Constrained Virtual Tailoring from Anthropometric Data, 3-D Shape and Data Mining

Eric Paquet[1,2], Herna L Viktor[2], Hongyu Guo[2] & Isis Peña Sanchez[2]
[1]National Research Council, Canada
[2]School of IT and Engineering, University of Ottawa, Canada
Email: eric.paquet@nrc-cnrc.gc.ca

Problem statement

- Clothes should tailor well, fit the body well and hide obvious body flaws
- What is the interrelationship between different body measurements, such as shoulder width, waist and neck circumference?
- Which measurements are important for a particular clothing size?
- Analyze an anthropometric data set to better understand the typical consumer
- Create rules <u>to be used as constraints</u> when designing and manufacturing clothes
- Verify finding against a set of 3-D body scans
- Of importance from Mass Market Manufacturer to Haute Couture

“Ideal” Measurements for Adult Males

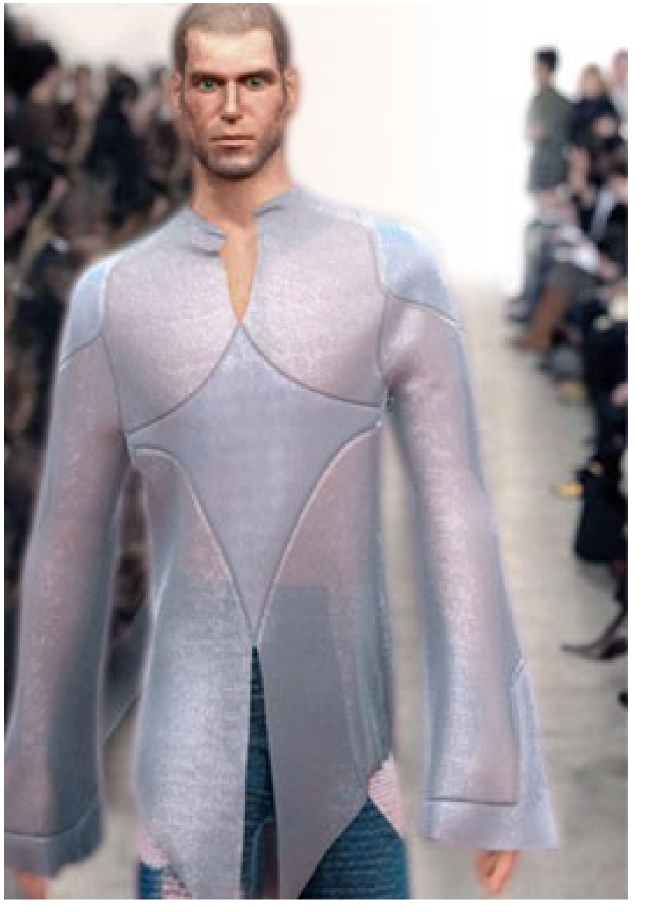
	Small		Medium		Large		X-Large		XX-Large	
	34	36	38	40	42	44	46	48	50	52
Chest	87	92	97	102	107	112	117	122	127	132
Waist	71	76	81	87	92	97	107	112	117	122
Hip	89	94	99	104	109	114	119	124	130	135
Neck	35.5	37	38	39.5	40.5	42	43	44.5	46	47
Sleeve	81	81	84	84	87	87	89	89	91	91
Stature	178	178	178	178	178	178	178	1.78	1.78	1.78

Typical Measurements for Adult Males

Chest	Waist	Hip	Neck	Stature
112.4	92.8	109.7	52.1	184.4
112.6	97.8	105	42.4	179.5
112.6	97.2	111.4	52	181.3
112.8	95.2	114	49.9	185.4
112.9	93.5	107.8	50.3	184.7
113	105.7	114.1	49.3	184.3
113.3	101.8	111.2	50.5	188.6
113.4	102.6	116.5	47	187.2
113.5	100.8	105.2	48.5	179.6
112	**97**	**114**	**42**	**178**

Outline of Talk

- CAESAR^{TM} case study
- Data acquisition
- Experimental approach
- Experimental results
- Conclusions

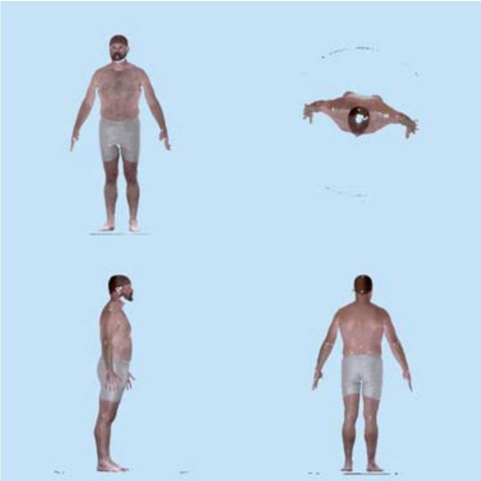


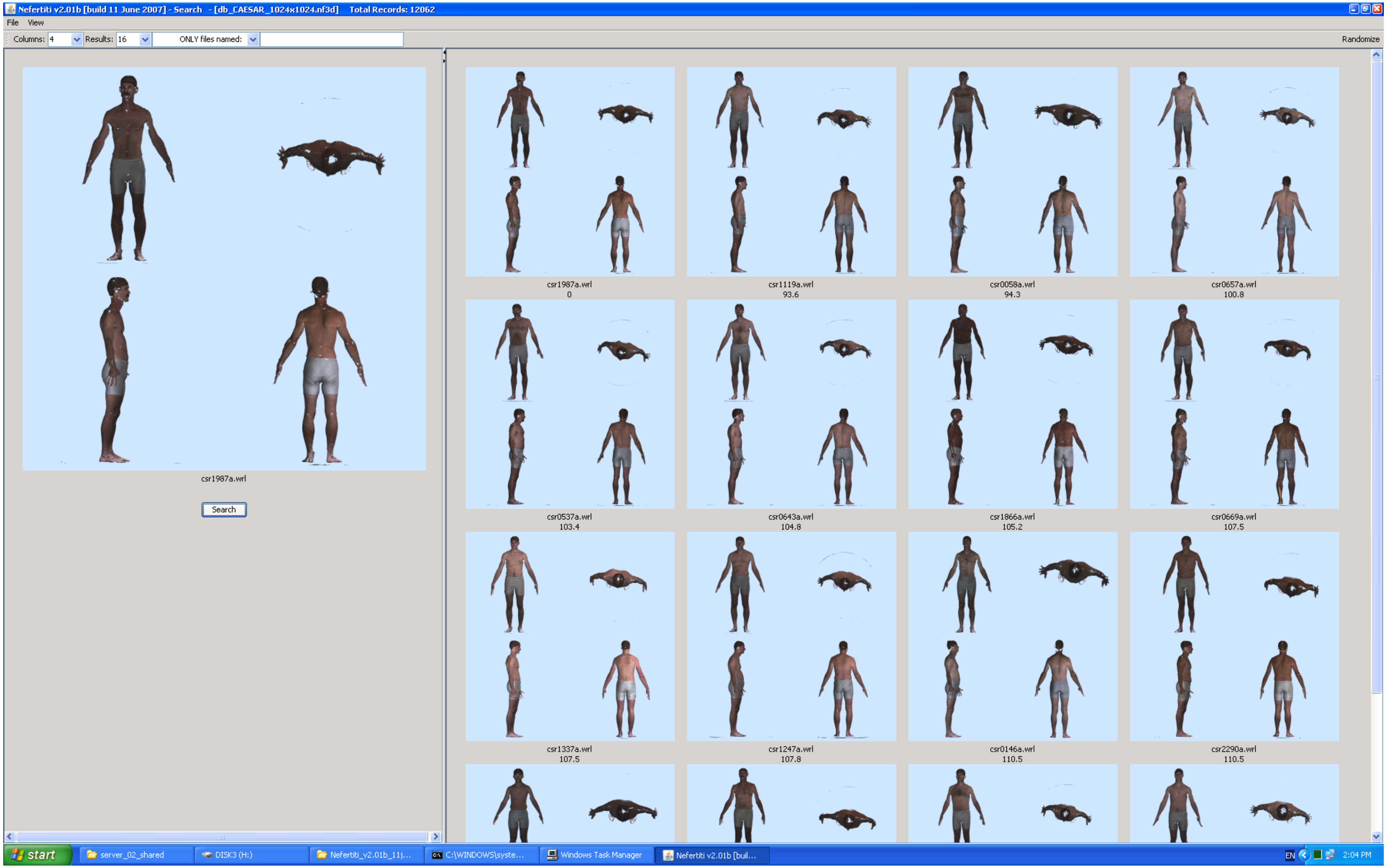
Case Study: CAESAR™ anthropometric database

- Anthropometric study in US, Netherlands, Canada and Italy
- Human subjects scanned using 3-D laser scanner technology
- Demographic and anthropometric details recorded
- Aim is to provide better fitting commercial products such as cars and clothing
- Contain 3-D and relational data

3-D Body Scan Indexing and Retrieval

- Content-based description
  - Shape-based descriptor based on the radial and angular distribution of triangles
  - Each person scanned in 3 positions
- Compact and abstract statistical description of the shape
- Can index the whole body or a part
- 120 bytes irrespectively of the tessellation of the body
- Automatic indexing
- Implemented in Cleopatra system





Nefertiti v2.01b [build 11 June 2007] - Search - [db_CAESAR_1024x1024.nf3d] Total Records: 12062
File View
Columns: 4
Results: 16
ONLY files named:
Randomize
csr1987a.wrl
Search
csr1987a.wrl
0
csr1119a.wrl
93.6
csr0058a.wrl
94.3
csr0657a.wrl
100.8
csr0537a.wrl
103.4
csr0643a.wrl
104.8
csr1866a.wrl
105.2
csr0669a.wrl
107.5
csr1337a.wrl
107.5
csr1247a.wrl
107.8
csr0146a.wrl
110.5
csr2290a.wrl
110.5
start
server_02_shared
DISK3 (H:)
Windows Task Manager
2:04 PM

Subset of Anthropometric Measurements

Measurement	Measurement
Acromial Height Sitting	Spine to Shoulder Length
Ankle Circumference	Spine to Elbow Length
Arm Length: Spine to Wrist	Arm Length: Shoulder to Wrist
Arm Length: Shoulder to Elbow	Arm Circumference
Bust Chest Circumference	Buttock Knee Length
Crotch Height	Eye Height Sitting
Face Length	Foot Length
Hand Length	Shoulder Breadth
Sitting Height	Vertical Trunk Circumference
Triceps Skinfold	Head Circumference
Knee Height Sitting	Thumb Tip Reach
Head Breadth	Hand Circumference

Subset of demographic measurements

Measurement
Family income
Age range
Fitness
Car brand
Car model
Car year
Education
Number of children
Marital Status
Occupation
Ethnic group

Also recorded details such as Perceived height and Perceived weight

CAESAR^{TM} Data acquisition and quality

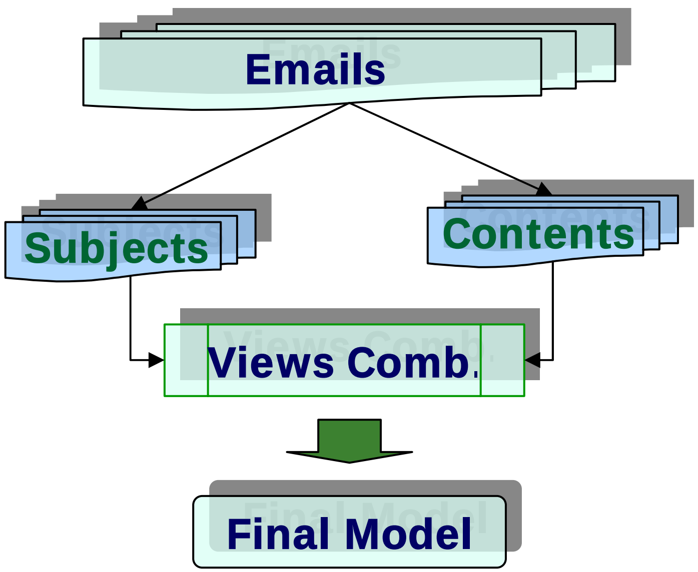
- Anthropometric data precision:
  - Precision of instrumentation is up to 1mm
  - Acquired (double-checked) by two anthropometric experts
- Anthropometric data accuracy:
  - Reference regions ill-defined for larger individuals due to lack of definition in body shape (e.g. crotch height)
- Demographic data:
  - Depends on the truthfulness and objectivity of the participants
  - E.g. 2.4% of perceived height and 6.2% of perceived weight of USA subjects was wrong

Experimental approach

- Data in IBM DB2 database
- Aim to <u>directly</u> mine this data, without extensive pre-processing
- Step 1: Cluster analysis based on anthropometric measures
- Step 2: Multi-view learning

Multi-view Learning: The General Idea

- Features are divided into disjoint subsets
- Each subset is used to form a separate view
- **An Example: Classifying Emails**

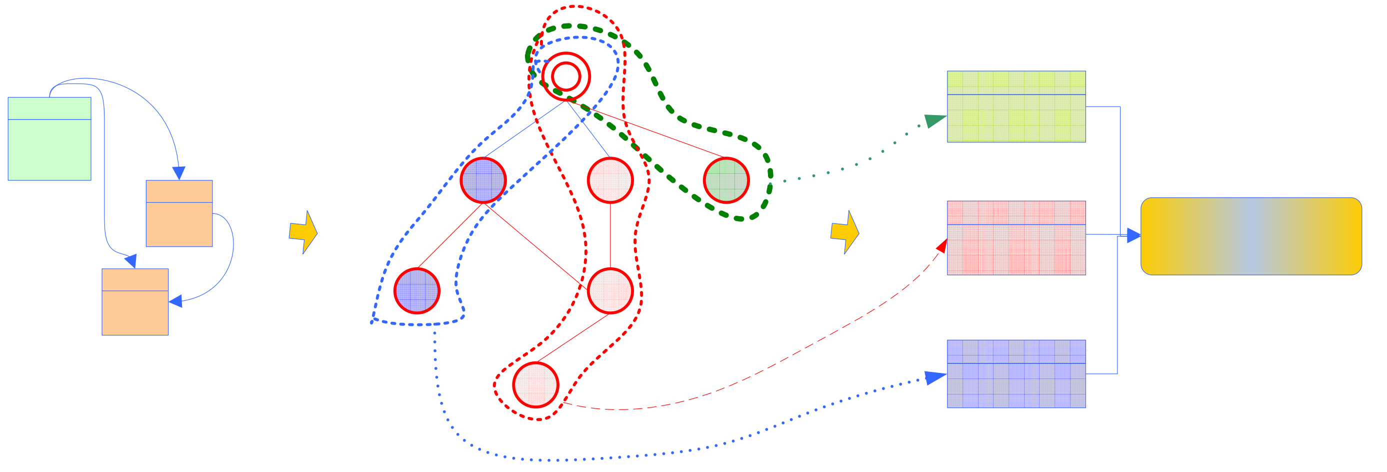


- Two views offer naturally divided subsets
- Combine for more “knowledge” about emails

Multi-view relational data mining:
Why not just apply a "traditional" method?

- Avoid "flattening" of data into single table
  - Computationally expensive
  - Does not scale well
  - May fail to converge
  - Many "null" values
  - Human error due to complex conversions; possible loss of semantic information
  - Difficult for domain experts, data mining novices

Multi-view Learning in Relational Databases



- Decompose the relational database into substructures
  - each maps to a subset of relations and joins (slot chain or joint path)
- Employ multi-view learners to form a set of strongly uncorrelated views
- A meta-learner is used to combine the multi-view learners

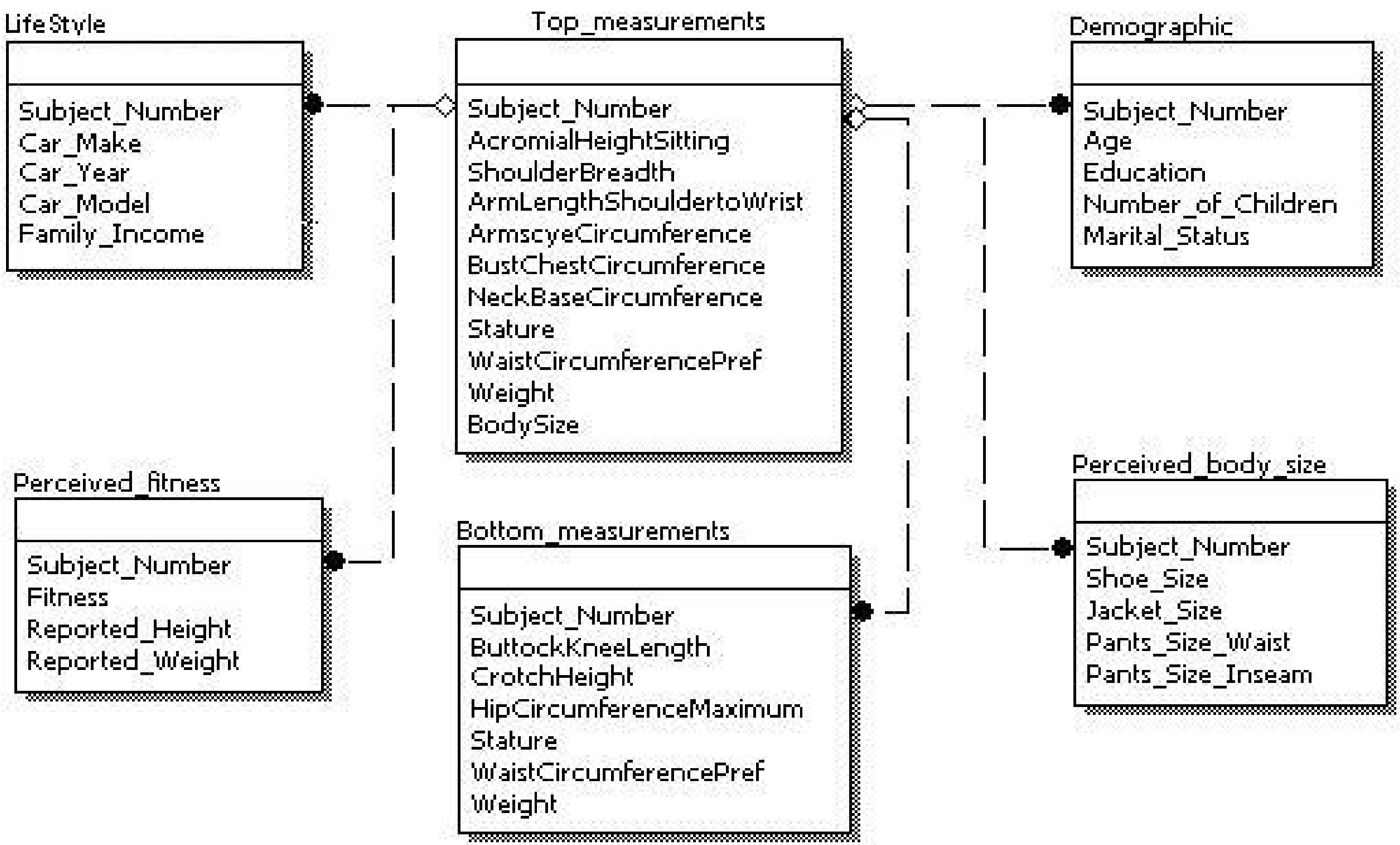
Multi-view learning: The Algorithm

Input: Relational database $\Re = R_t \times R_{b_1} \times \cdots \times R_{b_n}$,
Multi-view learner $\mathcal{L}$, Meta-learner $\mathcal{M}$,
Maximum length of join path $MaxJ$.

Output: Classification model $\mathcal{F}$.

1: Convert database schema $\Re$ into graph;
2: Extract join path set $\{\bowtie_n\}$ from graph;
3: Construct relational feature set $\varphi(t)$ for each join path in $\{\bowtie_n\}$; forming candidate view set $\{V_d^1, \cdots, V_d^n\}$;
4: Select a set $\{\mathcal{V}^i\}_1^{n'}$ from $\{V_d^1, \cdots, V_d^n\}$;
5: Train $\mathcal{L}$ with $\{\mathcal{V}^i\}_1^{n'}$, forming hypothesis set $\{\mathcal{H}^i\}_1^{n'}$;
6: Form final model $\mathcal{F}$ by combining $\{\mathcal{H}^i\}_1^{n'}$, using $\mathcal{M}$;
7: **return** $\mathcal{F}$.

ER diagram of CAESAR™ database



Experimental results: Analyzing the Male Population

- Integral data records of 418 male subjects (US citizens)
- Implemented using
  - IBM DB2 relational database
  - Cleopatra (3-D Body Scans for verification)
  - WEKA data mining system
    - Clustering: k-means (and others)
    - Classification: C4.5, RIPPER and PART
    - 10-fold cross-validation

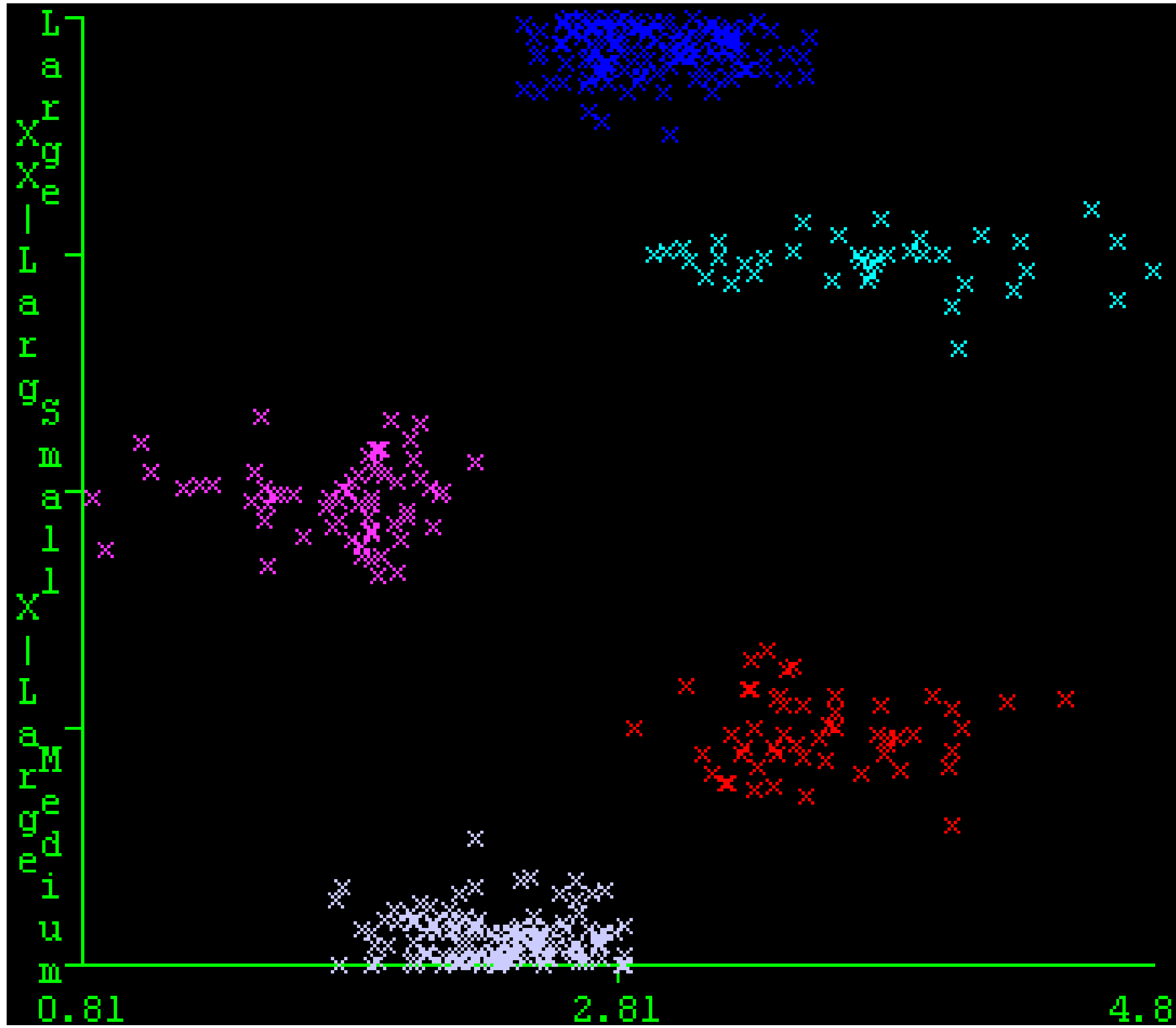
Cluster Analysis: Grouping into Clothing Sizes

- A number of clustering algorithms was applied to the data:
  - the EM algorithm, the CobWeb method, the k-means technique, the Farthest First approach, and an algorithm using a variation of the density-based clustering algorithm with k-means components
- K-means came out tops:
  - Similar sized clusters with convex shapes, numeric attributes.
  - Little noise or outliers.

Exploratory Cluster Analysis results

	Small	Medium	Large	X-Large	XX-Large
CobWeb	FTC	FTC	FTC	FTC	FTC
Farthest First	13 (3%)	252 (61%)	92 (22%)	55 (13%)	2 (0%)
EM	113 (27%)	122 (29%)	81 (20%)	70 (17%)	28 (7%)
k-means	62 (15%)	131 (31%)	131 (31%)	52 (13%)	38 (9%)
Density-based	85 (21%)	138 (33%)	78 (19%)	94 (23%)	19 (5%)

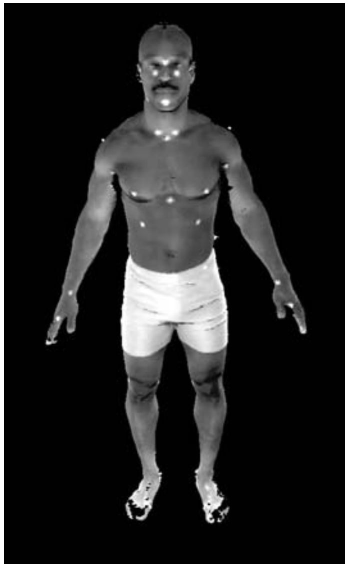
The Clusters..

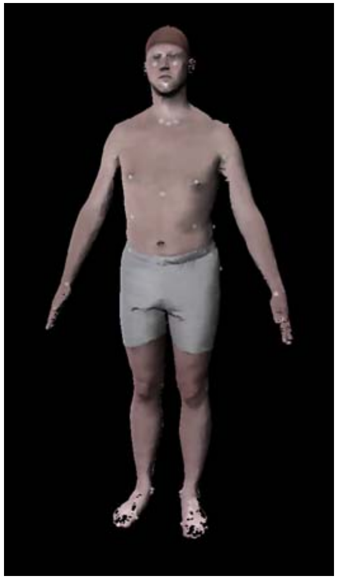


US Male Centroids Measurements

	Small	Medium	Large	X-Large	XX-Large
Chest	94.9 (5.9)	99.4 (6.8)	106.1 (6.19)	106.9 (6.6)	125.7 (10.9)
Waist	82.0 (6.8)	86.8 (6.8)	92.3 (7.5)	107.8 (7.0)	116.6 (16.0)
Hip	96.70 (4.80)	101.14 (4.93)	106.48 (5.42)	108.39 (4.90)	123.15 (14.33)
Neck	44.6 (1.9)	46.0 (2.0)	48.0 (2.19)	48.7 (2.0)	52.7 (3.04)
Arm Length	59.5 (2.03)	62.4 (1.79)	65.08 (1.71)	68.7 (1.96)	65.56 (3.02)
Weight (lbs)	151.17 (15.77)	172.19 (17.70)	198.16 (18.61)	212.85 (20.04)	274.99 (35.9)
Stature	167.0 (5.9)	174.7 (3.5)	180.6 (3.6)	190.0 (5.0)	181.9 (4.6)
Sitting Height	56.82 (2.47)	59.47 (2.45)	61.35 (2.26)	63.78 (2.84)	64.65 (2.41)
Hip Breadth Sitting	35.67 (3.13)	37.22 (2.09)	39.04 (2.3)	40.20 (2.23)	45.10 (4.55)
Number of Members	62 (15%)	131 (31%)	131 (31%)	52 (13%)	38 (9%)

The Cluster Centroids of the Male population:





Multi-view learning: Accuracy and number of rules

Learning task	Ripper		C4.5		PART	
Top measurements	78.7%	10	78.2%	30	76.6%	18
Bottom measurements	76.3%	11	78.2%	19	77.5%	19
Top and Bottom measurements	79.2%	12	80.1%	30	78.7%	21

Some high coverage rules

Rule	Size	Cover	%	Learner
(Stature <= 170.9) AND (BustChestCircumference <= 100.2)	**Small**	**(44/ 3)**	**71.0**	**RIPPER**
(167 < Stature <= 176) AND (Weight> 163) AND (HipCircumference <= 113.5) AND (ButtockKneeLength <= 63.5)	**Medium**	**(78/6)**	**59.5**	**C4.5**
(183 <Weight <=243) AND (62< ArmShouldertoWrist <=66.9) AND (CrotchHeight <= 87) AND (Stature > 176)	**Large**	**(94/8)**	**71.8**	**C4.5**
(ButtockKneeLength > 61.4) AND (BustChestCircumference <= 113)	**X Large**	**(31/5)**	**59.7**	**PART**
(Weight >= 239) and (BustChestCircumference >= 118.4)	**XX Large**	**(28/0)**	**73.7**	**RIPPER**

Interpretation of Results from a Tailoring Perspective

- Population grouped into clusters with well-defined Centroids or Archetypes
- Clusters correspond to reality, as verified against 3-D scans
- Describe interrelationship between measurements
- Identified which measurements are important
- Examples:
  - Medium: the Stature, Hip Circumference, Buttock to Knee Length and Crotch Height are important when designing pants
  - Large versus Xlarge: the chest and hip circumferences do not differ substantially, the waist and stature are of importance here
  - Small (or Thin) individuals generally have short legs

Next:
Demographic Profiling of Males

- Learning from the *Demographics, LifeStyle, Perceived_fitness* and *Perceived_bodysize* relations' perspectives
- Final three rule sets
  - 84 (PART – 73.5%)
  - 7 (RIPPER – 68.1%)
  - 35 (C4.5 – 73.6%)
- Many rules with low coverage; indicate diverse background
- Measured "interestingness"
  - percentage of population within a group that satisfies the rules

Some Interesting Rules

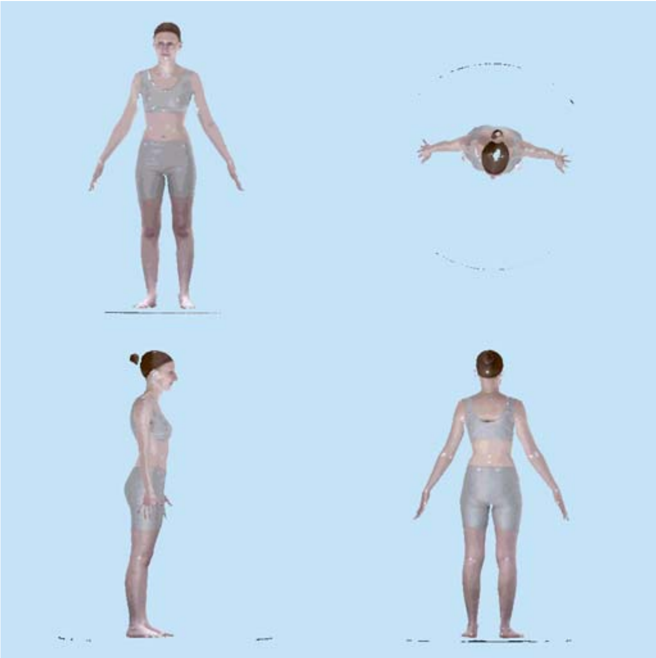
Age range	Fitness	Education	Children	Marital Status	Income (US$)	Clothing Size	Coverage (%)
25-33			3+	Married		XX-Large	86.7%
		Bachelors	0-3	Single or Married	60,000-99,999	X-Large	62.7%
		Bachelors		Single or Married	Over 100,000	Large	62.2%
25-40	Medium	Masters		Married	60,000-79,999	Medium	83.3%
40-50	High	Masters		Single		Small	76.4%

The Medium Sized Males

Age range	Fitness	Education	Number of Children	Marital Status	Family Income
25-40		High School	> 0	Single or Married	45,000-60,000
25-40	Low	Bachelors		Single	45,000-60,000
25-40	High	Bachelors		Single	100,000+
25-40	Medium	Masters	<= 2	Married	81,000-100,000
25-40	Medium	Masters	> 2	Married	61,000-80,000
25-40	Medium	Doctorate		Single	61,000-80,000
25-40		Doctorate		Married	100,000+
40+	Low	Masters		Married	61,000-80,000
40+		Technical Training			100,000+
40+				Single	100,000+

Analysis of US Female Population

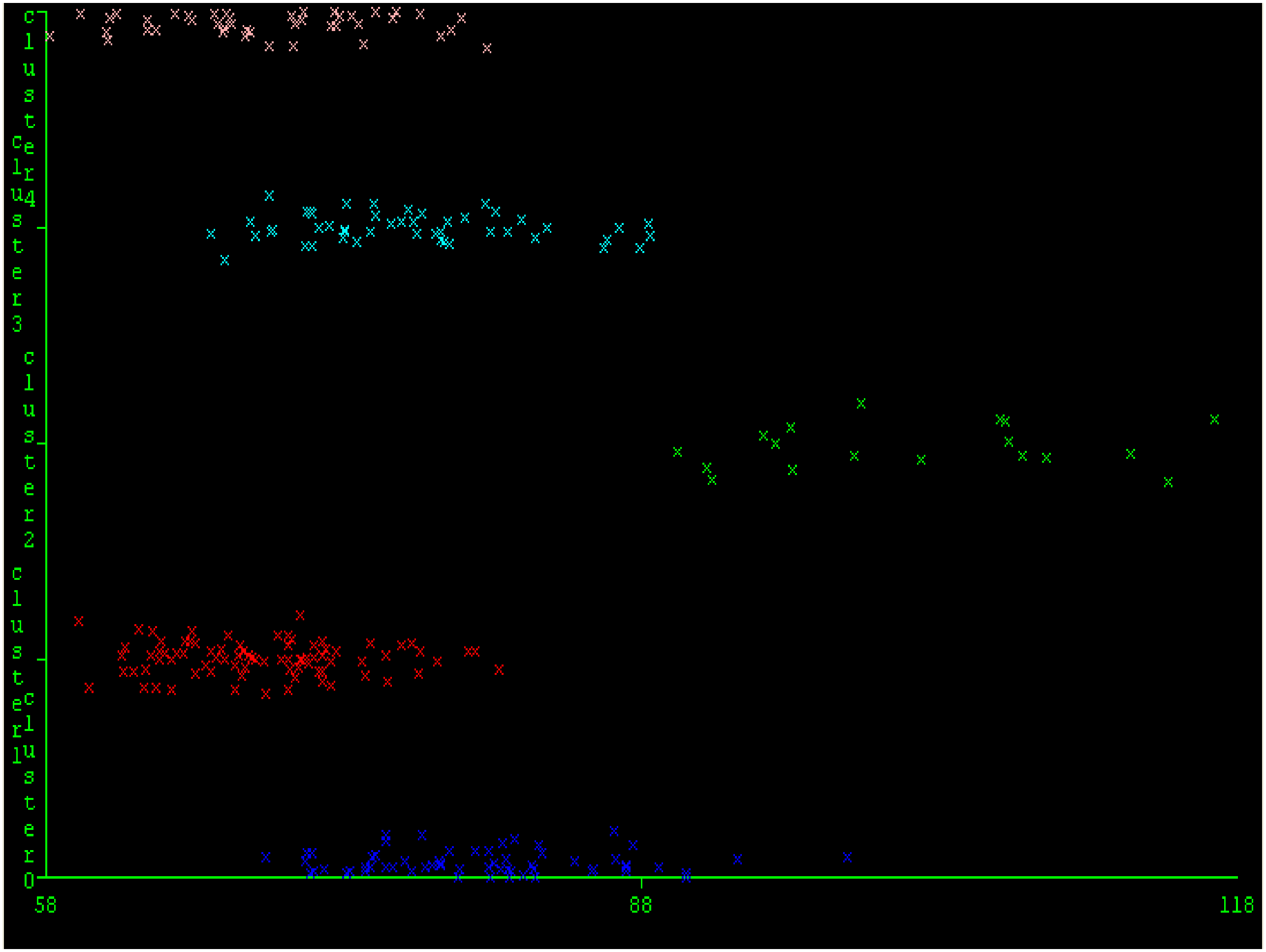
- Total of 256 subjects
- 49 measurements, including "under bust circumference"
- Followed same methodology as with US Male Subjects
  - WEKA system
    - Clustering via k-means
    - Classification RIPPER, Part C4.5
  - Verified using Cleopatra



Female Centroids Measurements

	Small	Medium	Large	X-Large	XX-Large
Chest	87.74 (5.84)	94.42 (4.5)	105.36 (6.9)	88.77 (4.64)	127.14 (9.64)
Waist	69.07 (6.14)	77.41 (6.27)	88.36 (9.1)	70.83 (4.99)	110.52 (5.59)
Hip	95.97 (5.04)	105.54 (5.27)	113.93 (7.69)	99.38 (5.11)	133.2 (9.81)
Arm Length	53.4 (1.98)	60.5 (2.29)	56.91 (1.68)	57.12 (1.68)	63.6 (2.7)
Weight (lbs)	118.98 (12.51)	154.72 (11.92)	178.4 (22.28)	130.44 (10.84)	278.9 (23.69)
Stature	155.87 (4.88)	171.94 (5.01)	162.04 (4)	163.69 (3.8)	180.58 (5.21)
Shoulder Breadth	41.22 (1.69)	44.45 (1.83)	46.79 (2.16)	42.41 (1.84)	55.44 (4.12)
Number of Members	53 (21%)	99 (39%)	60 (23%)	39 (15%)	5 (2%)

The Female Clusters..



Some high coverage rules

Rule	Size	Cover	%	Learner
(ArmLengthSpinetoWrist <= 74.5) and (VerticalTrunkCircumference <= 152.5)	**Small**	**(52/6)**	**98.1**	**RIPPER**
Weight <= 146 AND ArmLengthSpinetoWrist > 74.2 AND Stature <= 172.4 AND SpinetoElbow <= 52.4 AND ShoulderBreadth > 38.5 AND NeckBaseCircumference > 37.9	**Medium**	**(75/1)**	**75.8**	**C4.5**
KneeHeightSitting > 52.6 AND NeckBaseCircumference <= 44.3 AND ChestGirthatScye > 83.5	**Large**	**(49/1)**	**81.7**	**PART**
SubscapularSkinfold > 2.7 AND SpinetoElbow <= 54.5 AND ArmLengthSpinetoWrist <= 80.3 AND Weight > 145	**X Large**	**(30/0)**	**77.0**	**PART**

Note: Small number of XX-Large subjects lead to low coverage rules

Conclusions

- PKKK Innovative Application Award, Berlin, Germany, 2006
- Towards understanding the typical consumers' body profile
- Population grouped into 5 well-defined clusters
- Created rules to be used as constraints that should be satisfied when designing and manufacturing clothes
- Different measurements are important for different clothing sizes
- Results verified against 3-D body scans
- Future research:
  - The anthropometry of the disabled and elderly
  - Comparative study with other populations
- A thorough investigation into the general applicability and relevancy of demographic profiles
- Interestingness as a measure:
  - objective, semantically meaningful and/or subjective

PKDD INNOVATIVE APPLICATION AWARD

ECML PKDD 2006

17th European Conference on Machine Learning and
10th European Conference on Principles and Practice
of Knowledge Discovery in Databases

presented to

Herna Viktor, Eric Paquet, Hongyu Guo

for their paper

Measuring to Fit: Virtual Tailoring through Cluster Analysis and Classification

ECML PKDD 2006 Program Chairs

Johannes Fürnkranz | Tobias Scheffer | Myra Spiliopoulou

Berlin, Germany | September 18th - 22nd, 2006