Southeast University
School of Transportation

Urban Transportation Planning

Chinese-English course (2019)

Lecturer: Dr. Qiong Bao

Email: *baoqiong@seu.edu.cn*
Phone: *17351788445*

9:50AM, Friday, 10th May
School of Transportation | Southeast University of China | Campus Jiulonghu | Building Jizhong Y311

Lecture schedule

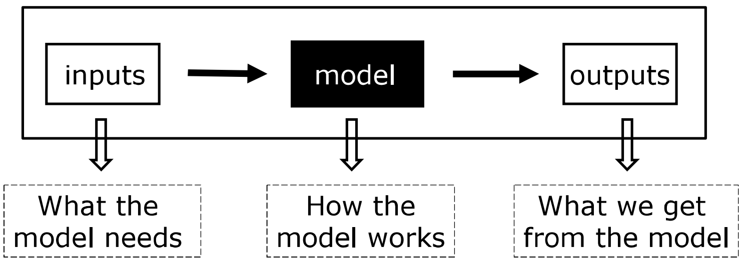
Lecture	Week	Date/Time	Topic
1	9	28 April 9：50-12：15	Transportation planning & demand and supply & trip-based model
2	10	5 May 9：50-12：15	ABM: data process
3	11	10 May 9：50-12：15	ABM: scheduling
4	12	17 May 9：50-12：15	ABM: uncertainty analysis
5	13	24 May 9：50-12：15	ABM: sensitivity analysis
6	14	31 May 9：50-12：15	Project Evaluation I
7	15	7 June 9：50-12：15	Festival
8	16	14 June 9：50-12：15	Project Evaluation II

Review of data processing in ABM

- Trip-based, tour-based, activity-based:
  - Difference among the trip-based approach, tour-based approach, and the activity-based approach;
- Activity-based model:
  - The output of ABM;
  - The input of ABM;
    - Synthetic population attribute data
      - IPF (Iterative proportional fitting algorithm)
      - IPU (Iterative proportional updating algorithm)

Methods of activity scheduling

Open the "black box"



Outline

- ABM scheduling process:
  - Scheduling process
  - Methodology
    - Rule based method
    - Decision tree technology
      - Gini Index
      - Information Gain
      - χ^2 contingency table statistic
  - Case study

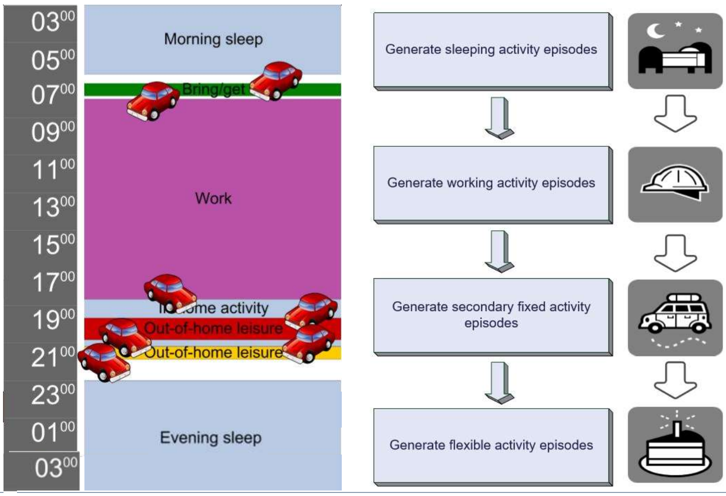
Methods of activity scheduling

- Famous activity based methods:
  - Constraint-based models
    (e.g. CARLA, MASTIC, PCATS)
  - Utility maximizing models
    (e.g. STARTCHILD, DAS, Tel Aviv)
  - Rule-based models (computational process models)
    (e.g. ALBATROSS, FEATHERS, TASHA, ADAPTS)
  - Micro-simulation models (mixed)
    (e.g. CEMDAP, HAPP, MATSim)

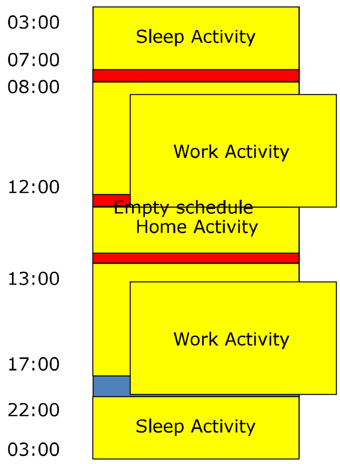
Methods of activity scheduling

- Utility maximizing method:
  - Models used to rely on utility-maximization.
  - The construction of models based strictly on the concept of utility maximization, neglecting substantial evidence relative to alternative decision strategies, e.g. habit formation, choice complexity.
- Rule-based model (computational process model):
  - Assume that choice behavior of individual is based on rules that are formed and continuously adapted through learning while the individual is interacting with the environment (reinforcement learning) or communicating with others (social learning).
    - rules are derived from decision trees
    - other rule-based learning algorithms can also be used

Agent-based schedule



FEATHERS scheduler



Action:

1. Firstly **ADD** sleep activity
2. Determine **<Y/N>** work activity
3. If <Y>, **CHOOSE** <Duration> work activity
4. CHOOSE <Y/N> **2nd** work activity
5. If <Y>, **CHOOSE** <Duration>
   - 2nd work activity
   - between work activities
6. CHOOSE <Beginning time> and <End time> **1st** work activity
7. CHOOSE <Beginning time> and <End time> **2nd** work activity
8. Estimate TT (travel time) **1st** work activity
9. Estimate TT after **1st** work activity
10. Estimate TT **2nd** work activity
11. If sum TTs < break time (reserve space) for new activity by inserting Home activity as dummy
12. Determine the location of activity

Constraints taken into account in ABM

- Basic constrains in ABM:
  - Situational constraints

  (e.g. activities can't happen in two places at the same time)
  - Institutional constraints (e.g. shopping only in the opening hours)
  - Household constraints (e.g. bringing children to school)
  - Spatial constraints

  (e.g. particular activities have to perform at particular locations)
  - Time constraints

  (e.g. activities require some minimum duration)
  - Spatial-temporal constraints

  (e.g. an individual cannot be at a particular location at the right time to conduct a particular activity)

Determine the schedule - General variables

Determine the schedule - Sleeping activity

Determine the schedule - Primary work/school activity

Inclusion? → Yes → Total duration → Number of episodes → Duration ratio episode 1 / episode 2 → Duration of break → Start time 1st episode

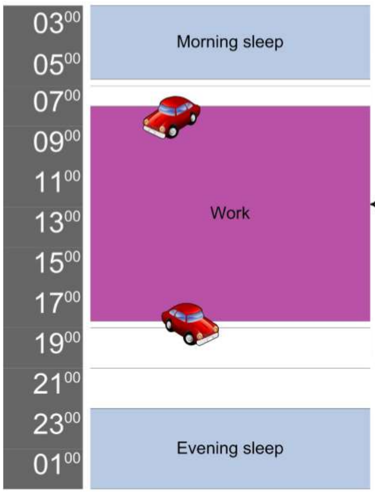
Condition variables:
- Number of employees working in different sectors within a certain range from home
- Distance to nearest employees working in different sectors
- Number of HH within a certain range from home
- Distance to nearest HH
- Sleep episodes characteristics
- Total duration work*
- Number of work episodes*
- Duration ratio of work episodes*
- Duration of break*

Determine the schedule - Primary work/school activity

Location of fixed activities → Transport mode work/school trips

Condition variables:
- Work/school episode characteristics
- Secondary fixed activity episodes characteristics
- Location characteristics
  * Availability of train connection
  * Car distance
  * ...
- Characteristics of transportation network, e.g.:
  * Cost ratio PT/Car
  * Parking price
  * ...

Determine the schedule - Primary work/school activity



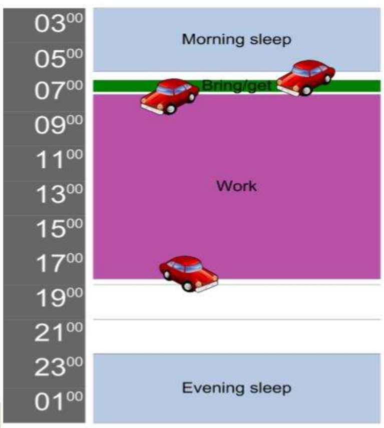
1. Include work episode:Y
2. Total duration: 650 min
3. Number of episodes: 1
4. Start time: 7:58
5. Location: 2326
6. Travel mode: car

Determine the schedule - Secondary fixed activities

Inclusion? → Yes → Number of episodes → Duration of episode → Linking episode to work/school? → Yes: Position of episode / No: Start time of episode → Location → Transport mode work tours

Condition variables:
- Number of employees working in different sectors within a certain range from home
- Distance to nearest employees working in different sectors
- Number of HH within a certain range from home
- Distance to nearest HH
- Sleep episodes characteristics
- Work episode characteristics
- Activity type
- Time available

Determine the schedule - Secondary fixed activities



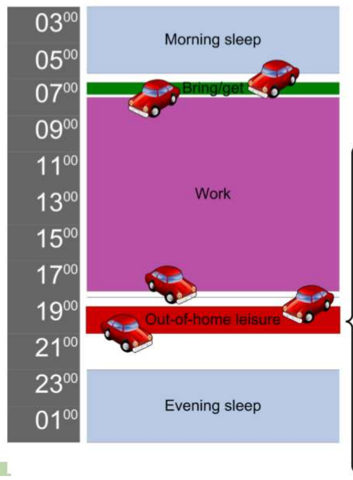
1. Include secondary fixed activity B/G?:Y
2. Number of episodes: 1
3. Duration: 6 min
4. Link to work?: Y
5. Position: before first work
6. Location: 2215
7. Transport mode: car

Determine the schedule - Flexible activities

Inclusion of an (additional) episode? → Yes → Travel party → Duration → Starting time → Trip chain → Location → Transport mode non-work tours

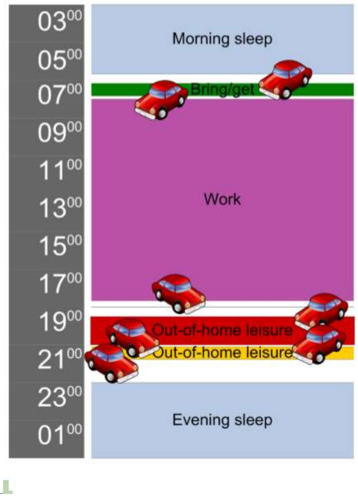
Condition variables:
- Number of employees working in different sectors within a certain range from home
- Distance to nearest employees working in different sectors
- Number of HH within a certain range from home
- Distance to nearest HH
- Activity type
- Available time
- Car availability
- Schedule characteristics:
  * Number of activities
  * Total duration
  * Feasibility of activity and duration
  * ...
- Travel characteristics (TM choice):
  * Travel time ratio PT/Car
  * Average parking cost
  * Distance of tour
  * ...

Determine the schedule - Flexible activities



1. Include flex activity OLeis?:Y
2. Travel party: alone
3. Duration: 90 min
4. Starting time: 19:24
5. Trip chain: single
6. Location: 2631
7. Transport mode: car

Determine the schedule - Flexible activities



1. Include additional flex activity OLeis?:Y
2. Travel party: alone
3. Duration: 40 min
4. Starting time: 21:09
5. Trip chain: single
6. Location: 2312
7. Transport mode: car

Rule-based method - decision tree technology

- Why we select decision tree technology?
  - Decision tree (DT) technology is one of the powerful and popular data mining tools for **classification** and **prediction**.
  - DTs represent rules, which can be understood by humans and used in knowledge discovery and decision making.
  - There is no requirement for the input data. The methodology could handle both continuous and categorical variables.

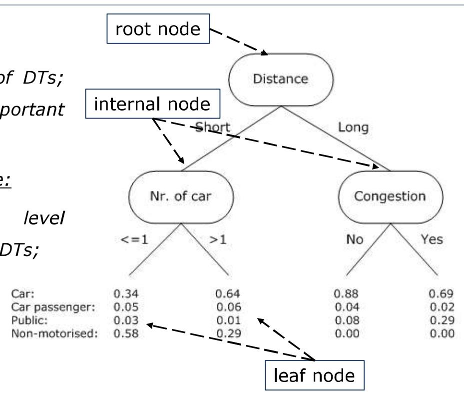
What is a decision tree?

__Root node:__

Start point of DTs; the most important attribute.

__Internal node:__

Secondary level attributes of DTs;



__Leaf node:__

Represents choice or choice distribution;

__Decision rules:__

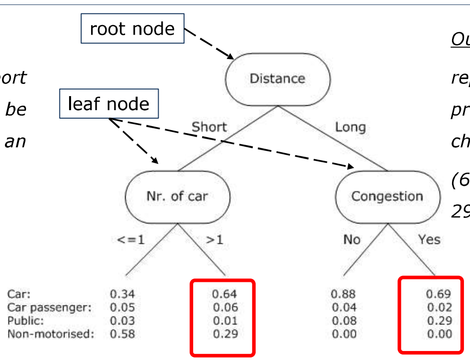
The path from root to leaf represents 'if-then' decision rules.

- A DT is a classifier in the form of a tree structure starting at the root of the tree and moving through it until a leaf node.

What is a decision tree?

__DT predicts:__

which transport mode will be selected by an individual.



__Output of DT:__

represents the probability of each choice alternative.

(64%, 6%, 1%, 29%)

- DT technology is a non-parametric learning method that uses a tree-like graph to predict the value of a target variable based on several input variables.

Requirements of DT technology

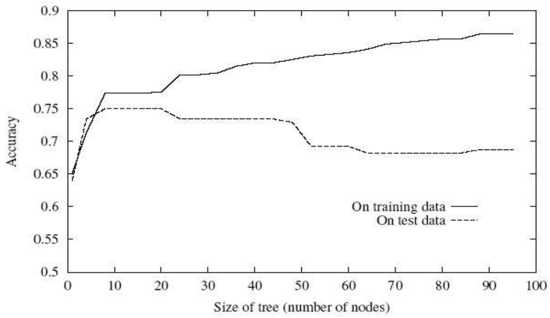
- Attribute-value description:
  - Variables must be expressible in terms of a fixed collection of properties or attributes (e.g., hot, mild, cold).
- Predefined choices or classes (target values):
  - the target alternative has discrete output values (e.g., 0/1 or multiclass)
- Sufficient data:
  - enough training data should be provided to learn the model.

Advantages and disadvantages of DT technology

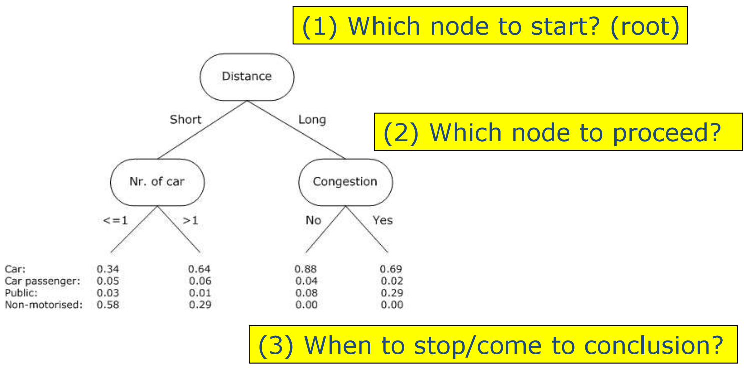
- Advantages:
  - Easy to use and understand
  - Produce rules that are easy to interpret & implement
  - Do not require the assumptions of statistical models
  - Can handle both continuous and categorical variables
  - Can work without extensive handling of missing data
- Disadvantages
  - May have a problem of overfitting
  - Since the process deals with one variable at a time, no way to capture interactions between variables

Overfitting

- One of the biggest problems with decision tree technology is overfitting



Research questions in DT building



Building decision trees

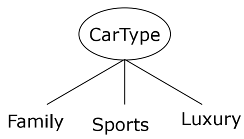
- Top-down tree construction
  - At start, all training records are at the root.
  - Partition the records recursively by choosing one attribute each time.
- Bottom-up tree pruning
  - Remove subtrees or branches, in a bottom-up manner, to improve the estimated accuracy on new cases.
- Strategy:
  - Split records based on an attribute test that optimizes certain criterion.
- Research questions:
  - Determine how to split the records (How to determine the best split?)
  - Determine when to stop splitting

Basic concepts

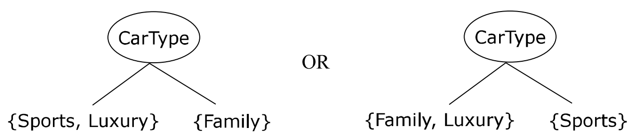
- Attribute types
  - Categorical
    - Nominal, e.g., Male vs. Female
    - Ordinal, e.g., Small-Medium-Large
  - Continuous
- The number of ways to split
  - 2-way split
  - Multi-way split

Splitting based on categorical attributes

- Multi-way split: Use as many partitions as distinct values.

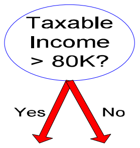


- Binary split: Divides values into two subsets.

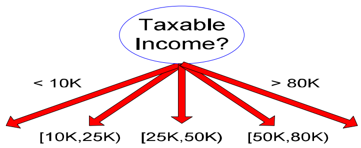


Splitting based on continuous attributes

- Discretization: to form an ordinal categorical attribute
  - ranges can be found by equal interval bucketing, equal frequency bucketing (percentiles), or clustering.



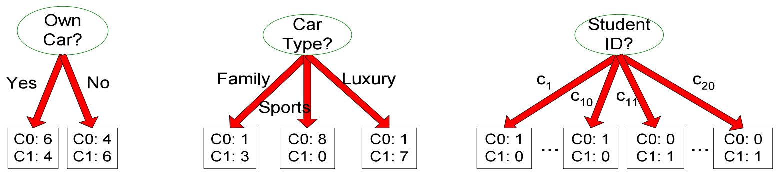
(i) Binary split



(ii) Multi-way split

How to determine the best split?

- Before Splitting: 10 records of class 0 (choice 1), 10 records of class 1 (choice 2)



- Which attribute should be select?

How to determine the best split?

- Principles:
  - Nodes with **homogeneous** class distribution are preferred
- Need a measure of node impurity:

C0: 5
C1: 5

Non-homogeneous,

High degree of impurity

C0: 9
C1: 1

Homogeneous,

Low degree of impurity

= High degree of purity

Main algorithms used in DT technology

- There are many specific DT algorithms:
  - ID3 (Iterative Dichotomiser 3)
  - C4.5 (Successor of ID3)
  - CART (Classification and Regression Tree)
  - CHAID (CHi-squared Automatic Interaction Detector).
    - Performs multi-level splits when computing classification trees.
  - MARS: extends decision trees to handle numerical data better
- Main difference: split criterion/attribute selection measure
  - Gini index (CART)
  - Information gain (ID3, C4.5)
  - χ^2 contingency table statistic (CHAID)

Gini Index

- Gini Index for a given node *t*:

$$Gini(t) = 1 - \sum_j [p(j|t)]^2$$

p(j | t) is the relative frequency of class j at node t.

C1	0
C2	6

P(C1) = 0/6 = 0　　P(C2) = 6/6 = 1

Gini = 1 – P(C1)^2 – P(C2)^2 = 1 – 0 – 1 = 0

C1	1
C2	5

P(C1) = 1/6　　P(C2) = 5/6

Gini = 1 – (1/6)^2 – (5/6)^2 = 0.278

C1	2
C2	4

P(C1) = 2/6　　P(C2) = 4/6

Gini = 1 – (2/6)^2 – (4/6)^2 = 0.444

C1	3
C2	3

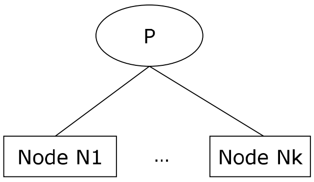
P(C1) = 3/6　　P(C2) = 3/6

Gini = 1 – (3/6)^2 – (3/6)^2 = 0.5

Gini Index

- Maximum ($1-1/n_c$)
  - when records are equally distributed among all classes, implying least interesting information.
- Minimum (0)
  - when all records belong to one class, implying most interesting information.
- Used in CART

Gini Index

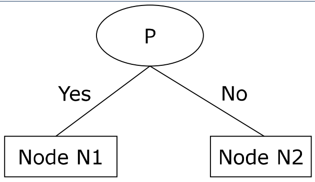


- When a node p is split into k partitions (children), the quality of split is computed as

$$Gini = \sum_{i=1}^{k} \frac{n_i}{n} Gini(i)$$

where, n_i = number of records at child i,
n = number of records at node p.

Example of Gini Index



Parent	
C1	6
C2	6

Gini(Parent)=0.5

Child N1	
C1	5
C2	2

Gini(N1) = 1 – (5/7)² – (2/7)² = 0.408

Child N2	
C1	1
C2	4

Gini (N2)= 1 – (1/5)² – (4/5)² = 0.32

Gini (Children)= 7/12 * 0.408 + 5/12 * 0.32= 0.371

Entropy-Maximizing Approach

Entropy-Maximizing Approach

- Example

zone	1	2
1	T_{11}	T_{12}
2	T_{21}	T_{22}

T={a, b, c, d, e}

If T_{11} contains 2 of these 5 trips, the possible combinations are $C_5^2 = 10$ as follows:

{a, b}; {a, c}; {a, d}; {a, e}; {b, c};
{b, d}; {b, e}; {c, d}; {c, e}; {d, e}.

Entropy-Maximizing Approach

- Now, suppose that there are totally *T* trips generated in *z* zones, then the possible combinations of trips in zone 1 that contains T_{11} trips ($T_{11} < T$) are:

$$\frac{T!}{T_{11}!(T-T_{11})!}$$

- We continue the process for the trips from zone 1 to zone 2 (T_{12}), for which (T-T_{11}) *trips* remain, so we have:

$$\frac{(T-T_{11})!}{T_{12}!(T-T_{11}-T_{12})!}$$

- …

Entropy-Maximizing Approach

- The product of all combinations gives the total number of possible microstates *W*:

$$W = \frac{T!}{T_{11}!(T-T_{11})!} \cdot \frac{(T-T_{11})!}{T_{12}!(T-T_{11}-T_{12})!} \text{L} = \frac{T!}{\prod_{ij} T_{ij}!}$$

- As it is assumed that all microstates are equally likely, the most probable T_{ij} would be the one that can be generated in a greater number of ways. So we try to maximize *W*.
- In doing so, we first take the logarithm of *W*:

$$\log W = \log \frac{T!}{\prod_{ij} T_{ij}!} = \log T! - \sum_{ij} \log T_{ij}!$$

Entropy-Maximizing Approach

- Stirling's approximation for $\log X! = X\log X - X$, can be used to make it easier to optimize W:

$$\log W = \log T! - \sum_{ij}(T_{ij}\log T_{ij} - T_{ij})$$

- Since the term $\log T!$ is usually a constant, it can be omitted from the optimization problem. The rest of the equation is often referred to as the *entropy function*:

$$\log W' = -\sum_{ij}(T_{ij}\log T_{ij} - T_{ij})$$

- Maximizing $\log W'$, subject to constraints corresponding to our knowledge about the macrostates, enables us to generate models to estimate the most likely T_{ij}.

Information Gain

- Entropy at a given node t:

$$Entropy(t) = -\sum_j p(j|t)\log p(j|t)$$

p(j | t) is the relative frequency of class j at node t.

Note: log(0) is not defined, but we evaluate 0*log(0) as zero

C1	0
C2	6

P(C1) = 0/6 = 0　　P(C2) = 6/6 = 1

Entropy (0,6)= – 0 log 0 – 1 log 1 = – 0 – 0 = 0

C1	1
C2	5

P(C1) = 1/6　　P(C2) = 5/6

Entropy (1,5) = – (1/6) $\log_2$ (1/6) – (5/6) $\log_2$ (5/6) = 0.65

C1	2
C2	4

P(C1) = 2/6　　P(C2) = 4/6

Entropy (2,4)= – (2/6) $\log_2$ (2/6) – (4/6) $\log_2$ (4/6) = 0.92

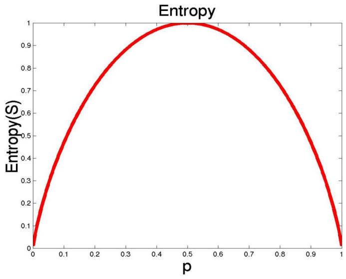
C1	3
C2	3

P(C1) = 3/6　　P(C2) = 3/6

Entropy (3,3)= – (3/6) $\log_2$ (3/6) – (3/6) $\log_2$ (3/6) = 1

Entropy

Entropy of a 2-class problem w.r.t. the portion of one class



- The entropy is 0 if the outcome is "certain".
  - i.e., all the records belong to one class.
- The entropy is maximum= $\log_2 n_c$ if records are equally distributed.
  - i.e., choice is equally possible.

Information Gain

- Information Gain:

$$Gain_{split} = Entropy(p) - \left(\sum_{i=1}^{k}\frac{n_i}{n}Entropy(i)\right)$$

where Parent Node p is split into k partitions;

n_i is the number of records in partition i

- – Measures reduction in Entropy achieved because of the split.
- – Choose the split that achieves the largest reduction (maximize GAIN)
- – Used in ID3 and C4.5
- Strategy: choose attribute results in greatest information gain

Example of information gain

Outlook	Temperature	Humidity	Windy	Class
sunny	hot	high	false	N
sunny	hot	high	true	N
overcast	hot	high	false	P
rain	mild	high	false	P
rain	cool	normal	false	P
rain	cool	normal	true	N
overcast	cool	normal	true	P
sunny	mild	high	false	N
sunny	cool	normal	false	P
rain	mild	normal	false	P
sunny	mild	normal	true	P
overcast	mild	high	true	P
overcast	hot	normal	false	P
rain	mild	high	true	N

For parent (before split):

P(C1)= 9/14　　P(C2)= 5/14

E(p)=Entropy (9+,5-)= – (9/14) $\log_2$ (9/14) – (5/14) $\log_2$ (5/14) = 0.940

Example of information gain

Compute the entropy for the attributes:

C1	2+
C2	3-

P(C1) = 2/5　　P(C2) = 3/5

Entropy (2,3)= – (2/5) $\log_2$ (2/5) – (3/5) $\log_2$ (3/5) =0.971

C1	4+
C2	0

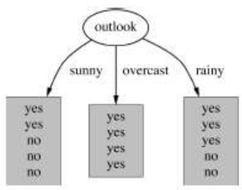
P(C1) = 4/4　　P(C2) = 0/4

Entropy (4,0) = – 1$\log_2$ 1 – 0 $\log_2$ 0 = 0

C1	3+
C2	2-

P(C1) = 3/5　　P(C2) = 2/5

Entropy (3,2)= – (3/5) $\log_2$ (3/5) – (2/5) $\log_2$ (2/5) =0.971



E(outlook)=5/14 * E(2,3)+4/14 * E(4,0)+5/14 * E(3,2)=0.694

Gain (outlook)=E(p)-E(outlook)=0.246

$$Gain(temperature) = 0.029$$
$$Gain(humidity) = 0.151$$
$$Gain(windy) = 0.048$$

χ^2 contingency table statistic

- Used in Chi-squared automatic interaction detector (CHAID)
  - Preparing predictors: create categorical predictors out of any continuous predictors by dividing the respective continuous distributions into a number of categories with an approximately equal number of observations.
  - Merging categories: cycle through the predictors to determine for each predictor the pair of categories that is least significantly different with respect to the dependent variable
    - where the dependent variable is categorical as well, it will compute **a Chi-square test**.

χ^2 contingency table statistic

- If the statistical significance for the respective pair of predictor categories is significant, then it will compute a Bonferroni-adjusted p-value for the set of categories for the respective predictor.
- Selecting the split variable: choose the variable with the smallest adjusted p-value, **i.e., the predictor variable that will yield the most significant split**
  - if the smallest adjusted p value for any predictor is greater than some **alpha-to-split value**, then no further splits will be performed and the respective node is a terminal node.

Determine when to stop splitting

- Stopping criteria for tree building
  - Stop expanding a node when all the records belong to the same class
  - Stop expanding a node when all the records have similar attribute values

Overfitting

- The generated tree may overfit the training data
  - too many branches
  - poor accuracy for unseen samples
- Reasons for overfitting
  - noise and outliers
  - too little training data
  - local maxima in the greedy search

Overfitting

- Two approaches:
  - prepruning: Stop the algorithm before it becomes a fully-grown tree
    - Stop if number of instances is less than some user-specified threshold
    - Stop if expanding the current node does not improve impurity measures (e.g., Gini or information gain).
    - Stop if class distribution of instances are independent of the available features (e.g., using χ^2 test)
  - postpruning: Remove branches from a “fully grown” tree
    - Trim the nodes of the decision tree in a bottom-up fashion
    - If generalization error improves after trimming, replace sub-tree by a leaf node.

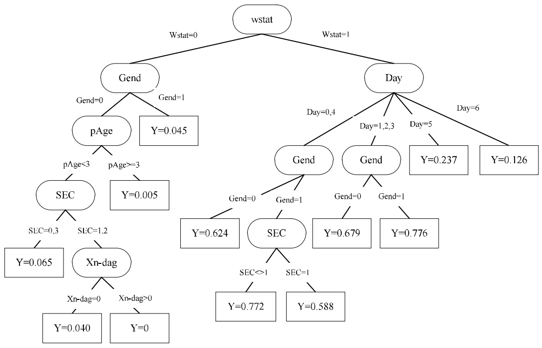
Other Methods

- Random Forest
- Bayesian classification
- Neural networks
- k-nearest neighbor classifier
- Case-based reasoning
- Genetic algorithm
- Rough set approach
- Fuzzy set approach

Rule-based method - Decision Tree

- Decision trees in FEATHERS:
  - a sequence of 26 decision trees based on CHAID: used in the scheduling process to predict individual's activities and travel choices.
  - It mainly used to determine for each individual: activity type (e.g., in-home, working, shopping, visiting), duration, start time, location, and transport mode.
- Pros: could represent the complicated interactions between different attributes (e.g. condition variables).
- Cons: hard to interpret the impacts of these attributes on the choice alternative.

DT: concerning work activity choice



Condition variable	Definition
Wstat	Work status
Day	Day of the week
Gend	Gender of individual
pAge	Age of the person
SEC	Income
Xn-dag	Number of employees

There are 6 input condition variables involved in this decision tree and collectively determine whether a work-related activity will be implemented or not.

Target Y: the probability of making the final decision, i.e. implement work activity or not

Case study of ABM

- Case 1: Leuven light rail project

  (Transportation supply planning)
  - Background: in this case study, the city of Leuven is selected as a study area to perform prediction of the travel demand. The city owns quite large transport potential, and is yet reasonably compact in size. Nevertheless, the city has no urban or regional light rail system so far.
- Case 2: Fuel-cost increase scenario

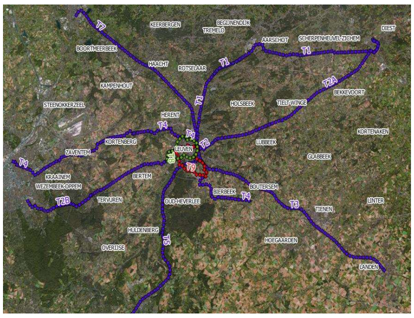
  (Travel demand management)
  - an increase of the fuel price by 20%
  - Influence on travel demand of different modes
  - Influence on road safety

Case study: Leuven light rail project

- Objective:
  - to investigate the potential impact of a regional light rail system on travel demand.
- Perform in 2 scenarios:
  - Null scenario:

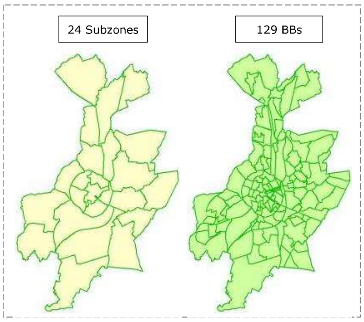
    limited to the situation where no light rail network is included. The public transport network contains only train and bus lines.
  - Light rail scenario:

    integrated with the proposed light rail network information.



The proposed regional light rail network surrounding Leuven consists of 10 different lines with a total length of about 250km.

City of Leuven



Geographical structure of Leuven

- Municipality Leuven:
  - a capital city in Flemish region of Belgium.
  - Superzone (ID: 124), consists of 24 Subzones and 129 BBs.

Restrain the size of study areas

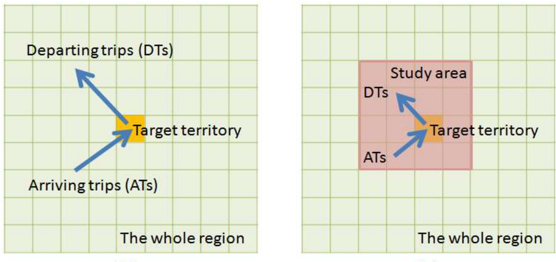
- One of the practical limitations of applying ABM: computation time. *especially when large amount of population and detailed geographical level are taken into account.*
  - 16 hours: for a single model run based on 10% of the full population of Flanders at BB level.
  - at least two days: based on 50% of the full population of Flanders at BB level.
  - the computation time will be magnified dramatically: by considering the effects of stochastic errors and therefore multiple model runs are required.

Restrain the size of study areas

- It is often the case that merely a small territory (e.g., a municipality) rather than the whole country is the focus of a specific study.
- Therefore, a relatively small study area surrounding the target territory is needed for investigation rather than to take the whole region into account.
- Solution (tradeoff): To reduce the computation time, one tradeoff can be made in the application. Which is to restrain the size of the study area and conduct the computation only for the selected region.
  – specific study: only a small territory (e.g., a municipality) rather than the whole Flanders is the focus.

Research question

Visualization of the DTs and ATs w.r.t. the target territory for the whole Flanders and for the study area.



(a) (b)

– In practice, to predict the travel demand variables, e.g., the total nr. of trips, ...he

Research question: What should be the minimum size of the study area surrounding the target territory and how to determine it?

– ...et territory within which most of the DTs and the ATs are generated, it is then not necessary to take the whole Flanders into account (see Figure (b)).

Methodology

- Approach:

  Apply an iterative approach to determine the minimum size of the study area.
- Study area:

  by adding one more zone to the target territory from close to distant until the accuracy level is reached.
- Accuracy of the model:

  defined as the difference (D_{ij}) between the occurrence of both the departing trips (DTs) and the arriving trips (ATs) derived based on the rebuilt study area (SA) and that based on the whole Flanders.

$$D_{ij} = \frac{\sqrt{(\#\,DTs\ in\ SA_{ij} - \#\,DTs\ in\ Flanders)^2 + (\#\,ATs\ in\ SA_{ij} - \#\,ATs\ in\ Flanders)^2}}{\sqrt{(\#\,DTs\ in\ Flanders)^2 + (\#\,ATs\ in\ Flanders)^2}}$$

$$A_{ij} = (1 - D_{ij}) \times 100\%$$

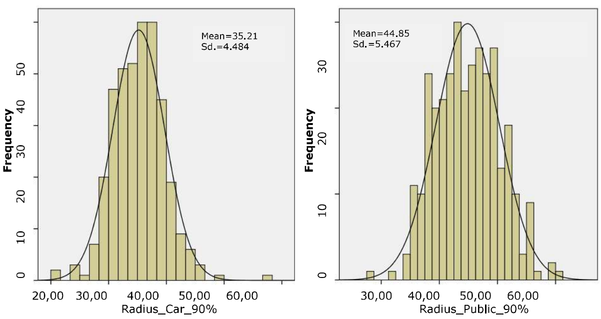
j: the added zone with the shortest centroid distance to the target territory i

Methodology

- In this study, we investigate the minimum size of the study area needed for each of the 327 municipalities (i.e., Superzone) in Flanders.
  – FEATHERS is executed at BB level,
  – based on 50% of the full population of Flanders,
  – DTs and ATs of each municipality is aggregated based on the BB level data,
  – accuracy level is predefined as 90%,
  – take into account 2 transport modes, i.e. car as driver and public transport.

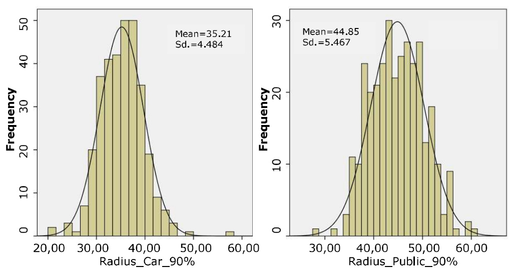
Distribution of the study area radius

The distribution of the minimum radius for 327 municipalities in Flanders w.r.t. car as driver mode and public transport mode



- The radius for both modes follows the normal distribution.
- Radius: refers in particular to the centroid distance between the last added municipality and the target one.

Distribution of the study area radius

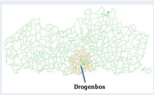
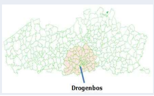
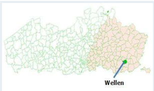


The distribution of the minimum radius for 327 municipalities in Flanders w.r.t. car as driver mode and public transport mode

- Average radius: car as driver mode < public transport mode (35.21km < 44.85km).
- In other words, to achieve the same 90% accuracy level, public transport mode in general needs a relatively larger study area compared to car as driver mode.

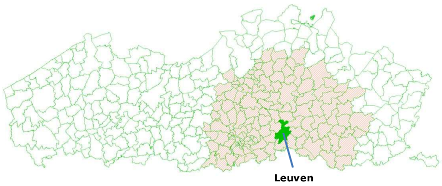
Validation of extreme cases

The municipality with the shortest and the longest radius of study area w.r.t. 2 transport modes

Travel mode	Municipality	Radius (km)		No. of persons within the SA (50%)	No. of households within the SA (50%)	Accuracy level	Running Time (hour)	Time saving
Car as driver		Shortest:	20.47	1,033,542	676,916	91.9%	15	76.7%
		Longest:	57.97	853,630	494,254	81.9%	10	84.5%
Public transport		Shortest:	27.97	1,307,798	837,150	96.2%	21	67.4%
		Longest:	60.40	1,215,802	702,082	86.9%	17	73.6%

64.5 hours for the whole Flanders

Study area of Leuven



Investigated study area for Leuven municipality

- Investigate the study area surrounding Leuven:
  - execute FEATHERS (ABM),
  - based on 50% of the full population of Flanders,
  - by comparing the travel demand difference between the study area and the whole Flanders, and predefine the accuracy level as 90%,
  - study area for Leuven: include 135 municipalities w.r.t the public transport.
  - It should be noted here: The restrained study area covers the proposed whole light rail network.

Validation of Leuven study area

Comparison between Leuven study area and the whole Flanders w.r.t. public transport mode

Transport mode	Study area	No. of persons within the SA (50%)	No. of households within the SA (50%)	Running Time (hour)	Accuracy level	Time saving
Public transport	Leuven study area	1021660	633216	28.8	89.6%	55%
	Flanders	2395514	1449213	64.5	-	-

- Validation results of the study area compared with the whole Flanders w.r.t. the public transport mode:
  - The data shows in the study area of Leuven, the population is almost equal to half of that of whole Flanders.
  - Results show a high accuracy level (89.6%). And 55% of the computation time could be saved compared to that for whole Flanders.

Comparison of two scenarios

The predicted daily travel demand (i.e., the number of trips) of 4 transport modes for 2 scenarios

		Car as driver	Car as passenger	Non-motorised mode	Public transport
Null scenario	Run1	1,384,997	333,402	818,210	154,913
	Run2	1,386,325	332,784	818,387	155,000
	Run3	1,383,779	332,003	820,958	154,692
	Run4	1,383,371	333,463	815,492	156,014
	Run5	1,384,407	334,249	817,158	154,650
	Run6	1,384,264	333,080	818,885	154,290
	Run7	1,380,316	333,205	816,984	155,437
	Average	**1,383,923**	**333,169**	**818,011**	**154,999**
Light rail scenario	Run1	1,384,585	334,431	779,447	167,609
	Run2	1,386,632	335,106	781,763	166,223
	Run3	1,386,480	331,012	780,364	167,564
	Run4	1,386,265	333,690	782,130	167,427
	Run5	1,388,175	332,199	781,431	166,685
	Run6	1,385,187	333,697	781,025	167,742
	Run7	1,388,022	332,826	779,965	166,980
	Average	**1,386,478**	**333,280**	**780,875**	**167,176**
% change		**[-0.13%, 0.57%] 0.18%**	**[-0.97%, 0.93%] 0.03%**	**[-5.06%, -4,09%] -4.54%**	**[6.54%, 8.72%] 7.86%**

Lecture summary

- ABM scheduling process:
  - Scheduling process
  - Methodology
    - Rule based method
    - Decision tree technology
      - Gini Index
      - Information Gain
      - χ^2 contingency table statistic
  - Case study

Questions

1. Please describe the commonly used methods in activity scheduling.
2. Calculation of Gini Index
3. Calculation of Information Gain

Thanks for your attention!

baoqiong@seu.edu.cn