



MRMR

Michael
Matějů

Motivation

Introduction

MRMR
Explained

End of Story

Maximum Relevancy, Minimum Redundancy

Michael Matějů

AI Squad

October 5, 2024



Outline

MRMR

Michael
Matějů

Motivation

Introduction

MRMR
Explained

End of Story

- 1 Motivation
- 2 Introduction
- 3 MRMR Explained
- 4 End of Story



Motivation

MRMR

Michael
Matějů

Motivation

Introduction

MRMR
Explained

End of Story

- Recent popularity of the algorithm due to UBER's paper (2019).
- The original paper from 2005.



Introduction

MRMR

Michael
Matějů

Motivation

Introduction

MRMR
Explained

End of Story

Feature selection methods allow you to choose the best features in your dataset, with respect to some criteria. Feature selection methods can be divided into three families:

- Filter Methods: Feature selection is made as part of the pre-processing, that is, before we train a model. We filter out features that perform poorly based on some criteria (e.g. correlation).
- Wrapper Methods: Iteratively choose a subset of features, train your model, and choose the best combination. Very time consuming.
- Embedded Methods: Embedded methods take advantage of the feature importance estimations which are embedded in the algorithm. E.g. Random Forest.



Introduction

MRMR

Michael
Matějů

Motivation

Introduction

MRMR
Explained

End of Story

Feature selection methods allow you to choose the best features in your dataset, with respect to some criteria. Feature selection methods can be divided into three families:

- **Filter Methods:** Feature selection is made as part of the pre-processing, that is, before we train a model. We filter out features that perform poorly based on some criteria (e.g. correlation).
- **Wrapper Methods:** Iteratively choose a subset of features, train your model, and choose the best combination. Very time consuming.
- **Embedded Methods:** Embedded methods take advantage of the feature importance estimations which are embedded in the algorithm. E.g. Random Forest.



Introduction

MRMR

Michael
Matějů

Motivation

Introduction

MRMR
Explained

End of Story

Feature selection methods allow you to choose the best features in your dataset, with respect to some criteria. Feature selection methods can be divided into three families:

- **Filter Methods:** Feature selection is made as part of the pre-processing, that is, before we train a model. We filter out features that perform poorly based on some criteria (e.g. correlation).
- **Wrapper Methods:** Iteratively choose a subset of features, train your model, and choose the best combination. Very time consuming.
- **Embedded Methods:** Embedded methods take advantage of the feature importance estimations which are embedded in the algorithm. E.g. Random Forest.



Introduction

MRMR

Michael
Matějů

Motivation

Introduction

MRMR
Explained

End of Story

Feature selection methods allow you to choose the best features in your dataset, with respect to some criteria. Feature selection methods can be divided into three families:

- **Filter Methods:** Feature selection is made as part of the pre-processing, that is, before we train a model. We filter out features that perform poorly based on some criteria (e.g. correlation).
- **Wrapper Methods:** Iteratively choose a subset of features, train your model, and choose the best combination. Very time consuming.
- **Embedded Methods:** Embedded methods take advantage of the feature importance estimations which are embedded in the algorithm. E.g. Random Forest.



Cheat Sheet

MRMR

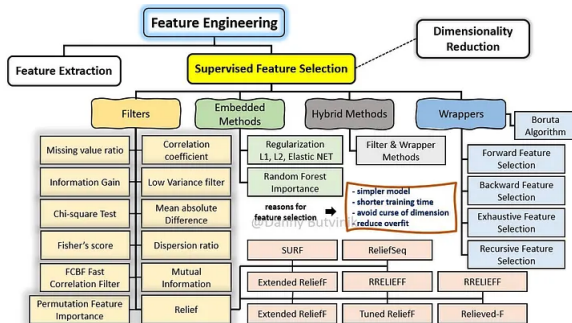
Michael
Matějů

Motivation

Introduction

MRMR
Explained

End of Story





MRMR

MRMR

Michael
Matějů

Motivation

Introduction

MRMR
Explained

End of Story

MRMR works iteratively. At each iteration, it identifies the best feature (according to a rule) and adds it to the basket of selected features. Once a feature goes into the bucket, it cannot ever come out.

In each iteration i a score is computed for each feature f according to formula:

$$score_i(f) = \frac{relevance(f|target)}{redundancy(f|features\ selected\ until\ (i-1))} ,$$

where we have to specify the relevancy and redundancy functions.



MRMR

MRMR

Michael
Matějů

Motivation

Introduction

MRMR
Explained

End of Story

MRMR works iteratively. At each iteration, it identifies the best feature (according to a rule) and adds it to the basket of selected features. Once a feature goes into the bucket, it cannot ever come out.

In each iteration i a score is computed for each feature f according to formula:

$$score_i(f) = \frac{relevance(f|target)}{redundancy(f|features\ selected\ until\ (i-1))} ,$$

where we have to specify the relevancy and redundancy functions.



Feature importance

MRMR

Michael
Matějů

Motivation

Introduction

MRMR
Explained

End of Story

There are couple of variants to calculate the feature importance from the original UBER paper. For discrete features:

- Mutual Information Difference (MID):

$$f^{MID}(X_i) = I(Y, X_i) - \frac{1}{|S|} \sum_{X_s \in S} I(X_s, X_i)$$

- Mutual Information Quotient (MIQ):

$$f^{MID}(X_i) = I(Y, X_i) / \left(\frac{1}{|S|} \sum_{X_s \in S} I(X_s, X_i) \right)$$



Feature importance

MRMR

Michael
Matějů

Motivation

Introduction

MRMR
Explained

End of Story

There are couple of variants to calculate the feature importance from the original UBER paper. For discrete features:

- Mutual Information Difference (MID):

$$f^{MID}(X_i) = I(Y, X_i) - \frac{1}{|S|} \sum_{X_s \in S} I(X_s, X_i)$$

- Mutual Information Quotient (MIQ):

$$f^{MID}(X_i) = I(Y, X_i) / \left(\frac{1}{|S|} \sum_{X_s \in S} I(X_s, X_i) \right)$$



Feature importance

MRMR

Michael
Matějů

Motivation

Introduction

MRMR
Explained

End of Story

For continuous features:

- F-test Correlation Difference (FCD):

$$f^{MID}(X_i) = F(Y, X_i) - \frac{1}{|S|} \sum_{X_s \in S} \rho(X_s, X_i)$$

- F-test Correlation Quotient (FCQ):

$$f^{MID}(X_i) = F(Y, X_i) / \left(\frac{1}{|S|} \sum_{X_s \in S} \rho(X_s, X_i) \right)$$

where ρ is Pearson correlation.



Feature importance

MRMR

Michael
Matějů

Motivation

Introduction

MRMR
Explained

End of Story

For continuous features:

- F-test Correlation Difference (FCD):

$$f^{MID}(X_i) = F(Y, X_i) - \frac{1}{|S|} \sum_{X_s \in S} \rho(X_s, X_i)$$

- F-test Correlation Quotient (FCQ):

$$f^{MID}(X_i) = F(Y, X_i) / \left(\frac{1}{|S|} \sum_{X_s \in S} \rho(X_s, X_i) \right)$$

where ρ is Pearson correlation.



Feature importance

MRMR

Michael
Matějů

Motivation

Introduction

MRMR
Explained

End of Story

Another extensions:

- RDC algorithm for redundancy function (The randomized dependence coefficient)
- Model-based Feature importance for relevance:
 - Random Forest Correlation Quotient
 - Random Forest RDC Quotient
 - Pure Random Forest

From benchmarks, RF variances are good if down-stream model is Random Forest. FCQ is pretty close and works on more models.



Feature importance

MRMR

Michael
Matějů

Motivation

Introduction

MRMR
Explained

End of Story

Another extensions:

- RDC algorithm for redundancy function (The randomized dependence coefficient)
- Model-based Feature importance for relevance:
 - Random Forest Correlation Quotient
 - Random Forest RDC Quotient
 - Pure Random Forest

From benchmarks, RF variances are good if down-stream model is Random Forest. FCQ is pretty close and works on more models.



Feature importance

MRMR

Michael
Matějů

Motivation

Introduction

MRMR
Explained

End of Story

Another extensions:

- RDC algorithm for redundancy function (The randomized dependence coefficient)
- Model-based Feature importance for relevance:
 - Random Forest Correlation Quotient
 - Random Forest RDC Quotient
 - Pure Random Forest

From benchmarks, RF variances are good if down-stream model is Random Forest. FCQ is pretty close and works on more models.



Example

MRMR

Michael
Matějů

Motivation

Introduction

MRMR

Explained

End of Story

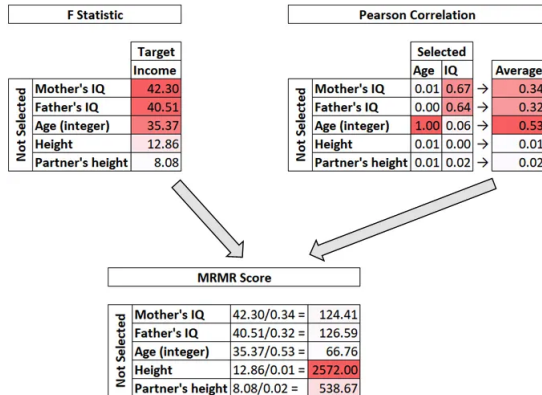


Illustration of what happens at iteration 3. [Figure by Author]



The End

MRMR

Michael
Matějů

Motivation

Introduction

MRMR
Explained

End of Story

Thank you for your attention and patience.