R Data Packages of Multi-Reader Multi-Case Studies and Simulation Tools to Support the Development of Reader Performance Evaluation Methods

Si Wen, Qi Gong, Brandon Gallas
FDA/CDRH/OSEL Division of Imaging, Diagnostics, and Software Reliability
Outline

• MRMC studies and analysis
  – Multi-reader, Multi-case
• ROC: Receiver Operating Characteristic curves
• R Data packages
  – What and how
• Show examples of MRMC data analysis

• Why
  – Disseminate our research data, methods, and tools
  – Enables other investigators to develop methods and tools
Background

- Medical device evaluation
  - Assessing new device performance
  - Assessing computer aided diagnosis devices

Comparing two modalities in a medical task

AI detected potential lesion

Motivation

• Design a study to compare two modalities for a medical task
  – Ask clinicians to review cases under the two different modalities

• Random factors in the study:
  – **Cases**: randomly select samples from intended case population
  – **Readers**: randomly select samples from intended reader population

• **Multi-Reader Multi-Case (MRMC) Study**
  – Each case is reviewed by multiple clinicians/readers
  – Each reader review multiple cases

• **MRMC analysis**
  – Account for reader and case variability
  – Generalize results to the population of readers and cases
Multi-Reader Multi-Case Studies

• Application
  – Comparing manual scores from clinicians by using different medical device
  – Comparing machine learning algorithm result with manual scores
  – Comparing the results from different machine learning algorithms

Medical Imaging and other technologies
## Multi-Reader Multi-Case Studies

### Data Structure

<table>
<thead>
<tr>
<th>Truth</th>
<th>Modality A</th>
<th>Modality B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 1 1 1 1 0 0 0 0</td>
<td>90 70 100 40 50 10 20 80 60 30</td>
<td>70 80 100 90 50 20 60 40 30 10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Truth</th>
<th>Modality A</th>
<th>Modality B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 1 1 1 1 0 0 0 0</td>
<td>90 70 100 40 50 10 20 80 60 30</td>
<td>70 80 100 90 50 20 60 40 30 10</td>
</tr>
</tbody>
</table>

### Reader Study Result

<table>
<thead>
<tr>
<th>readerID</th>
<th>caseID</th>
<th>modalityID</th>
<th>confidenceScore</th>
</tr>
</thead>
<tbody>
<tr>
<td>reader1</td>
<td>case1</td>
<td>A</td>
<td>90</td>
</tr>
<tr>
<td>reader1</td>
<td>case2</td>
<td>A</td>
<td>70</td>
</tr>
<tr>
<td>reader1</td>
<td>case3</td>
<td>A</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>reader1</td>
<td>case1</td>
<td>A</td>
<td>70</td>
</tr>
<tr>
<td>reader2</td>
<td>case2</td>
<td>A</td>
<td>80</td>
</tr>
<tr>
<td>reader2</td>
<td>case3</td>
<td>A</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>reader2</td>
<td>case1</td>
<td>B</td>
<td>80</td>
</tr>
<tr>
<td>reader1</td>
<td>case2</td>
<td>B</td>
<td>50</td>
</tr>
<tr>
<td>reader1</td>
<td>case3</td>
<td>B</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>reader2</td>
<td>case7</td>
<td>B</td>
<td>70</td>
</tr>
<tr>
<td>reader2</td>
<td>case8</td>
<td>B</td>
<td>20</td>
</tr>
<tr>
<td>reader2</td>
<td>case9</td>
<td>B</td>
<td>50</td>
</tr>
<tr>
<td>reader2</td>
<td>case10</td>
<td>B</td>
<td>10</td>
</tr>
</tbody>
</table>

### Ground Truth

<table>
<thead>
<tr>
<th>caseID</th>
<th>Truth</th>
</tr>
</thead>
<tbody>
<tr>
<td>case1</td>
<td>1</td>
</tr>
<tr>
<td>case2</td>
<td>1</td>
</tr>
<tr>
<td>case3</td>
<td>1</td>
</tr>
<tr>
<td>case4</td>
<td>1</td>
</tr>
<tr>
<td>case5</td>
<td>1</td>
</tr>
<tr>
<td>case6</td>
<td>0</td>
</tr>
<tr>
<td>case7</td>
<td>0</td>
</tr>
<tr>
<td>case8</td>
<td>0</td>
</tr>
<tr>
<td>case9</td>
<td>0</td>
</tr>
<tr>
<td>case10</td>
<td>0</td>
</tr>
</tbody>
</table>
Multi-Reader Multi-Case Studies

• Receiver Operating Characteristic (ROC) analysis

<table>
<thead>
<tr>
<th>t=90</th>
<th>Truth=1</th>
<th>Truth=0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score≥90</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Score&lt;90</td>
<td>3</td>
<td>5</td>
</tr>
</tbody>
</table>

Sensitivity = 2/5 = 0.4  Specificity = 5/5=1

• AUC represents the probability that a random positive case is positioned to the right(>) of a random negative case
Multi-Reader Multi-Case Studies

- MRMC ROC analysis for two modalities

<table>
<thead>
<tr>
<th>Truth</th>
<th>1 1 1 1 1 0 0 0 0 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modality A</td>
<td>90 70 100 40 50 10 20 80 60 30</td>
</tr>
<tr>
<td>Modality B</td>
<td>70 80 100 90 50 20 60 40 30 10</td>
</tr>
<tr>
<td>Modality A</td>
<td>80 50 100 70 20 30 10 60 40 90</td>
</tr>
<tr>
<td>Modality B</td>
<td>30 80 100 90 40 60 70 20 50 10</td>
</tr>
</tbody>
</table>

- Difference of reader-averaged AUC = 0.88 − 0.72 = 0.16
- Variance of difference of reader-averaged AUC - hard

Reader-averaged AUC in modality A
= \( \frac{1}{2} (0.80 + 0.96) = 0.88 \)

Reader-averaged AUC in modality B
= \( \frac{1}{2} (0.68 + 0.76) = 0.72 \)
Multi-Reader Multi-Case Studies

Difficulties

- Time consuming and labor intensive not much data is publicly available
- Complicated correlation structure when computing variance of reader-averaged AUC

Solutions

- MRMC data storage and distribution
  - R data packages
- MRMC performance analysis
  - iMRMC R package
R data package for MRMC study

• To store and **distribute** data along with its **documentation**
  – Well-organized documentation: description, format, source, etc.
  – Public available: CRAN(The Comprehensive R Archive Network [https://cran.r-project.org/](https://cran.r-project.org/)) or GitHub ([https://github.com/](https://github.com/))

• To provide **use case** for existing and future functions/methods
  – Pre-processed reproducible data that can be easily accessed in R

• To help establishing **future studies**
  – Acting as a baseline for similar study
  – Providing information for sizing future study in this area
Building up an R data package

• Tool: devtools ([https://www.rstudio.com/products/rpackages/devtools/]())
• Steps:
  – devtools::create() create the package skeleton
  – devtools::use_data_raw() create data-raw/
  – Put data preprocessing script into .R file, which reads in raw data and at calls usethis::use_data() to save .rda formatted data files in data/
MRMC R data packages

• Structure
  – colorScaleStudyData (https://github.com/DIDSR/colorScaleStudyData)

  • **Description** (required as a R package) contains the package name, title, and description
  • **Readm.md** (for GitHub) contains the instruction information of the package
MRMC R data packages

- **Structure**
  - colorScaleStudyData
    - [https://github.com/DIDSR/colorScaleStudyData](https://github.com/DIDSR/colorScaleStudyData)

  **data**
  - Cardio Study_truth.csv
  - CardioCTstudy_3mod_12obs_210cases_rawdata.csv
  - Prostate_Mode_09obs_165cases_wGSDFrawdata_Allreaders.csv
  - Prostate_Truth.csv

  **data-ready**
  - cardioCTGroundTruth.R
  - cardioCTReaderData.R
  - prostateMRIGroundTruth.R
  - prostateMRIReaderData.R

- **Data-ready** contains the raw dataset and functions to output the R format data
MRMC R data packages

• Structure
  – colorScaleStudyData (https://github.com/DIDSR/colorScaleStudyData)

  ```r
  > library(colorScaleStudyData)
  > head(cardioCTReaderData)
  readerID  caseID  modalityID confidenceScore
  1    Lowl    141       Grayscale         31
  2    Lowl    142       Grayscale         54
  3    Lowl    143       Grayscale         88
  4    Lowl    144       Grayscale         88
  5    Lowl    145       Grayscale         54
  6    Lowl    146       Grayscale         32
  ```

• Data contains the R format dataset can be read in R easily
MRMC R data packages

• Documentation
  – colorScaleStudyData (https://github.com/DIDSR/colorScaleStudyData)

• The documentation is edited in the data.R file in R
• devtools::document() generate .Rd manuals in man
MRMC R data packages

• Build the package

  – `devtools:: build()` build source package and generate .tar.gz file

    • No ./data-raw and .Rproj in the zip file

  – In Rstudio, you can also build your package by click Build & Reload
MRMC R data packages - Summary

• To store and distribute data along with its documentation
  – Structure
  – Documentation
  – How to create

• To provide use case for functions in existing and future packages
  – How to use : apply doIMRMC in iMRMC package to the data
MRMC Data Analysis

• iMRMC package ([https://github.com/DIDSR/iMRMC](https://github.com/DIDSR/iMRMC))
  – A stand-alone, precompiled Java application developed by Dr. Brandon Gallas’ lab

  – To analyze and size MRMC reader studies that compare the difference AUCs from two modalities, including the U-statistics estimation of variance of the difference of reader-averaged AUC

  – iMRMC R package: an R package that utilizes the core Java application and output the R list object that contains all the analysis result. ([https://cran.r-project.org/web/packages/iMRMC/index.html](https://cran.r-project.org/web/packages/iMRMC/index.html))
MRMC Data Analysis

• iMRMC R package core functions:
  – doIMRMC:
    • Takes ROC data as a data frame and runs a MRMC analysis based on U-statistics
  – sim.gRoeMetz:
    • Simulate an MRMC data set of an ROC study comparing two modalities
    • Allow testing other new methods
MRMC Data Analysis

- Apply doIMRMC in iMRMC package to color scale study data

We need to merge both `cardioCTGroundTruth`

<table>
<thead>
<tr>
<th>caseID</th>
<th>truth</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>positive</td>
</tr>
<tr>
<td>2</td>
<td>positive</td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>209</td>
<td>209</td>
</tr>
<tr>
<td>210</td>
<td>210</td>
</tr>
</tbody>
</table>

and `cardioCTReaderData`

<table>
<thead>
<tr>
<th>readerID</th>
<th>caseID</th>
<th>modalityID</th>
<th>confidenceScore</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Low1</td>
<td>141</td>
<td>Grayscale</td>
</tr>
<tr>
<td>2</td>
<td>Low1</td>
<td>142</td>
<td>Grayscale</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1920</td>
<td>Low8</td>
<td>139</td>
<td>Rainbow</td>
</tr>
<tr>
<td>1921</td>
<td>Low8</td>
<td>140</td>
<td>Rainbow</td>
</tr>
</tbody>
</table>

1. Cardiac CT study

- 3 modalities color scales: Grayscale, Rainbow, Hotiron.
- 12 readers (observers) Readers had different experience level, which is referenced in the reader nickname as High and Low, for high and low level of experience, respectively.
- 210 cases 105 positive for lesion.
- Study was not fully crossed. There were three groups of readers, and each group read 35 positive and 35 negative cases.

`combinedDf`

<table>
<thead>
<tr>
<th>readerID</th>
<th>caseID</th>
<th>modalityID</th>
<th>confidenceScore</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>truth</td>
<td>1</td>
<td>truth</td>
</tr>
<tr>
<td>2</td>
<td>truth</td>
<td>2</td>
<td>truth</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2130</td>
<td>Low8</td>
<td>139</td>
<td>Rainbow</td>
</tr>
<tr>
<td>2131</td>
<td>Low8</td>
<td>140</td>
<td>Rainbow</td>
</tr>
</tbody>
</table>

[https://github.com/DIDSR/colorScaleStudyData/blob/master/inst/extra/applyIMRMC.pdf](https://github.com/DIDSR/colorScaleStudyData/blob/master/inst/extra/applyIMRMC.pdf)
MRMC Data Analysis

- Apply doIMRMC in iMRMC package to color scale study data

```r
library(iMRMC)
result <- doIMRMC(data = combinedDf)
```

- AUC and variance of AUC for each modality by U-statistics method

```r
data.frame(rbind(result$Ustat$AUCA[1:3], result$Ustat$varAUCA[1:3],
sqrt(result$Ustat$varAUCA[1:3])),
row.names = c("AUC", "variance of AUC", "SE of AUC"))
```

Table 6: AUC for different modalities: Ustat

<table>
<thead>
<tr>
<th></th>
<th>Grayscale</th>
<th>Hot</th>
<th>Rainbow</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUC</td>
<td>0.5902954</td>
<td>0.5671724</td>
<td>0.5176793</td>
</tr>
<tr>
<td>variance of AUC</td>
<td>0.0010402</td>
<td>-0.0000219</td>
<td>0.0000338</td>
</tr>
<tr>
<td>SE of AUC</td>
<td>0.0322528</td>
<td>NaN</td>
<td>0.0058162</td>
</tr>
</tbody>
</table>

https://github.com/DIDSR/colorScaleStudyData/blob/master/inst/extra/applyIMRMC.pdf
MRMC Data Analysis

- Apply doIMRMC in iMRMC package to color scale study data

```r
library(iMRMC)
result <- doIMRMC(data = combinedDf)
```

- Difference of AUC and variance of difference of AUC for pairwise comparison of different modalities by U-statistics method

<table>
<thead>
<tr>
<th></th>
<th>Grayscale vs. Hot</th>
<th>Grayscale vs. Rainbow</th>
<th>Hot vs. Rainbow</th>
</tr>
</thead>
<tbody>
<tr>
<td>difference of AUC</td>
<td>0.0231230</td>
<td>0.0726161</td>
<td>0.0494932</td>
</tr>
<tr>
<td>variance of difference of AUC</td>
<td>0.0012101</td>
<td>0.0014347</td>
<td>0.0010837</td>
</tr>
<tr>
<td>SE of difference of AUC</td>
<td>0.0347870</td>
<td>0.0378770</td>
<td>0.0329203</td>
</tr>
<tr>
<td>95% CI lower bound</td>
<td>-0.0450596</td>
<td>-0.0016227</td>
<td>-0.0150307</td>
</tr>
<tr>
<td>95% CI upper bound</td>
<td>0.0913055</td>
<td>0.1468550</td>
<td>0.1140170</td>
</tr>
</tbody>
</table>

https://github.com/DIDSR/colorScaleStudyData/blob/master/inst/extra/applyIMRMC.pdf
MRMC Data Analysis

• R-markdown files
  – Demonstrate the use of functions and data
  – colorScaleStudyData
    (https://github.com/DIDSR/colorScaleStudyData/blob/master/inst/extra/applyIMRMC.Rmd)

• More package application
  – viperData R data package (https://didsr.github.io/viperData/)
MRMC Data Analysis

• iMRMC R package core functions:
  – **doIMRMC:**
    • Takes ROC data as a data frame and runs a MRMC analysis based on U-statistics
  – **sim.gRoeMetz:**
    • Simulate an MRMC data set of an ROC study comparing two modalities
    • Allow testing other new methods
Simulation Tool for MRMC Data

• Roe and Metz model for MRMC ROC analysis

\[ X_{ijkt}^{R&M} = \mu_t + \tau_{it} + R_jt + C_{kt} + [RC]_{jkt} + [\tau R]_{ijt} + [\tau C]_{ikt} + [\tau RC]_{ijkt} + E_{ijkt} \]

– where \( X_{ijkt} \) denotes the score for modality \( i \), reader \( j \), case \( k \), and truth state \( t \)

– modality and truth state are fixed factors (\( \mu_t \) and \( \tau_{it} \) are fixed effects)

– reader and case are random factors (the remaining terms are independent zero-mean Gaussian random variables with its own variance)
Simulation Tool for MRMC Data

- Apply `sim.gRoeMetz` to simulate MRMC data

```r
> config<-sim.gRoeMetz.config()
> df.MRMC <- sim.gRoeMetz(config)
> View(df.MRMC)
> ggplot(subset(df.MRMC,modalityID %in% c("testA","testB")),
+ aes(x=score,
+     color=factor(unlist(lapply(as.character(caseID),
+                               function(x){strsplit(x,'case')[[1]][1]}))))
+ + geom_density(position = "identity", alpha=0.2) +
+ facet_grid(rows = "modalityID") +
+ labs(x = "MRMC Reading Score", y = "Density") +
+ theme(legend.title = element_blank())
> result <- doIMRMC(df.MRMC)
> ROC.data <- rbind(data.frame(FPF = result$ROC$testA: Diagonal Average $fpf,
+ TPF = result$ROC$testA: Diagonal Average $tpf),
+ data.frame(FPF = result$ROC$testB: Diagonal Average $fpf,
+ TPF = result$ROC$testB: Diagonal Average $tpf))
> ROC.data$modality = rep(c("testA","testB"), each = nrow(ROC.data)/2)
> ggplot(ROC.data,aes(x=FPF,y=TPF,color=modality)) +
+ geom_point(size=0.5) +
+ labs(x="False Positive Fraction",y="True Positive Fraction")
```
Summary

• Multi-Reader Multi-Case (MRMC) study
  
  – Data storage and distribution: \texttt{R data Package}
  
  – MRMC ROC analysis: \texttt{doIMRMC in iMRMC package}
  
  – MRMC simulation tool: \texttt{sim.gRoeMetz in iMRMC package}
Reference Links

- colorScaleStudyData R data package [https://github.com/DIDSR/colorScaleStudyData](https://github.com/DIDSR/colorScaleStudyData)
- viperData R data package [https://didsr.github.io/viperData/](https://didsr.github.io/viperData/)
- iMRMC package [https://github.com/DIDSR/iMRMC](https://github.com/DIDSR/iMRMC)
- iMRMC R package [https://cran.r-project.org/web/packages/iMRMC/index.html](https://cran.r-project.org/web/packages/iMRMC/index.html)
- References:
Thank you!

Q&A