subjective choice sets (chosen routes as well as known alternatives) using the branch&bound algorithm separately for partial networks. Reported subjective alternatives appeared to be generated by the proposed choice set generation algorithm for the majority of the 708 respondents, thus implying a very high coverage level (87% for chosen alternatives and 81% for reported subjective alternatives). The coverage level is defined as the percentage of chosen or known alternatives generated by the algorithm).

Table 2 Violation of constraints for chosen alternatives (CA) and reported subjective alternatives (RSCS).

type of rule	*acceptable distance ranges*	*CA %*	*RSCS %*
walking distance	home-end: [0 km, 2 km] activity-end: [0 km, 3 km]	0.1 1.5	1.0 3.8
cycling distance	home-end: [0.8 km, 4 km] activity-end: [0.9 km, 5 km]	3.4 1.6	**6.8** 3.5
car distance	home-end: [1.5 km, 10 km] activity-end: [0.7 km, 12 km]	0.0 0.0	0.0 0.0
route-factor compared to minimum travel time (complete trips)	$T_{travel} \in$ [20 min, 45 min]: $[\rho_1,\rho_2]$=[0, 1.8] $T_{travel} \geq$ 45 min: $[\rho_1,\rho_2]$=[0.1.7]	2.6 0.9	3.3 0.0
route-factor compared to minimum number of transfers (complete trips)	$[\nu_1,\nu_2]$=[0, 2]	**8.9**	**15.5**
distance to tram stop	home-end: [0, 600 m] activity-end: [0, 600 m]	1.8 0.0	3.7 1.3
distance to bus stop	home-end: [0, 600 m] activity-end: [0, 600 m]	0.0 0.0	0.0 2.4
distance to metro stop	home-end: [0, 1000 m] activity-end: [0, 1000 m]	0.0 2.5	0.0 0.0
distance to local railway station	home-end: [0, 1.5 km] activity-end: [0, 1.5 km]	3.3 **5.2**	**6.3** **10.9**
distance to express railway station	home-end: [0, 5.25 km] activity-end: [0, 5.25 km]	0.0 0.0	0.0 0.0
distance to IC railway station	home-end: [0, 7.5 km] activity-end: [0, 7.5 km]	0.0 0.3	0.0 0.0

Besides computing the coverage level, we have tried to explain why certain reported alternatives were not generated. Such information can be used to further improve the choice set generation algorithm. To this end, we determined for each of the missing alternatives which constraint(s) had to be loosened (and to which extent) in order to generate the non-covered routes. Table 2 shows the percentages of cases of violation of different types of constraints for chosen alternatives (CA) and reported subjective alternatives (RSCS). The first column shows types of constraints, while the second column shows ranges for parameter values used in the constraints. If relevant, a difference has been made between the home-end and activity-end of a trip or between routes of different lengths. For example, the first constraints means that walking distances to railway stations should be at most 2 km and 3 km at the home-end and activity-end, respectively. Note that most constraints are based on 10% and 90% percentiles of observed travel behaviour, implying that a 20% violence of constraints (in columns 3 and 4) could be expected. It turns out that violation numbers for chosen as well as known alternatives are much lower than 20%. Violation numbers for reported subjective alternatives are slightly higher than for chosen alternatives. This can be explained from the fact that chosen alternatives relate to actual travel behaviour (closest to observed travel behaviour used to determine constraints and parameters). Travellers have less knowledge of reported subjective alternatives than of chosen alternatives and might never have used the reported subjective alternatives at all. The violation of constraints can in most cases be explained from typical geographical conditions, resulting in among other matters:

- longer access distances by foot or by bike to railway stations (e.g. travellers have to cross the river Maas in Rotterdam to reach a railway station);
- counter-intuitive modal combinations;
- UPT to railway stations that take more time than the fastest available UPT (e.g. a relatively slow metro-metro alternative in Rotterdam instead of faster direct bus or tram alternative).

These results clearly show that this branch&bound approach meets the practical requirements for choice set generation. The generated choice sets have been used successfully for the estimation of different types of choice models (see (*8*, *16*, *17*)).

14 Summary and conclusions

In this paper, a new choice set generation algorithm, applicable to mixed public and private transport networks, was presented. Its principle is to enumerate all feasible route alternatives satisfying a varied set of constraints (so-called constrained enumeration). In order to apply the algorithm first an extended diachronic super-network needs to be built. Subsequently, a set of alternatives is determined in two steps, namely a tree search step and a re-evaluation step. Prior to the development of the algorithm, constraints (i.e. logical, feasibility and behavioural constraints) were formulated that reflect observed choice behaviour. The developed generation algorithm can be used to generate different types of choice sets suitable for both choice model estimation and travel demand prediction purposes, and may apply to one specific traveller or a group of travellers. The type of choice set that results from the generation process is determined by the set of constraints that is used in the tree research and the re-evaluation step.

The choice set generation algorithm can be applied to a complete multi-modal network at once. However, this paper shows that by exploiting knowledge about the structure of multi-modal trips, separate application of the algorithm to partial networks and consecutive concatenation of subroutes into complete door-to-door trips substantially reduces computation times without resulting in incomplete choice sets. To illustrate this, the choice set generation algorithm was applied to generate choice sets to be used in the estimation of choice models for realistic, large-scale multi-modal transport networks. Comparing generated objective choice sets with reported chosen and known alternatives showed that generated objective choice sets indeed include these alternatives. Since the constraints included in the algorithm are based on observed behaviour, unrealistic alternatives are excluded from generated objective choice sets.

References

1. Abdelghany, K.F. & H.S. Mahmassani (2001). Dynamic Trip Assignment-Simulation Model for Inter-modal Transportation Networks, Transportation Research Record 1771, pp 52-60.
2. Abdelghany, K.F. & H.S. Mahmassani (1999). Multi-objective shortest path algorithm for large-scale inter-modal networks. Presented at the INFORMS Fall annual meeting, Philadelphia, USA.
3. Abdelghany, K.F. (2001). Stochastic dynamic traffic assignment for the inter-modal transportation networks with consistent information supply strategies. PhD thesis, Austin, University of Texas, USA.
4. Azevedo, J.A., M.E.O. Santos Costa, J.J.E.R. Silvestre Maderia & E.Q. Vieira Martins (1993). An algorithm for the ranking shortest paths, European Journal of Operational Research 69, 97-106.
5. Ben-Akiva, M.E., M.J. Bergman, A.J. Daly & R. Ramaswamy (1984). Modelling inter-urban route choice behaviour. In: J. Volmuller & R. Hamerslag (eds.). Proceedings Ninth International Symposium on Transportation and Traffic Theory, VNU Science Press, Utrecht, The Netherlands, 299-330.
6. Benjamins M., C.D.R. Lindveld & R. van Nes (2002). Multi-modal travel choice modeling using a super network approach, Paper # 02-2948, CD-ROM of the 81th Annual Meeting of the Transportation Research Board, Washington.
7. Bovy, P.H.L. & S. Fiorenzo-Catalano (2006). Stochastic route choice set generation; behavioural and probabilistic foundations. CD-ROM of the 11th International Conference on Travel Behaviour Research, Kyoto, August 2006.
8. Bovy, P.H.L. & S. Hoogendoorn-Lanser (2005). Modeling route choice behaviour in multi-modal transport networks, Transportation 32(4), pp 341-368.
9. Cascetta E. (2001). Transportation systems engineering: Theory and methods, Kluwer Academic Publishers, Dordrecht, The Netherlands.
10. De Cea, J. & E. Fernandez (1993). Transit assignment for congested public transport system: An assignment model, Transportation Science 27(2), 133-147.

11. De La Barra, T., B. Perez & J. Anez (1993). Multidimensional path search and assignment, Proceedings of the 21st PTRC Summer Meeting, Manchester, Great Britain.
12. Friedrich, M., I. Hofsass & S. Wekeck (2001). Timetable-based transit assignment using branch and bound, Transportation Research Record 1752, 100-107.
13. Fiorenzo-Catalano S., R. van Nes, P.H.L. Bovy (2004). Choice set generation for multi-modal travel analysis, European Journal of Transport and Infrastructure Research 4(2), 195-209.
14. Hoogendoorn-Lanser S. (2005). Modeling travel behaviour in multi-modal networks, PhD Thesis Delft University of Technology, TRAIL Thesis Series T2005/4, The Netherlands.
15. Hoogendoorn-Lanser, S., R. van Nes & P.H.L. Bovy (2006a), A rule-based approach to multi-modal choice set generation, CD-ROM of the 11th International Conference on Travel Behaviour Research, Kyoto, August 2006.
16. Hoogendoorn-Lanser, S., R. van Nes, S.P. Hoogendoorn & P.H.L. Bovy (2006b), Modeling transfers in multi-modal trips: Explaining correlations. Paper # 06-1442, CD-ROM of 85th Annual Meeting of Transportation research Board. Accepted for publication.
17. Hoogendoorn-Lanser, S., R. van Nes, P.H.L. Bovy & S.P. Hoogendoorn (2006c), Home-activity approach to multi-modal travel choice modeling. Paper # 06-1720, CD-ROM of 85th Annual Meeting of Transportation research Board. Accepted for publication.
18. Lawler, E.L. (1976). Combinatorial optimization: Networks and matroids, Holt, Rinehart & Winston, New York.
19. Nuzzolo, A. & F. Russo (1994). An equilibrium assignment model for intercity transit networks, Proceedings of the TRISTAN III conference, San Juan, Puerto Rico.
20. Park, D. & L.R. Rilett (1997). Identifying multiple and reasonable path in transportation networks: A heuristic approach, Transportation Research Records 1607, 31-37.
21. Prato, C.G.. & S. Bekhor (2006). Applying Branch-and-Bound Technique to Route Choice Set Generation, Paper # 06-0517, CD-ROM of the 85th Annual Meeting of the Transportation Research Board, Washington.
22. Ramming, M.S. (2002). Network knowledge and route choice, PhD thesis, MIT, Cambridge.
23. Shier, R.D. (1979). On algorithms for finding the k-shortest path in a network, Networks 9, 195-214.
24. Train, K.E. (2002). Discrete choice methods with simulation, Cambridge University Press.
25. Van der Waerden, P., A. Borgers & H.J.P. Timmermans (2004). Choice set composition in the context of pedestrians' route choice modeling, CD-ROM of the 83th Annual Meeting of the Transportation Research Board, Washington.
26. Van der Zijpp, N.J. & M.S. Fiorenzo Catalano (2005). Path enumeration by finding the constrained k-shortest, Transportation Research Part B 39(6), 545-563.
27. Van Nes, R. (2002). Design of multi-modal transport networks: A hierarchical approach, PhD Thesis Delft University of Technology, Trail Thesis Series T2002/5, Delft, The Netherlands.