Modelling Activity-Diary Data: Complexity or Parsimony?

Elke Moons

Transportation Research Institute -

Instituut voor Mobiliteit (IMOB)

Universiteit Hasselt - campus Diepenbeek

Belgium

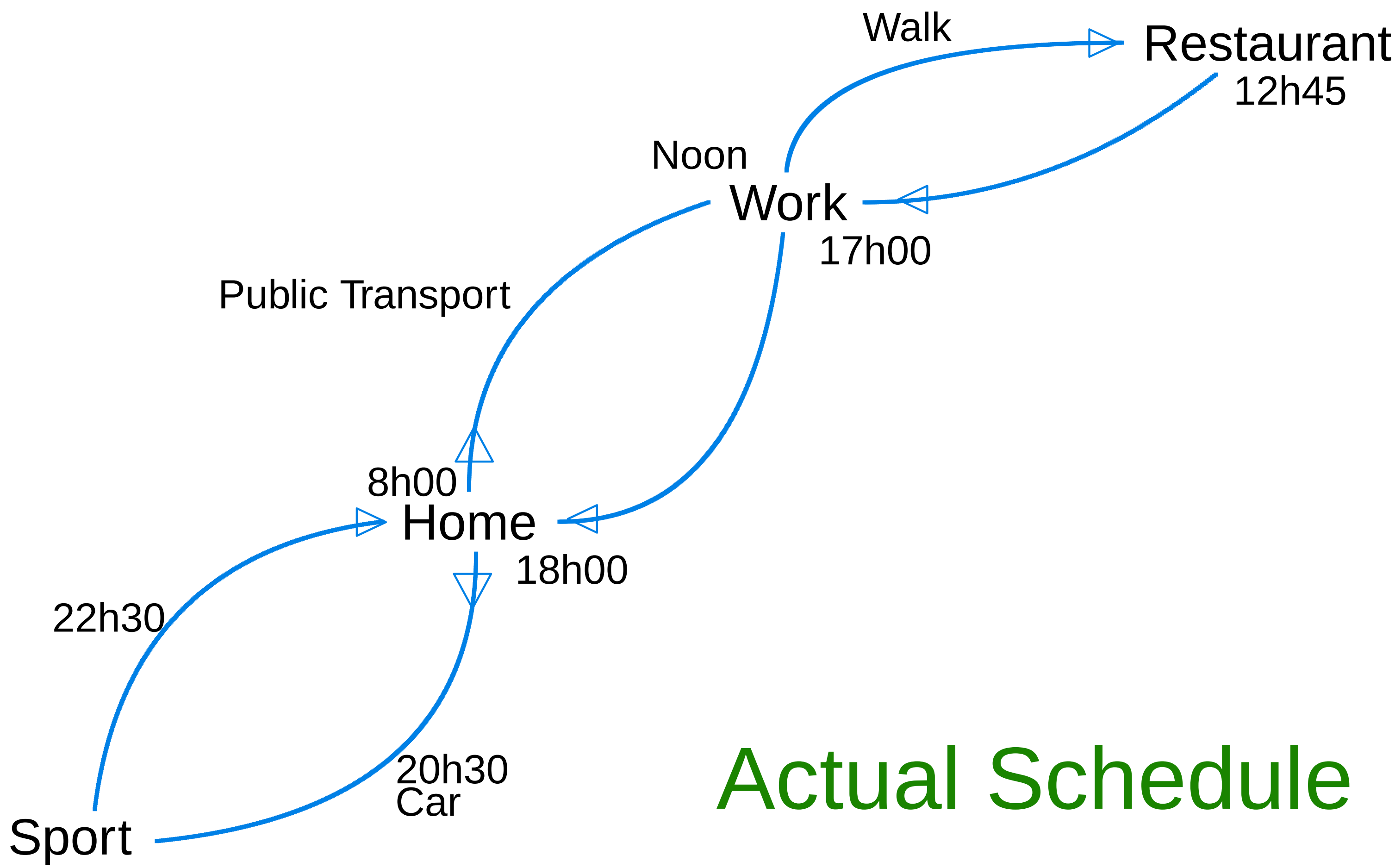
Overview

- Activity-Based Models
  - History
  - Albatross Model
  - Complexity or Parsimony?
  - Application
- Mode Choice Models
  - Tree-based Lack-of-Fit Test
  - Semi- & Nonlinear Models

History: Transport Modelling

- 50's: car use increased $\Rightarrow$ Need for models to predict travel demand $\Rightarrow$ Trip-based (Four Step) Model
- Drawback: focus on individual trips, relationship between trips is ignored $\Rightarrow$ mid 70's: Tour-based Models
- Home- & work-based chains $\Rightarrow$ Drawback: no relationship between travel and non-travel aspects
$\Rightarrow$ 90's: Activity-based Models
- Shift from travel demand to activity schedules: travel is derived demand $\Rightarrow$ purpose is to predict which activities are conducted where, when, for how long, with whom and which transport mode was used to arrive there

History: Transport Modelling (2)



Actual Schedule

History: Transport Modelling (3)

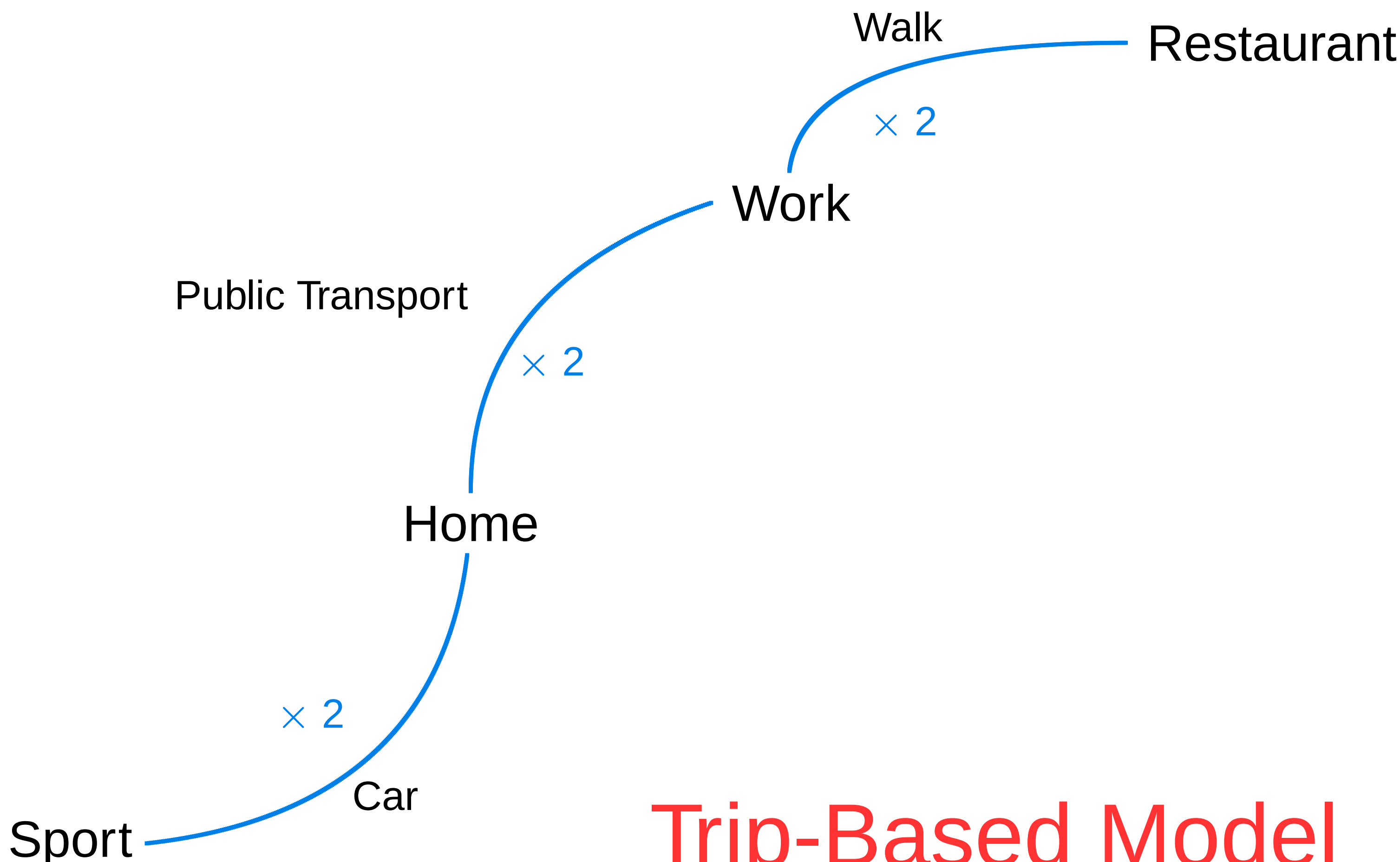
Restaurant

Work

Home

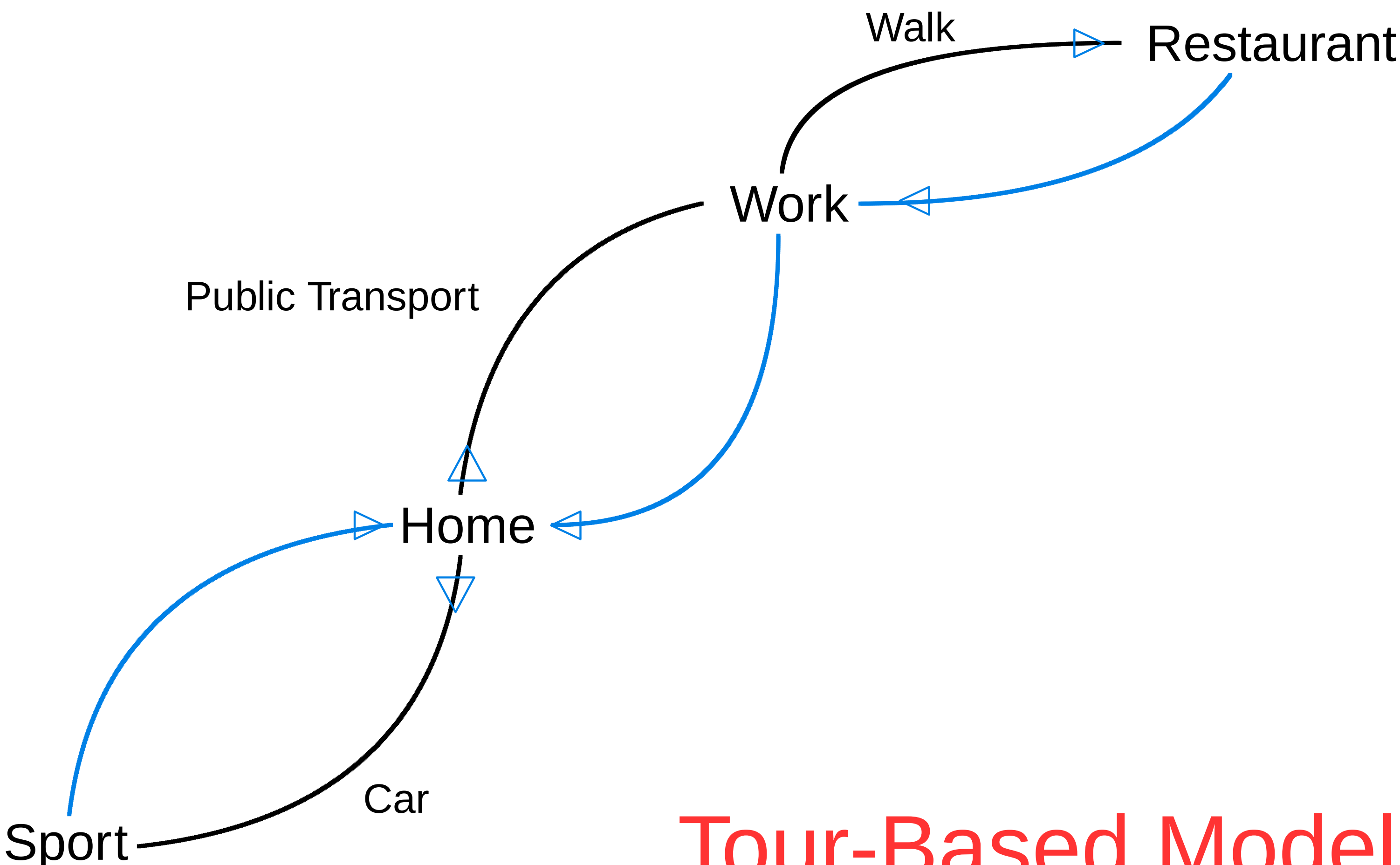
Sport

History: Transport Modelling (3)



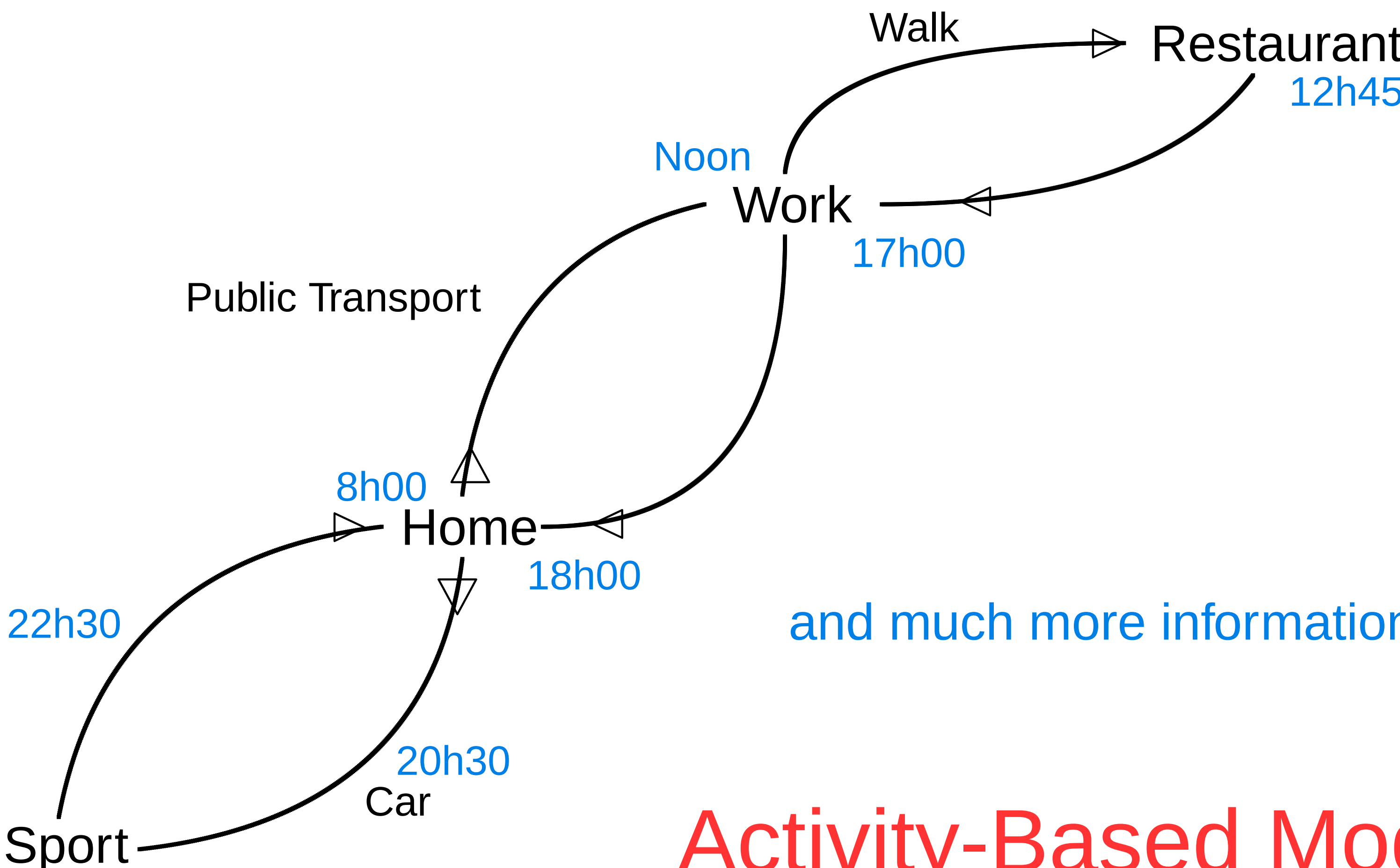
Trip-Based Model

History: Transport Modelling (3)



Tour-Based Model

History: Transport Modelling (3)



and much more information...

Activity-Based Model

History: Activity-Based Models

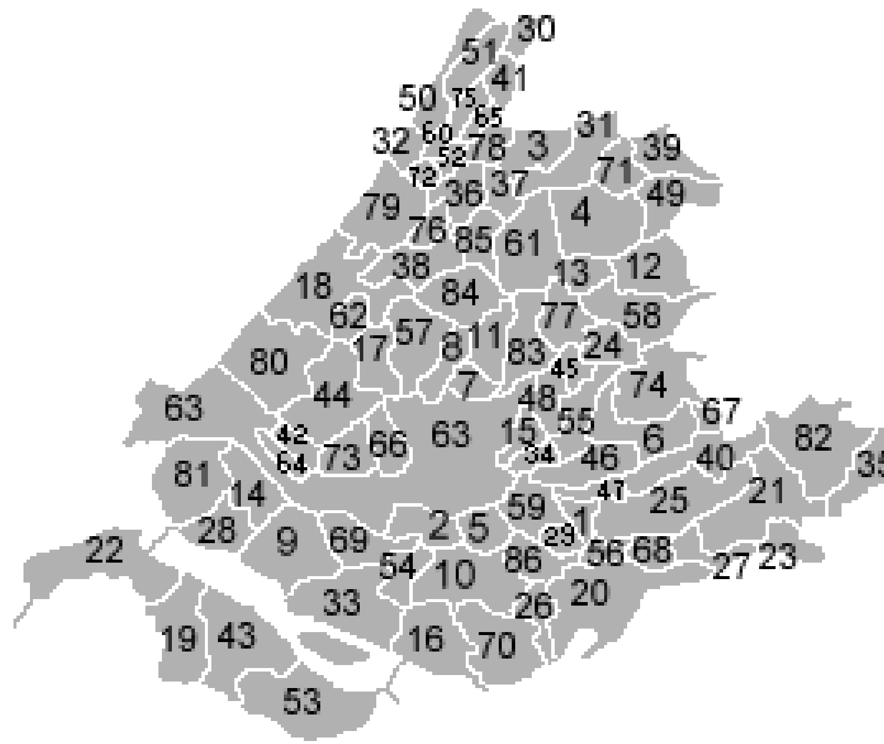
- Constraints-based Models
  - Travel time
  - Duration & timing
  - Sequencing
- Utility-maximising or Simultaneous Models
  - One day pattern
  - Simultaneous
  - Econometric
- Computational Process or Sequential Models
  - How does one arrive at the sequence of activities?
  - ‘Optimal’ choice
  - IF - THEN rules

Albatross Model

- A Learning BAsed TRansportation Oriented Simulation System
- Only fully operational computational process model to date
- Dutch Ministry of Transportation
- Data
  - Collected in Hendrik-Ido-Ambacht and Zwijndrecht (South-Rotterdam region)
  - In February 1997, random sample of 1649 respondents
  - Activity Diary: nature, location, day, begin & end time, transport mode (chain), travel times, accompanying persons
  - General characteristics: age, gender, type of household, children, car availability (ratio), etc.

(Arentze and Timmermans, 2000)

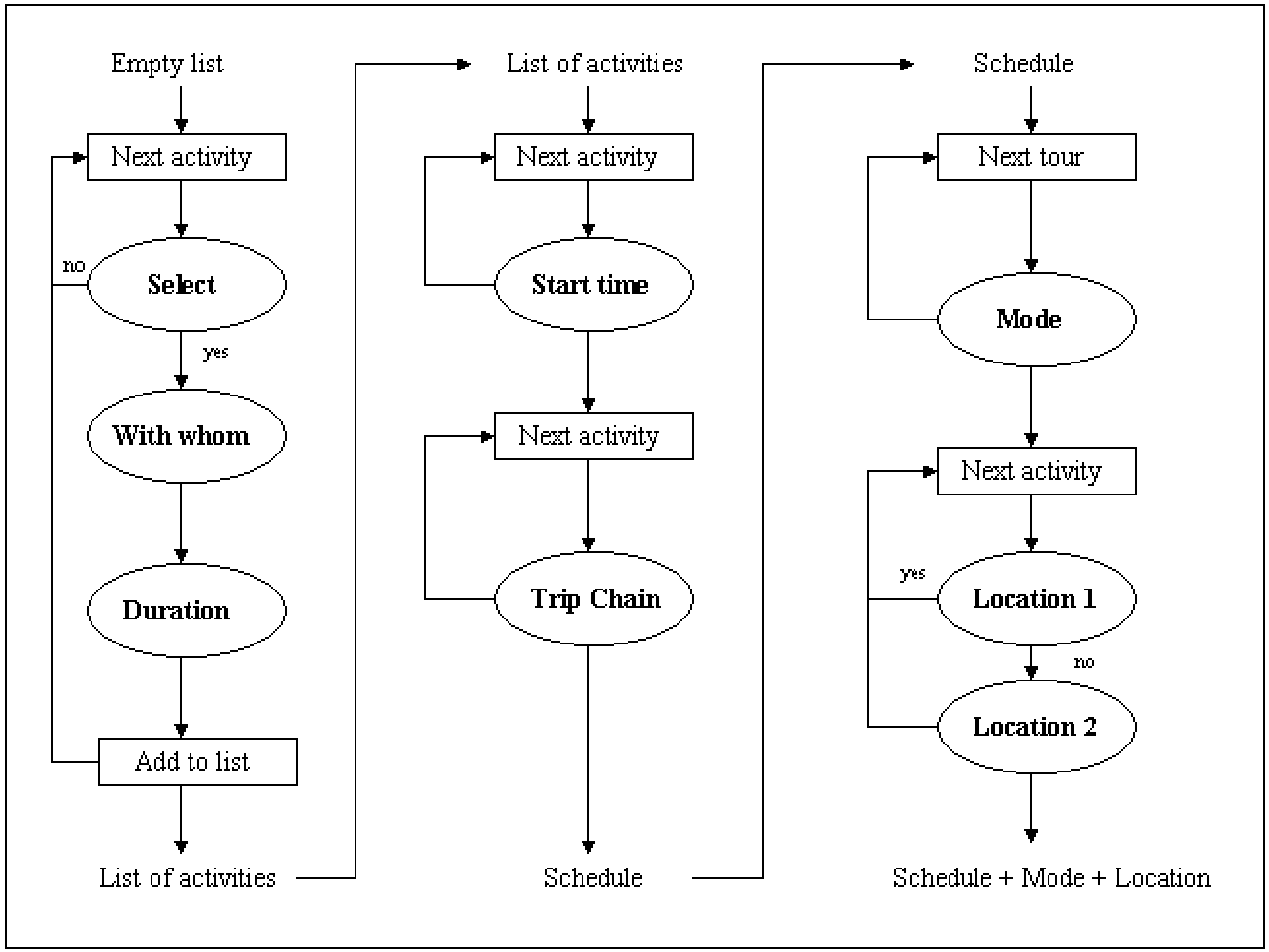
Albatross Model (2)



Albatross Model (3)

- Purpose: Schedule activities
- Interactions between individuals in household
- Constraints
  - Situational
  - Institutional
  - Household
  - Spatial
  - Time
- 9 dimensions or choice facets that need to be modelled!

Albatross Model (4)



Model Comparison

- Choice Facet Level: Complexity & Accuracy

Dimension	$\sharp$ cases	$\sharp$ alternatives	$\sharp$ variables
Mode for Work	858	4	32
Selection	14190	2	40
Travel Party	2970	3	39
Duration	2970	3	41
Start Time	2970	6	63
Trip Chaining	2651	4	53
Mode Other	2602	4	35
Location1	2112	7	28
Location2	1027	6	28

Model Comparison (2)

- Activity Pattern Level: Sequence Alignment Methods (Joh *et al.*, 2001, 2002)
  - Dissimilarity measures between observed and predicted sequences (type, location, mode & travel party)
  - Effort required to make two sequences identical
    $\Rightarrow$ lower SAM measures are better
  - Insertion, deletion and substitution operators
  - 4 uni-dimensional SAM, UDSAM & MDSAM
- Trip Matrix Level: correlation coefficient
  - Origin-Destination matrix: trip is basic unit
  - Disaggregation on day, primary activity and transport mode

Complexity or Parsimony

- Why (or why not)?
  - High predictive performance: Complex $\leftrightarrow$ Strongest effects, no (disturbing) details: Simple - Parsimonious Models
  - Middle ages: Occam's razor (but beware!)
  - Psychology: Human Behaviour
  - What is best in Activity-Based Transportation context?
- Apply complex and parsimonious models within Albatross and compare results

Complexity or Parsimony (2)

- 2 Different Ways of Attaining Parsimony:
  - Simple Heuristics
    - One R
    - Naïve Bayes
  - Feature / Variable Selection: Relief-F
- Combination of Simple Models
  - Bagging
  - Boosting

Simple Heuristics

- One R(ule)
  - For each explanatory variable a: find the majority class c of the response per value v in the domain of the predictor a
  - Rule: If a has value v then assign class c
  - Choose the rule with the highest accuracy
  - *Example on Transport Mode*

    Distance: Short $\longrightarrow$ Slow transport
    Long $\longrightarrow$ Car
    Accuracy of 70%

    Parking: Bad $\longrightarrow$ Public Transport
    Good $\longrightarrow$ Car
    Accuracy of 45%

(Holte, 1993)

Simple Heuristics (2)

- Naïve Bayes
  - Bayes rule & naïve assumption of conditional independence

$$
\begin{aligned}
P(C_i|X) &= \frac{P(X|C_i) \cdot P(C_i)}{P(X)} && \text{by Bayes rule} \\
&\propto P(A_1, \ldots, A_n|C_i) \cdot P(C_i) \\
&\propto \prod_{j=1}^{n} P(A_j|C_i) \cdot P(C_i)
\end{aligned}
$$

  - Prediction: Class with largest probability
  - Robust to violations of independence assumption
- Zero R: choose the majority class

(Langley *et al.*, 1992)

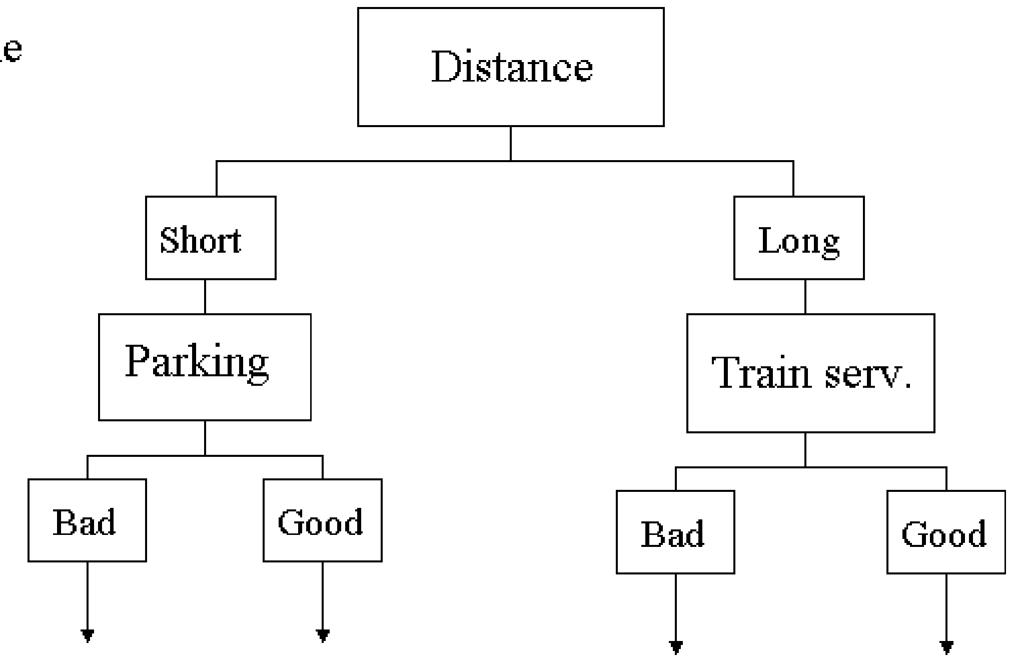
Decision Tree Induction

- Complex model
- Root node → Daughter nodes → Terminal node
- Stopping rule or pruning strategies
- Splitting criterion
  - Chi-square measure: CHAID (Kass, 1980)
  - Entropy: CART (Breiman *et al.*, 1984)
  - Gain ratio: C4.5 (Quinlan, 1993)
- Binary or multi-way splits

Decision Tree Induction (2)

Decision tree

example



Slow	0.90	0.71	0	0
Car	0.08	0.29	0.70	0.40
Car Pool	0	0	0.30	0.20
Public	0.02	0	0	0.40

Feature / Variable Selection

- Too many variables: tree may over-fit the data (unstable)
- Tree structure: sensitive to highly correlated variables (problem of multicollinearity)
- Smaller tree (exclude redundant variables) $\Rightarrow$ higher probability of misclassification?
- Relief-F
  - Ranks variables according to their ability in distinguishing between the classes of the response, while mutually the variables should have a correlation as low as possible
  - Handles multi-class response variables

(Kononenko, 1994)

Application - Part I

- Simple Heuristics & Trees after FS & Complex Models in Albatross
- No dramatic differences at Choice Facet Level: CHAID and C4.5 perform best, but FS approach and simple heuristics come close (difference of 5-12% in accuracy)
- Important variables for
  - Mode for work: Travel time by bike (!)
  - Activity Selection: Type of activity, day of the week
  - Travel Party: Household composition, day, type of activity
  - Duration: Travel party, type of activity
  - Start Time: Available time
  - Trip Chain: Enough time to include activity in schedule?
  - Mode Other: Household and transport characteristics
  - Location 1 & 2: Time-related and transport mode

Application - Part I (2)

- Activity Pattern Level

Measure	Mean Distance					
	CHAID	Zero R	One R	NB	Full	FS
SAM (AT)	2.777	3.130	3.027	3.022	2.903	2.929
SAM (TP)	3.168	3.464	3.312	3.225	3.210	3.208
SAM (L)	3.127	3.251	3.184	3.107	3.166	3.033
SAM (TM)	4.626	5.018	4.592	4.781	4.497	4.600
UDSAM	16.475	17.993	17.142	17.156	16.678	16.699
MDSAM	8.333	8.951	8.474	8.671	8.374	8.373

  - One R, FS, C4.5 ('full') and CHAID perform well
  - 1 substitution and/or 1-2 insertion/deletion operations suffice
  - MDSAM: reasonable amount of association between elements across dimensions

Application - Part I (3)

- Trip Matrix Level
  - Total number of trips from certain origin to certain destination

Matrix	Observed	Predicted					
		Zero R	One R	NB	CHAID	Full	FS
Day							
Weekday	2359	2572	2510	2414	2454	2564	2413
Saturday	356	276	277	380	416	331	262
Sunday	287	203	248	172	297	214	157
Transport mode							
Car	1609	1580	1609	1465	1771	1573	1466
Slow	814	1020	1013	1031	999	1038	920
Public	79	83	81	102	83	113	107
Car pas.	294	356	321	357	305	375	333

Application - Part I (4)

Matrix	Observed	Predicted					
		Zero R	One R	NB	CHAID	Full	FS
Primary activity							
Work out	970	901	937	944	960	942	971
Med. visit	44	32	34	30	36	36	38
Bring/get	538	485	497	508	526	524	502
Non-leis.	106	83	89	91	94	89	85
Non-groc.	251	287	204	320	231	260	220
Grocery	319	281	316	321	398	399	264
Leisure	466	329	440	264	447	371	269
Soc. visit	241	353	378	304	352	331	280
Service	59	243	95	133	96	121	168
Other out	18	63	52	61	36	48	46

Application - Part I (5)

- Trip Matrix Level

Matrix	ρ(o, p)					
	CHAID	Zero R	One R	NB	Full	FS
None	0.937	0.925	0.928	0.917	0.942	0.947
Mode	0.836	0.787	0.862	0.842	0.856	0.849
Day	0.944	0.925	0.937	0.919	0.950	0.946
Primary activity	0.830	0.766	0.801	0.800	0.861	0.840

  - Aggregated number of trips reflects in correlation coefficients
  - Fit decreases with an increasing number of cells
  - One R, FS and C4.5 give a good performance

Application - Part I (6)

- Conclusions - so far. . .
  - Simple models do not perform better, but are also not inferior to more complex models
  - Strong reduction in size of the trees can be obtained by first applying FS
  - Findings endorse primary belief that people do not rely on a complex series of rules to make a decision
- What is next?

Combination of Simple Models

- 2 powerful learning ideas introduced in the last decade
- Bagging
  - Re-sample training set 50 times (sampling with replacement)
  - Average the result of the classifier over the different data sets
  - Reduces the variance of the prediction & improves the stability
- Boosting (AdaBoost)
  - Build model on the data
  - Increase influence of misclassified instances by giving them a weight
  - Combine results of 10 different classifiers through a weighted majority vote (more accurate classifiers have a higher weight)

(Breiman, 1996; Freund and Schapire, 1997)

Combination of Simple Models (2)

- General: Feature Selection approach provided the best results of all the simple models
- Wickramaratna *et al.* (2001): use Bagging & Boosting only with weak classifiers
- $\Rightarrow$ Application of Bagging & Boosting to One R and FS models

Application - Part II

- Choice Facet Level

Dimension	1R	1R Bag.	1R Bo.	FS	FS Bag.	FS Bo.	Best
Mode for work	0.595	0.605	0.640	0.595	0.611	0.614	0.648
Selection	0.677	0.678	0.734	0.669	0.672	0.673	0.724
With-whom	0.408	0.430	0.408	0.467	0.566	0.564	0.509
Duration	0.348	0.418	0.348	0.368	0.452	0.451	0.431
Start time	0.227	0.318	0.227	0.172	0.704	0.703	0.408
Trip chain	0.699	0.699	0.807	0.811	0.896	0.894	0.833
Mode other	0.413	0.724	0.413	0.508	0.931	0.930	0.528
Location 1	0.435	0.435	0.501	0.513	0.564	0.561	0.575
Location 2	0.234	0.296	0.234	0.312	0.708	0.710	0.372

  - Clearly improvement in predictive performance per dimension!
  - How about aggregate behaviour?

Application - Part II (2)

- Activity Pattern Level

Measure	Mean Distance						
	1R	1R Bag.	1R Bo.	FS	FS Bag.	FS Bo.	Best
SAM (AT)	3.027	2.821	2.805	2.929	2.988	3.022	2.777
SAM (TP)	3.312	3.204	3.088	3.208	3.270	3.342	3.168
SAM (L)	3.184	2.889	2.844	3.033	3.158	3.096	3.033
SAM (TM)	4.592	4.347	4.189	4.600	4.588	4.540	4.497
UDSAM	17.142	16.081	15.732	16.699	16.990	17.022	16.475
MDSAM	8.474	7.944	7.782	8.373	8.471	8.356	8.333

  - One R Bagging and Boosting give overall the best performance at activity pattern level

Application - Part II (3)

- Trip Matrix Level

Measure	ρ(o,p)						
	1R	1R Bag.	1R Bo.	FS	FS Bag.	FS Bo.	Best
None	0.928	0.925	0.930	0.947	0.951	0.950	0.947
Mode	0.862	0.877	0.883	0.849	0.863	0.861	0.862
Day	0.937	0.927	0.929	0.946	0.948	0.950	0.950
Primary activity	0.801	0.807	0.803	0.840	0.818	0.798	0.861

  - Best performance at aggregate level $\neq$ best performance at dimensions separately

Application - Part II (4)

- Conclusions
  - Bagging and Boosting improve the performance of the Simple Models
  - (Combinations of) Simple Models do not necessarily perform worse than more complex models
  - They are able to capture the most important information in trying to predict activity-travel behaviour

Mode Choice Models

- Important issue: Modelling of Choice of Transport Mode
- Changes over the last decennium
  - Improvement in survey techniques $\Rightarrow$ Larger data sets
  - Emphasis on activities $\Rightarrow$ transportation characteristics + demographical variables + activity features
  - Variety of new techniques: semi- and nonlinear models
- Need for a test statistic that can assess the model fit of a logistic regression model in a high dimensional sample space
- How do (some of) these new techniques perform in this context?

Lack-of-Fit Test

- Selection of particular transport mode above others ⇒ Multiple Logistic Regression Setting

- Assessment of adequacy of model
  - Pearson test statistic (1900): all variables need to be categorical
  - Hosmer and Lemeshow (1980): continuous and categorical predictors
  - le Cessie and Van Houwelingen (1991): test based on smoothing
  - Aerts *et al.* (1999): statistic based on orthogonal series approximation
  - . . .

Lack-of-Fit Test (2)

- Most of previous methods are faced with curse of dimensionality and have practical difficulties with implementation

- Hosmer and Lemeshow
  - Pearson-like statistic $\sim \chi^2_8$ (simulations)
  - Form 10 equally sized groups (deciles of risk)
  - Grouping is based on the fitted probabilities under the null model
  - Deal with dimensionality problem, BUT at cost of power

- Tree-Based Lack-of-Fit Test

General Setting

- Null hypothesis:

$$H_0 : \text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = \beta\mathbf{X}$$

with $\pi = P(Y = 1)$ and $\mathbf{X}$ an $n \times (p+1)$ matrix consisting of the n measurements on the p variables

- Alternative hypothesis H_1: not a specific alternative model

- Omnibus test for H_0 against a wide range of alternative models

Tree-Based Lack-of-Fit Test

- Test statistic:

$$T = \sum_{i=1}^{g} \frac{(y_i - n_i\hat{\pi}_i)^2}{n_i\hat{\pi}_i(1-\hat{\pi}_i)}$$

  with $\hat{\pi}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \hat{\pi}_j^0$ the average of the probabilities $\hat{\pi}_j^0$ for all covariate values $(x_{j1}, \ldots, x_{jp})$ in group i, fitted under the null model

- Grouping is according to recursive partitioning algorithm underlying CART, which can be considered as nonparametric alternative model

- Number of groups will affect power characteristics

(Breiman *et al.*, 1984)

Tree-Based Lack-of-Fit Test (2)

- Unbalanced groups
- Two variants
  - Weighted version

$$T_W = \sum_{i=1}^{g} w_i \frac{(y_i - n_i\hat{\pi}_i)^2}{n_i\hat{\pi}_i(1-\hat{\pi}_i)}$$

with weight $w_i = gn_i/N$, giving less weight to small groups

  - Cressie-Read version

$$T_{CR} = \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^{g} \left\{ y_i \left(\left(\frac{y_i}{n_i\hat{\pi}_i} \right)^{\lambda} - 1 \right) + (n_i - y_i) \left(\left(\frac{n_i - y_i}{n_i(1-\hat{\pi}_i)} \right)^{\lambda} - 1 \right) \right\}$$

with $-\infty < \lambda < \infty$

Tree-Based Lack-of-Fit Test (3)

- Partitioning of sample space is random $\Rightarrow$ Hard to determine null distribution theoretically
- As in Hosmer and Lemeshow's approach: simulations learn that $\chi^2_{2\times g-p}$ is a good approximation
- Null distribution can always be simulated by a parametric bootstrap method
- Tree may reveal a particular deviation from the null model

Simulation Study

Null model:

$$H_0 : \text{logit}(\pi(x_i, z_i)) = \beta_0 + \beta_1 x_i + \beta_2 z_i$$

with $Y_i \sim$ Ber$(\pi(x_i, z_i)), i = 1, \ldots, 100, x_i, z_i \sim U(-6, 6)$,
$\beta_0 = 0.0, \beta_1 = 0.8$ and $\beta_2 = 0.3$

Null model			
Test	0.10%	0.05%	0.01%
HL	12.46	15.97	28.85
$\chi^2(8)$	13.36	15.51	20.09
T	20.86	25.13	32.45
T_W	16.40	18.93	23.68
T_{CR}	19.76	22.97	27.60
$\chi^2(12)$	18.55	21.03	26.22

Simulation Study (2)

First alternative model:

$$H_1 : \text{logit}(\pi(x_i, z_i)) = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 x_i z_i$$

with $\beta_3 = 0.1$ and 0.2

Interaction Model				
	$\beta_3 = 0.10$		$\beta_3 = 0.20$	
	Test(0.10)	Test(0.05)	Test(0.10)	Test(0.05)
HL	27.14	14.29	68.47	51.20
T	36.33	24.90	93.17	84.54
T_W	45.71	32.04	97.19	94.58
T_{CR}	40.20	26.73	96.18	91.16
Oracle Test	77.02	73.08	99.59	99.59

Simulation Study (3)

Second alternative model:

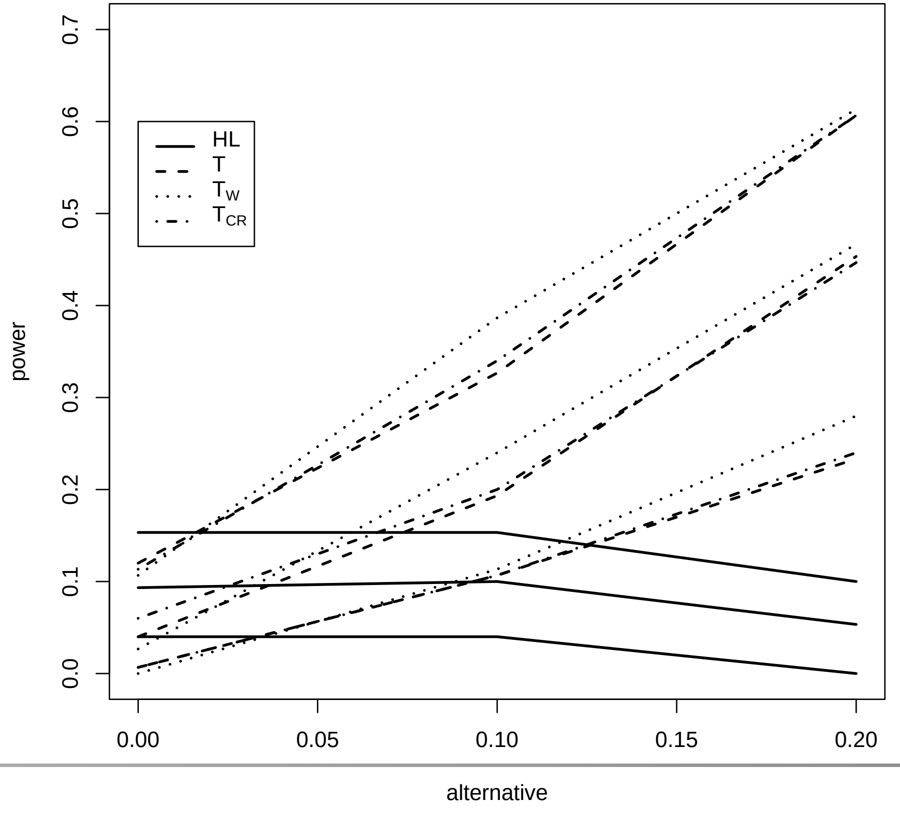
$$H_1 : \text{logit}(\pi(x_i, z_i)) = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 x_i^2$$

with $\beta_3 = 0.1$ and 0.2

Quadratic Model				
	$\beta_3 = 0.10$		$\beta_3 = 0.20$	
	Test(0.10)	Test(0.05)	Test(0.10)	Test(0.05)
HL	13.83	5.01	17.60	6.80
T	29.66	15.43	89.2	78.4
T_W	47.70	32.06	95.40	91.40
T_{CR}	36.47	23.85	94.40	86.60
Oracle Test	86.97	84.52	99.59	99.59

Other Examples

- GVHD and POPS data: Results comparable to those obtained in the literature
- High-dimensional simulation settings: very low power for HL in detecting interaction ↔ reasonable results for tree-based statistic



Other Examples (2)

- Dutch Car Driver Data: 1025 cases and 39 variables
- Goodness-of-Fit Test as Model Selection Tool

Variables used in model	HL		T_W		AIC
	Stat.	P-value	Stat.	P-value	
Model 1: 25	6.09	0.648	64.71	0	926.98
Model 2: 26	4.47	0.800	64.77	0.002	928.91
Model 3: 28	14.33	0.058	64.89	0.006	925.42
Model 4: 30	5.31	0.670	45.02	0.028	910.23
Model 5: 32	5.63	0.616	45.14	0.034	912.38
Model 6: 34	7.35	0.434	39.63	0.044	910.41
Model 7: 36	8.46	0.358	38.62	0.048	902.98
Model 8: 38	8.62	0.328	35.64	0.066	896.94
Model 9: 39	9.55	0.274	35.14	0.074	898.56

Conclusion

- Continuous and categorical explanatory variables
- Non-parametric method can be used to confirm or improve a parametric null model
- Simulation results: very promising power characteristics in detecting incorrectly modelled variables, omitted interaction or quadratic effects, . . .
- Further research on approximate null distribution

Nonlinear Models

- Field of Transportation Research: dominated by linear models
- The Right Choice to Make?
  - Semi-linear: Fractional polynomials (mfp)
  - Nonlinear
    - Support Vector Machines (SVM)
    - Classification and Regression Trees (CART)

Fractional Polynomial Approach

Fractional polynomial of degree m

$$\zeta_0 + \sum_{i=1}^{m} \zeta_i H_i(x)$$

for $(p_1, \ldots, p_m)$ vector of positive or negative integers or fractions

$$H_i(x) = \begin{cases} x^{(p_i)} & \text{if } p_i \neq p_{i-1} \\ H_{i-1}(x)\ln(x) & \text{if } p_i = p_{i-1} \end{cases}$$

Set $H_0(x) = 1$, $p_0 = 0$ and

Box-Tidwell: $$x^{(p_i)} = \begin{cases} x^{p_i} & \text{if } p_i \neq 0 \\ \ln(x) & \text{else} \end{cases}$$

(Royston and Altman, 1994)

Fractional Polynomial Approach (2)

Recommendations

- Restrict powers to $p_i \in \{-2, -1, -\frac{1}{2}, 0, \frac{1}{2}, 1, 2, \max(3, m)\}$
- $m > 2$ rarely required in practice

Examples

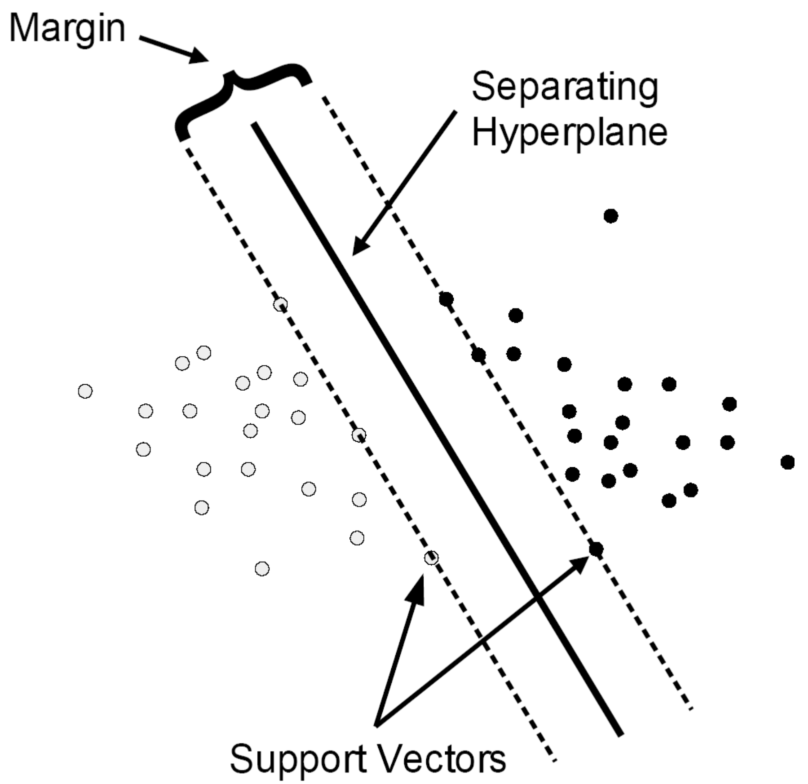
- $(\frac{1}{2}, 1) : \zeta_0 + \zeta_1 \sqrt{x} + \zeta_2 x$
- $(1, 1) : \zeta_0 + \zeta_1 x + \zeta_2 x \ln x$

Setting

- Multiple Logistic Regression
- Iterative Procedure: Stepwise Regression

Support Vector Machines

- Optimal separating hyperplane
- Example



(Vapnik, 1996)

Support Vector Machines (2)

- Non-separable Case
  - Maximise the margin
  - Allow for some points to be at the wrong side of the margin

- Nonlinear boundary: Transform original space and find linear boundaries in transformed space

- Kernel functions $K(x, x') = \langle h(x), h(x') \rangle$
  - d-th degree polynomial: $K(x, x') = (\kappa + \gamma \langle x, x' \rangle)^d$
  - Radial basis: $K(x, x') = \exp(-\frac{\|x - x'\|^2}{\gamma})$
  - Neural network: $K(x, x') = \tanh(\kappa_1 \langle x, x' \rangle + \kappa_2)$

Data and Model Comparison

- Southeast Florida: 14527 cases and 15 variables
- Albatross: 1025 cases and 20 variables

	Slow	Public	Car driver	Other
Dutch data	18.93%	12.29%	68.78%	0.00%
Southeast Florida	2.71%	9.67%	87.13%	0.49%

- Slow and Public Transport will be hard to predict!
- Take Sensitivity $= P(y_{predicted} = 1 | y_{observed} = 1)$ as measure
- Accuracy will always be high, even if no ‘true’ case is predicted correctly!

Example: Dutch Data: Car Driver

$$\begin{aligned}\text{logit}(\pi(x)) &= \beta_0 + \beta_1(x_6 = 2) + \beta_2 x_7 + \beta_3(x_{14} = 1) + \beta_4(x_{24} = 1) \\ &+ \beta_5(x_{26} = 2) + \beta_6(x_{26} = 8) + \beta_7 x_{30}^{-1} + \beta_8 x_{30}^{3} + \beta_9 x_{31} \\ &+ \beta_{10} x_{33}^{-1} + \beta_{11} x_{33}^{-\frac{1}{2}} + \beta_{12}(x_{38} = 1) + \beta_{13}(x_{39} = 1)\end{aligned}$$

- Mfp model: 2 fractional polynomials used; lowest AIC and BIC values among 3 ‘linear’ models
- Interpretation: Number of cars, at least 1 shopping, service, social, leisure, bring/get activity, travel time by bike, woman
- Classification tree: same variables
- Comparison of models occurs on the test set (70-30% split)

Examples: Summary

- Results of all the analyses on the test set

	D-CD	D-PT	D-ST	SEF-PT	SEF-ST
Linear	0.730	0.824	0.882	0.649	0.627
Mfp	0.811	0.029	0.882	0.000	0.659
Interaction	0.788	-	-	-	-
SVM - Linear	0.928	0.000	0.549	0.000	0.000
SVM - Polynomial	0.793	0.294	0.588	-	-
SVM - Radial basis	0.946	0.000	0.510	0.052	0.008
SVM - Neural net	0.901	0.000	0.569	0.000	0.000
CART	0.864	0.118	0.549	0.097	0.056

Conclusions

- (Semi-)linear: Best interpretable
- SVM and CART: over-fit training data $\Rightarrow$ lower performance on test data
- Very skew data sets: better performance of (semi-)linear models
- More balanced data sets: nonlinear models

Thanks!

Everything should be made as simple as possible,
but not simpler - Albert Einstein