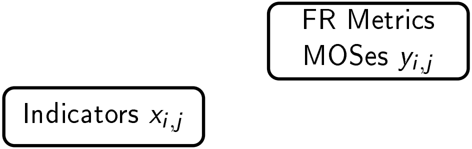
Data Analysis Proposition

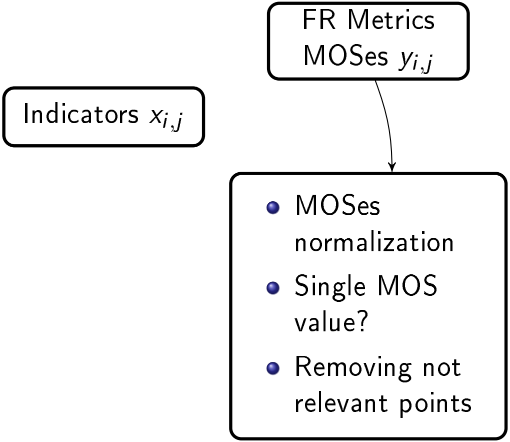
Lucjan Janowski, AGH

June 11, 2012

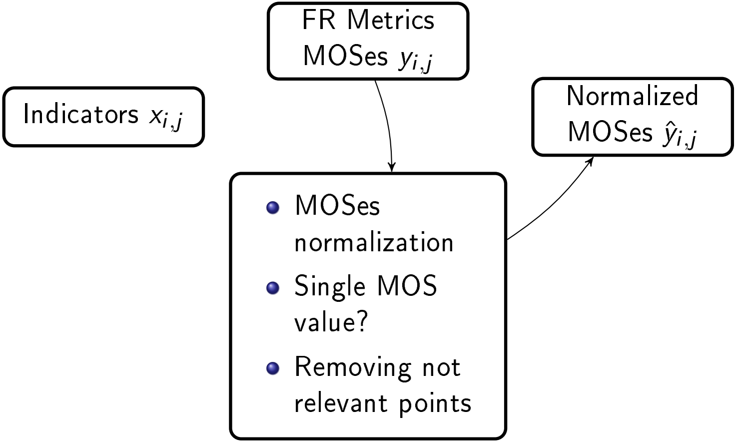
Data Flows



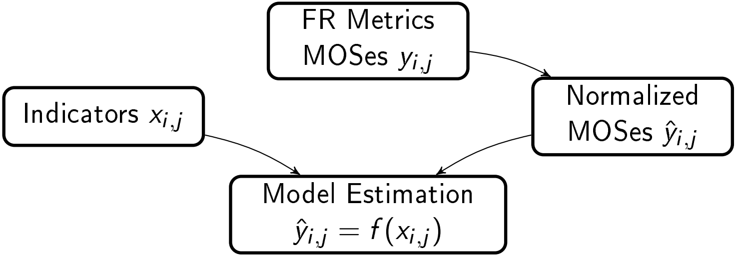
Data Flows


FR Metrics
MOSes $y_{i,j}$
Indicators $x_{i,j}$
MOSes normalization
Single MOS value?
Removing not relevant points

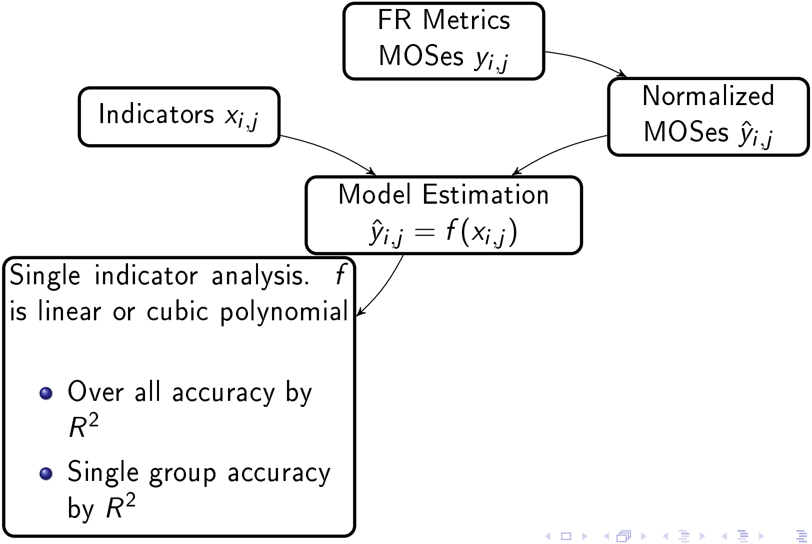
Data Flows


FR Metrics MOSes $y_{i,j}$
Indicators $x_{i,j}$
Normalized MOSes $\hat{y}_{i,j}$
MOSes normalization
Single MOS value?
Removing not relevant points

Data Flows


FR Metrics
MOSes $y_{i,j}$
Indicators $x_{i,j}$
Normalized
MOSes $\hat{y}_{i,j}$
Model Estimation
$\hat{y}_{i,j} = f(x_{i,j})$

Data Flows


FR Metrics MOSes $y_{i,j}$
Normalized MOSes $\hat{y}_{i,j}$
Indicators $x_{i,j}$
Model Estimation $\hat{y}_{i,j} = f(x_{i,j})$
Single indicator analysis. f is linear or cubic polynomial
Over all accuracy by R^2
Single group accuracy by R^2

Data Flows

FR Metrics MOSes $y_{i,j}$

Normalized MOSes $\hat{y}_{i,j}$

Indicators $x_{i,j}$

Model Estimation $\hat{y}_{i,j} = f(x_{i,j})$

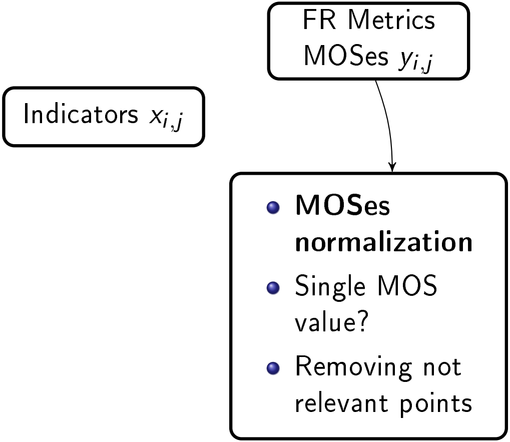
Single indicator analysis. f is linear or cubic polynomial

- Over all accuracy by R^2
- Single group accuracy by R^2

Multiple indicators analysis. f is linear or linear with interactions

- Indicator accuracy by p-value
- Overall model accuracy by adjust R^2

Data Flows


FR Metrics
MOSes $y_{i,j}$
Indicators $x_{i,j}$
MOSes normalization
Single MOS value?
Removing not relevant points

Normalization

- Why do we need it?

Normalization

- Why do we need it?
- If two metrics are correlated then one metric value can be computed by: $MOS_{i,1} = aMOS_{i,2} + b$

Normalization

- Why do we need it?
- If two metrics are correlated then one metric value can be computed by: $MOS_{i,1} = aMOS_{i,2} + b$
- If $|MOS_{i,1} - MOS_{i,2}| > \epsilon$ it means nothing

Normalization

- Why do we need it?
- If two metrics are correlated then one metric value can be computed by: $MOS_{i,1} = aMOS_{i,2} + b$
- If $|MOS_{i,1} - MOS_{i,2}| > \epsilon$ it means nothing
- if $|MOS_{i,1} - (aMOS_{i,2} + b)| > \epsilon$ it means that $MOS_{i,1}$ and $MOS_{i,2}$ are not well correlated at least for ith PVS

Normalization

- Why do we need it?
- If two metrics are correlated then one metric value can be computed by: $MOS_{i,1} = aMOS_{i,2} + b$
- If $|MOS_{i,1} - MOS_{i,2}| > \epsilon$ it means nothing
- if $|MOS_{i,1} - (aMOS_{i,2} + b)| > \epsilon$ it means that $MOS_{i,1}$ and $MOS_{i,2}$ are not well correlated at least for ith PVS
- If more than two metrics are used a base metric has to be chosen

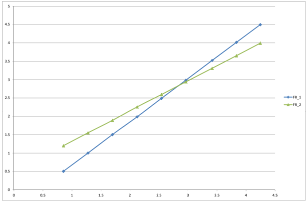
Normalization

- Why do we need it?
- If two metrics are correlated then one metric value can be computed by: $MOS_{i,1} = aMOS_{i,2} + b$
- If $|MOS_{i,1} - MOS_{i,2}| > \epsilon$ it means nothing
- if $|MOS_{i,1} - (aMOS_{i,2} + b)| > \epsilon$ it means that $MOS_{i,1}$ and $MOS_{i,2}$ are not well correlated at least for ith PVS
- If more than two metrics are used a base metric has to be chosen
- For HDTV it was set of subjective MOSes, for JEG-Hybrid it cannot be assumed

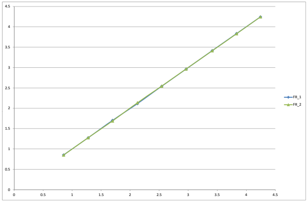
Normalization

- Why do we need it?
- If two metrics are correlated then one metric value can be computed by: $MOS_{i,1} = aMOS_{i,2} + b$
- If $|MOS_{i,1} - MOS_{i,2}| > \epsilon$ it means nothing
- if $|MOS_{i,1} - (aMOS_{i,2} + b)| > \epsilon$ it means that $MOS_{i,1}$ and $MOS_{i,2}$ are not well correlated at least for ith PVS
- If more than two metrics are used a base metric has to be chosen
- For HDTV it was set of subjective MOSes, for JEG-Hybrid it cannot be assumed
- Mean MOS of all metrics works the best probably

Example

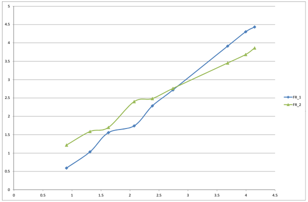


Input data

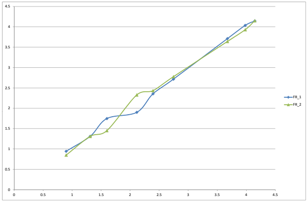


Normalized metrics

Example with Errors

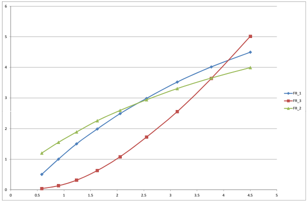


Input data

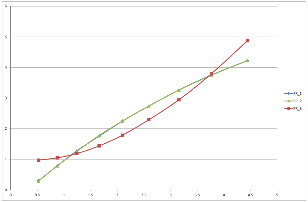


Normalized metrics

Example Non Linear

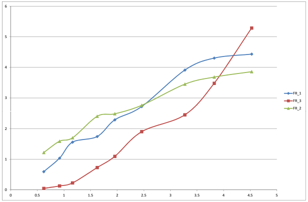


Input data

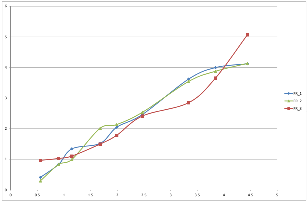


Normalized metrics

Example Non Linear with Errors

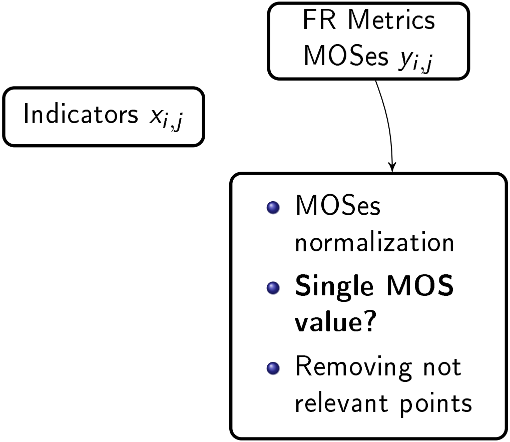


Input data



Normalized metrics

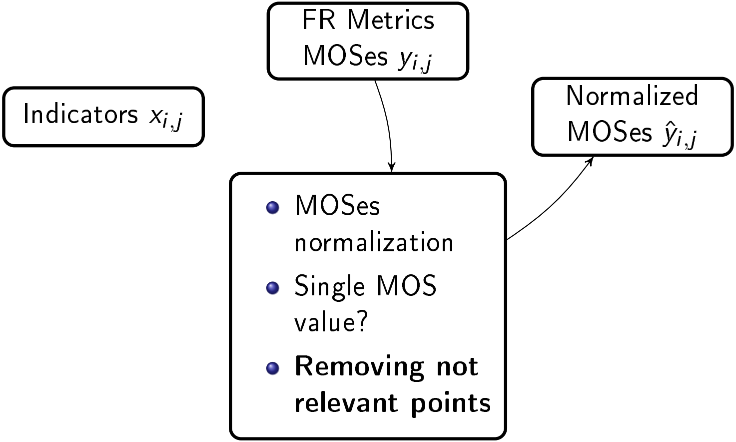
Data Flows


FR Metrics
MOSes $y_{i,j}$
Indicators $x_{i,j}$
MOSes normalization
Single MOS value?
Removing not relevant points

Single MOS or Multiple MOSes

- More points more information
- An alternative solution is regression including errors
- Is an error for 3-4 points really meaning full?
- If all points are used a perfect metric has R^2 different from 1 since the metrics inaccuracy makes the target function irrelevant

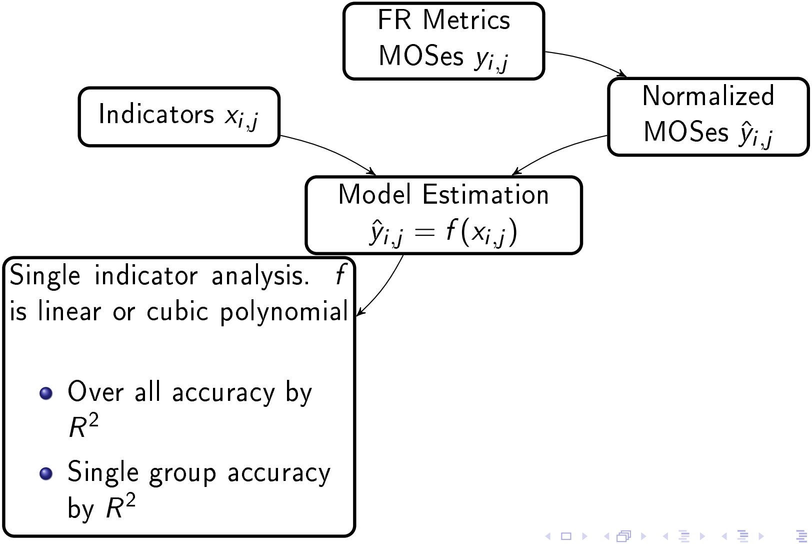
Data Flows


FR Metrics
MOSes $y_{i,j}$
Indicators $x_{i,j}$
Normalized
MOSes $\hat{y}_{i,j}$
MOSes normalization
Single MOS value?
Removing not relevant points

Removing not Relevant Points

- $|MOS_{i,1} - (aMOS_{i,2} + b)| > \epsilon$ is changed to $|\hat{y}_{i,1} - \hat{y}_{i,2}| > \epsilon$.
- For many metrics PVS_j is removed if $\max_j |\hat{y}_{i,j} - \hat{y}_{i,j}| > \epsilon$
- What should be ϵ value?
- Typical subjective experiment error? Which is?

Data Flows


FR Metrics MOSes $y_{i,j}$
Normalized MOSes $\hat{y}_{i,j}$
Indicators $x_{i,j}$
Model Estimation $\hat{y}_{i,j} = f(x_{i,j})$
Single indicator analysis. f is linear or cubic polynomial
Over all accuracy by R^2
Single group accuracy by R^2

PVS Grouping

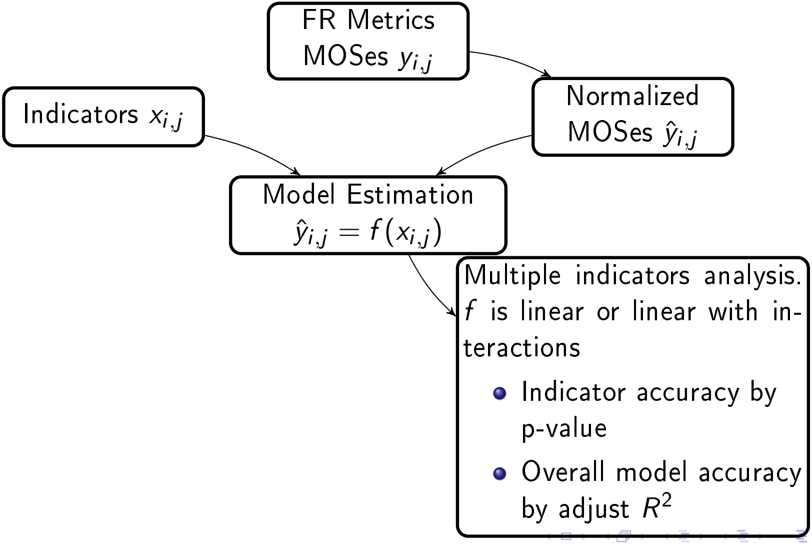
- An indicator can work just for some PVSes (e.g. frame rate change only)
- Single PVS can be a member of many different groups
- Each analysis is performed on each group including all data
- Group of PVSes is a set of i indexes I_k

Analysis

Let us consider metric number 1

- Estimate a and b of a linear fit $\hat{y}_{i,j} = ax_{i,1} + b$ for $i \in I_k$ and all j and k
- Calculate R^2 for the obtained fit
- $x_{\cdot,1}$ normalization by $\frac{x_{i,1} - \bar{x}_1}{\sigma_{x_{i,1}}}$ where $\sigma_{x_{i,1}}$ is the standard deviation
- Estimate a cubic fit for $\hat{y}_{i,j} = \sum_{l=0}^{3} a_l x_{i,1}^l$ for $i \in I_k$ and all j and k
- Chucking p-values of a_l - is the metric non linear?
- Calculate R^2 for the obtained fit

Data Flows


FR Metrics
MOSes $y_{i,j}$
Indicators $x_{i,j}$
Normalized
MOSes $\hat{y}_{i,j}$
Model Estimation
$\hat{y}_{i,j} = f(x_{i,j})$
Multiple indicators analysis. f is linear or linear with interactions
Indicator accuracy by p-value
Overall model accuracy by adjust R^2

Analysis

- $x_{\cdot,1}$ normalization by $\frac{x_{i,1}-\bar{x}_1}{\sigma_{x_{i,1}}}$ where $\sigma_{x_{i,1}}$ is standard deviation
- Estimate a linear fit $\hat{y}_{i,j} = \sum_l a_l x_{i,b_l} + b$ for $i \in I_k$ and all j, k and all possible combinations of b_l values - if possible $\hat{y}_{i,j} = a_1 x_{i,1} + a_2 x_{i,2} + b$; $\hat{y}_{i,j} = a_1 x_{i,1} + a_2 x_{i,3} + b$; $\hat{y}_{i,j} = a_1 x_{i,2} + a_2 x_{i,3} + b$; $\hat{y}_{i,j} = a_1 x_{i,1} + a_2 x_{i,2} + a_3 x_{i,3} + b$
- First step is p-values analysis: is particular indicator statistically important?
- Removing not statistically important indicators and than adjust R^2 computation

Analysis

- Metrics "cooperation" $\hat{y}_{i,j} = a_0 + a_1 x_{i,1} + a_2 x_{i,2} + a_3 x_{i,1} x_{i,2}$ for $i \in I_k$ and all j and k
- First step is p-values analysis: is particular indicator statistically important? Looking especially on the metrics products coefficients
- Removing not statistically important indicators and than adjust R^2 computation