Relational Event Modeling (REM) Project README

2024 May

Contents

Relational Event Modeling (REM) Project

Overview

This document provides an overview of the Relational Event Modeling (REM) project, including the required packages, data preparation, methodology, and results analysis.

Required Packages

The following R packages are necessary to run the REM analysis:

```
if (!require("igraph")) install.packages("igraph")
```

```
## Loading required package: igraph
```

```
##
## Attaching package: 'igraph'
```

```
## The following objects are masked from 'package:stats':
##
##     decompose, spectrum
```

```
## The following object is masked from 'package:base':
##
##     union
```

```
if (!require("rem")) install.packages("rem")
```

```
## Loading required package: rem
```

```
if (!require("network")) install.packages("network")
```

```
## Loading required package: network
```

```
##
## 'network' 1.18.2 (2023-12-04), part of the Statnet Project
## * 'news(package="network")' for changes since last version
## * 'citation("network")' for citation information
## * 'https://statnet.org' for help, support, and other information
```

```
##
## Attaching package: 'network'
```

```
## The following objects are masked from 'package:igraph':
##
##     %c%, %s%, add.edges, add.vertices, delete.edges, delete.vertices,
##     get.edge.attribute, get.edges, get.vertex.attribute, is.bipartite,
##     is.directed, list.edge.attributes, list.vertex.attributes,
##     set.edge.attribute, set.vertex.attribute
```

```
if (!require("tidyr")) install.packages("tidyr")
```

```
## Loading required package: tidyr
```

```
##
## Attaching package: 'tidyr'
```

```
## The following object is masked from 'package:igraph':
##
##     crossing
```

```
if (!require("caret")) install.packages("caret")
```

```
## Loading required package: caret
```

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

```r
if (!require("survival")) install.packages("survival")
```

```
## Loading required package: survival
```

```
##
## Attaching package: 'survival'
```

```
## The following object is masked from 'package:caret':
##
##     cluster
```

```r
if (!require("dplyr")) install.packages("dplyr")
```

```
## Loading required package: dplyr
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:igraph':
##
##     as_data_frame, groups, union
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
if (!require("ggplot2")) install.packages("ggplot2")
if (!require("ggraph")) install.packages("ggraph")
```

```
## Loading required package: ggraph
```

```r
library(knitr)
# do not echo or run code
knitr::opts_chunk$set(echo = FALSE, eval = FALSE)
```

Overview

Relational Event Modeling (REM) is a statistical framework for analyzing time-stamped interaction data. It is particularly useful in social network analysis where the goal is to understand the dynamics of interactions between actors over time.

Data Preparation

Before fitting a relational event model, data must be prepared in a specific format. The data should include:

- **Timestamp and Location:** Each interaction event should be time-stamped
- **Speaker Sequence:** The sequence of speakers or actors involved in each event.
- **Performance Score:** An indicator of the performance level (e.g., high or low) associated with the event.
- **Dialogue Act Classification:** Each interaction should be classified according to its dialogue act type (e.g., statement, question, backchannel).

Data processing involves cleaning and preparing the interaction data and actor data:

1. **Interactions Data**: Contains information about the sender, receiver, time of interaction, and type of interaction (dialogue act type).
2. **Actors Data**: Contains attributes of each actor, such as name and gender.

Refer to `data_processing.Rmd` for detailed steps on data cleaning and preparation.

Example Code

Project Directory Structure

- **data_processing.Rmd**: RMD file for data processing steps, including data cleaning and preparation.
- **rem_survival_analysis.Rmd**: RMD file for conducting survival analysis within the REM framework; unclude visualizations
- **finaldata.Rmd**: RMD file documenting the final dataset preparation and cleaning.
- **low_surv_analysis.Rmd**: RMD file for survival analysis of low-performing sessions.
- **high_survival_analysis.Rmd**: RMD file for survival analysis of high-performing sessions.
- **data**: Directory containing raw and processed data files used in the analysis.

Visualization

see `rem_survival_analysis.Rmd` for visualizations

Kaplan-Meier Survival Curves

Kaplan-Meier survival curves are used to estimate the survival function from the time-to-event data. This helps in visualizing the proportion of interactions that occur over time.

Hazard Function Plot

The hazard function describes the instant risk of an event occurring at a given time point.

Interaction Plot

- interactions between senders and receivers over time provide insights into the dynamics of communication.

These visualizations provide an overview of the interaction dynamics and survival probabilities within the REM framework. Adjust the visualizations based on your specific dataset and analysis requirements.

Methodology

Survival Analysis

Survival analysis is used to understand the likelihood of interaction events over time. The analysis involves creating survival objects and fitting Cox proportional hazards models.

Refer to `rem_survival_analysis.Rmd`, `high_survival_analysis.Rmd`, and `low_surv_analysis.Rmd` for detailed analysis.

Example Code

Incremental Model Building

Incremental model building involves fitting several models to understand the effect of different factors on interaction events:

- **Model 0**: Baseline hazard function.
- **Model 1**: Effect of sender attributes.
- **Model 2**: Effect of receiver attributes.
- **Model 3**: Combined sender and receiver effects.
- **Model 4**: Effect of dialogue act types.
- **Model 5**: Combined sender and dialogue act effects.

Refer to `modelresult.Rmd` for detailed steps and results of the incremental model building.

Results and Analysis

Comparative Insights

The analysis compares high-performing and low-performing sessions based on the REM models. Key findings include:

- **High-Performing Sessions**: Characterized by more active and engaging conversations with influential actors driving the interactions.
- **Low-Performing Sessions**: Show less dynamic interactions, often influenced by less popular participants.

Hazard Ratios

- **HR $>$ 1**: Increases the likelihood of the event.
- **HR $<$ 1**: Decreases the likelihood of the event.
- **HR $=$ 1**: No effect on the event likelihood.

For example, a high hazard ratio for questions indicates that asking questions significantly drives subsequent interactions.

Future Work

- **Incorporate Degree Centrality:** To understand the influence of key members in triggering interactions.
- **Predict Session Performance:** Using features like participant gender, type and sequence of dialogues, and interaction frequency.
- **Tune Classification Models:** For better accuracy in classifying sessions as high or low performing.