Bayesian analysis of the NESTA study of interventions against verbal aggression online Technical Report

Rafal Urbaniak

Contents

Data and exploration

```
Hate <- readRDS(file = "datasets/RAWNESTA/Hate.rds")
Comments <- readRDS(file = "datasets/RAWNESTA/Comments.rds")
summaries <- read.csv(file = "datasets/Summaries.csv")

dates <- colnames(Hate)[-1]
dates <- as.Date(dates)
startDate <- dates[1]
interventionDate <- "2020-07-08"
observationDate <- "2020-09-09"
end <- dates[length(dates)]

periods <- numeric(length(dates))
periods <- ifelse(dates < interventionDate, "pre-treatment", periods)
periods <- ifelse(dates >= interventionDate & dates < observationDate,
    "treatment", periods)
periods <- ifelse(dates >= observationDate, "post-treatment", periods)

hateTS <- as.data.frame(colSums(Hate[, -1]))
hateTS$date <- as.Date(rownames(hateTS))
rownames(hateTS) <- NULL
colnames(hateTS) <- c("attacks", "date")
hateTS$periods <- periods

interventions <- readRDS(file = "datasets/interventions.rds")
interventionsTS <- as.data.frame(table(interventions$day))
interventionsTS$Var1 <- as.Date(interventionsTS$Var1)
```

```
colnames(interventionsTS) <- c("date", "interventions")

periodsDF <- merge(x = hateTS, y = interventionsTS, by = "date", all.x = TRUE)

idx <- c(1, diff(periodsDF$date))
i2 <- c(1, which(idx != 1), nrow(periodsDF) + 1)
periodsDF$grp <- rep(1:length(diff(i2)), diff(i2))

periodsDF$interventions[is.na(periodsDF$interventions) & periodsDF$periods ==
    "treatment"] <- 0

periodsPlot <- ggplot(periodsDF) + geom_line(aes(x = date, y = attacks,
    group = grp), alpha = 0.8, size = 0.6) + geom_line(aes(x = date, y = interventions,
    group = grp), alpha = 0.8, size = 0.6) + geom_vline(xintercept = startDate,
    lty = 2, size = 0.2, alpha = 0.5) + geom_vline(xintercept = as.Date(interventionDate),
    lty = 2, size = 0.2, alpha = 0.5) + geom_vline(xintercept = as.Date(observationDate),
    lty = 2, size = 0.2, alpha = 0.5) + geom_vline(xintercept = as.Date(end),
    lty = 2, size = 0.2, alpha = 0.5) + labs(title = "Attacks and interventions time series",
    subtitle = "no line at data gaps", caption = "days with data: 81 (pre-treatment), 62 (treatment), 72 (post-treatment)") +
    theme_tufte() + theme(axis.title.x = element_blank(), plot.caption = element_text(hjust = 0.5,
    face = "italic")) + scale_x_date(date_labels = "%b %d", breaks = c(startDate,
    as.Date(startDate), as.Date(interventionDate), as.Date(observationDate),
    end), limits = c(startDate - 30, end + 10)) + ylab("count") + annotate("rect",
    xmin = as.Date(interventionDate), xmax = as.Date(observationDate),
    ymin = -1, ymax = 360, alpha = 0.2, fill = "darkgreen") + ylim(c(-1,
    370)) + annotate("text", label = "pre-treatment", x = as.Date(startDate) +
    2, y = 370, hjust = 0) + annotate("text", label = "treatment", x = as.Date(interventionDate) +
    2, y = 370, hjust = 0) + annotate("text", label = "post-treatment",
    x = as.Date(observationDate) + 2, y = 370, hjust = 0) + annotate("text",
    label = "interventions:", x = as.Date(interventionDate) - 52, y = 15,
    hjust = 0) + annotate("text", label = "attacks:", x = as.Date(startDate) -
    30, y = 215, hjust = 0)

periodsDF$weekdays <- weekdays(as.Date(periodsDF$date))
periodsDF$weeks <- week(as.Date(periodsDF$date))

periodsDF$weekdays <- as.factor(periodsDF$weekdays)
levels(periodsDF$weekdays) <- c("Monday", "Tuesday", "Wednesday", "Thursday",
    "Friday", "Saturday", "Sunday")
weeksPlot <- ggplot(periodsDF) + geom_smooth(aes(x = weekdays, y = attacks,
    group = 1)) + geom_line(aes(x = weekdays, y = attacks, group = weeks),
    alpha = 0.1) + theme_tufte() + labs(title = "Weekly attacks over six months",
    subtitle = "No weekly patterns") + xlab("") + ylab("count")
```

For the duration of the project we selected 486 Reddit users and tracked their activity, starting on `rstartDate‘, beginning the intervention period on 2020-07-08, leading to a further observation period starting on 2020-09-09 and ending on 2020-11-20. The time series of attacks observed and of interventions conducted can be inspected in Figure 1.

```
# head(summaries) %>% kable( 'latex', booktabs = T) %>%
# kable_styling(latex_options = c('striped', 'scale_down') ,font_size
# = 9)
```

The basic variables we are dealing with are in the following table.

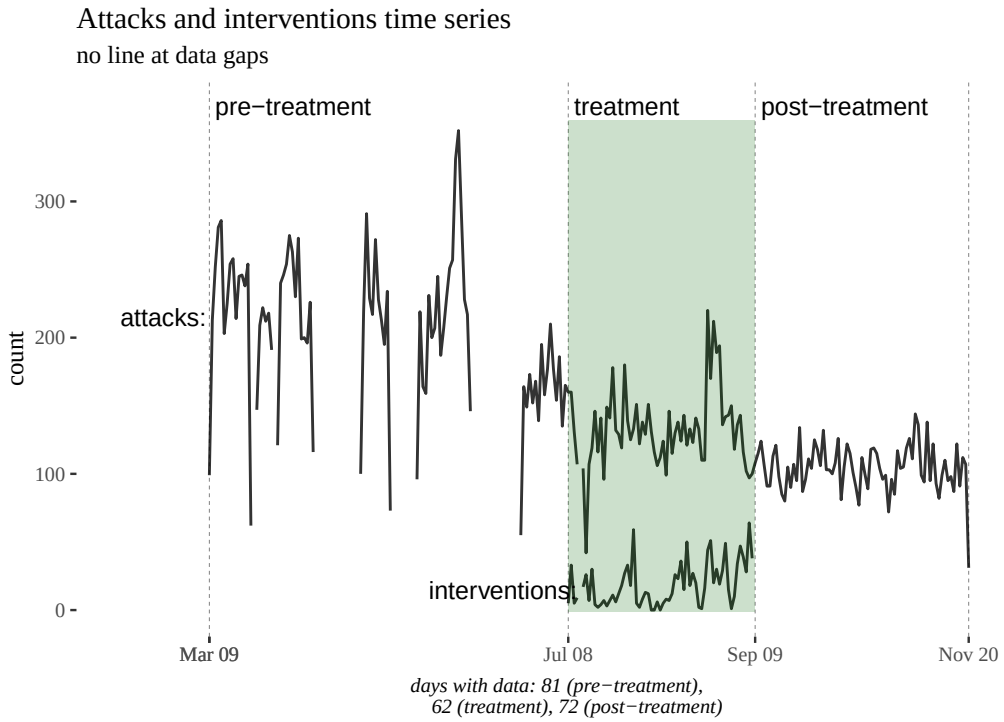

Figure 1: Daily sums of attacks and interventions throughout the three experimental periods.

variable	explanation
AB	attacks before (pre-treatment)
AD	attacks during (the treatment period)
AA	attacks after (post-treatment)
CB	comments before
CD	comments during
CA	comments after
group	treatment group
IC	intervention count