

RSDexport

Yichen Han

2024-08-09

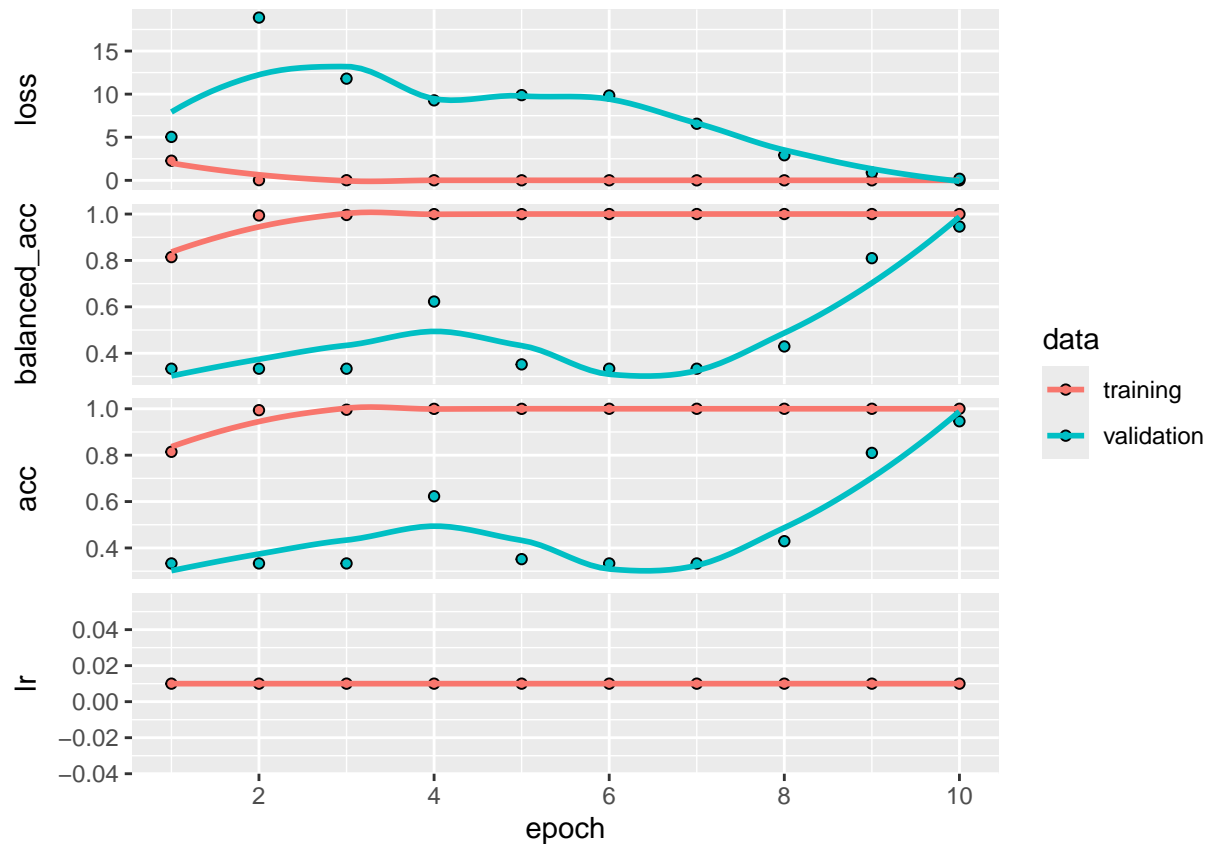
Experiment 39

Spec.region excluded for “abnormal”. Functional change not guaranteed.

Call: GenePermutation(triplets, keyed, num.perm=15000, min.subs=10, max.subs=30, spec.region=30:60)

Model: maxlen = 477, batch_size = 64, steps_per_epoch = 45, epochs = 10, step = c(1, 1, 1)

```
# load outputs/hist39.rds
hist39 <- readRDS(file="outputs/hist39.rds")
plot(hist39)
```



```
checkpoint_path <- file.path("checkpoints")
dir_path <- file.path("outputs")
```

```
run_name <- "rsd-permutation_39"
model <- load_cp(paste0(checkpoint_path, "/", run_name),
                 cp_filter = "last_ep")
```

```
## Using checkpoint checkpoints/rsd-permutation_39/Ep.010-val_loss0.18-val_acc0.946.hdf5
```

Model at last Epoch

```
path_special_test <- file.path("special/test")
path_normal_test <- file.path("normal/test")
path_abnormal_test <- file.path("abnormal/test")
eval_model <- evaluate_model(path_input = c(path_normal_test,
      path_abnormal_test, path_special_test),
      model = model,
      batch_size = 64,
      step = 5,
      vocabulary_label = list(c("normal", "abnormal", "special")),
      number_batches = 10,
      mode = "label_folder",
      verbose = FALSE
)

eval_model
```

```
## [[1]]
## [[1]]$confusion_matrix
##           Truth
## Prediction normal abnormal special
##   normal      254        26        0
##   abnormal      0        213        0
##   special       2         17       256
##
## [[1]]$accuracy
## [1] 0.9414062
##
## [[1]]$categorical_crossentropy_loss
## [1] 0.1702538
##
## [[1]]$AUC
## NULL
##
## [[1]]$AUPRC
## NULL
```

```
special_seq <- permute_sequence(triplets, keyed, type="ok", min.subs=5, max.subs=30,
                               dict=codon.dict, spec.cond=FALSE, spec.region=NULL)
special_seq <- permute_sequence(special_seq, keyed, type="func", min.subs=5, max.subs=30,
                               dict=codon.dict, spec.cond=TRUE, spec.region=30:60)
special_seq <- paste(special_seq, collapse = "")
onehot_instance <- seq_encoding_label(char_sequence = special_seq,
```

```

maxlen = 477,
start_ind = 1,
vocabulary = c("A", "C", "G", "T"))

head(onehot_instance[1,,])

```

```

##      [,1] [,2] [,3] [,4]
## [1,]    1    0    0    0
## [2,]    0    0    0    1
## [3,]    0    0    1    0
## [4,]    0    1    0    0
## [5,]    0    0    0    1
## [6,]    0    0    0    1

```

```

pred <- predict(model, onehot_instance, verbose = 0)
pred

```

```

##      [,1]      [,2] [,3]
## [1,] 2.989441e-09 3.559504e-18 1

```

```

ig <- integrated_gradients(
  input_seq = onehot_instance,
  baseline_type = "shuffle",
  target_class_idx = 3,
  model = model,
  num_baseline_repeats = 50)

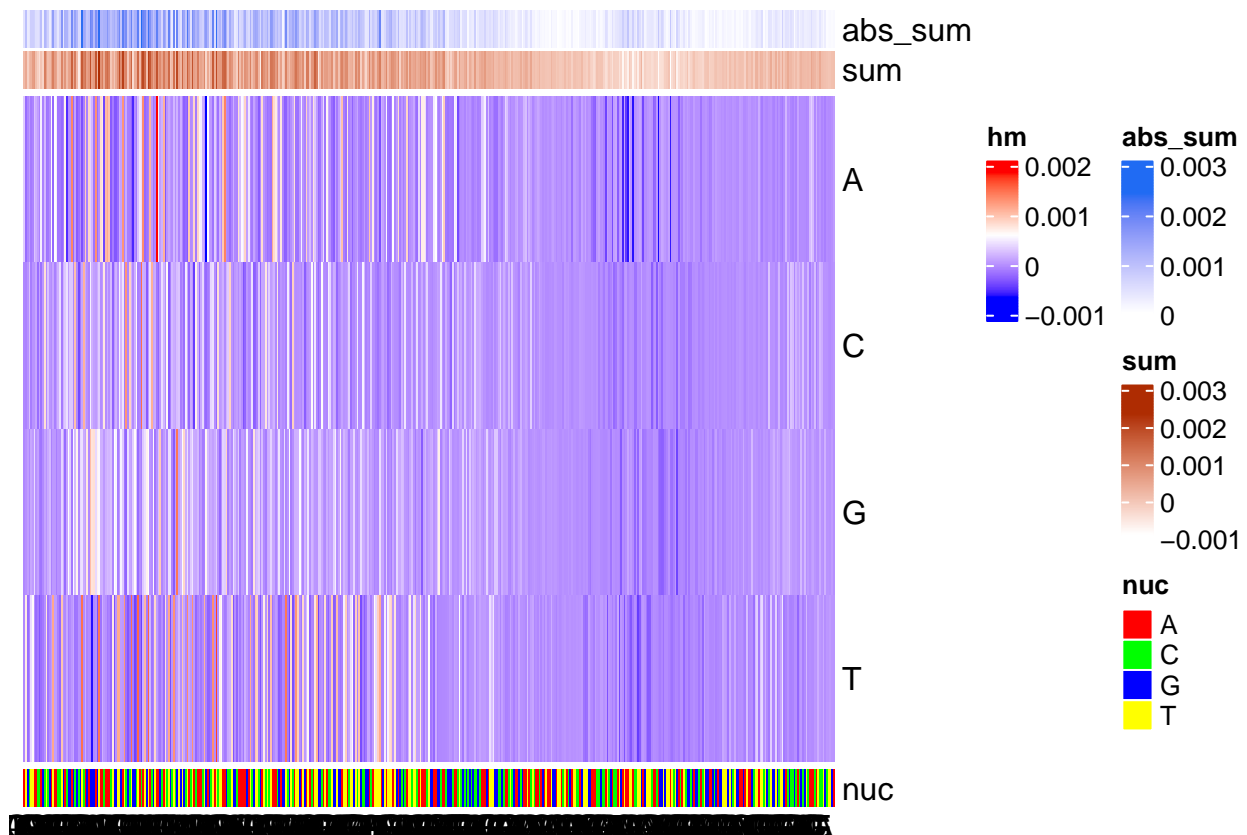
heatmaps_integrated_grad(integrated_grads = ig,
  input_seq = onehot_instance)

```

```

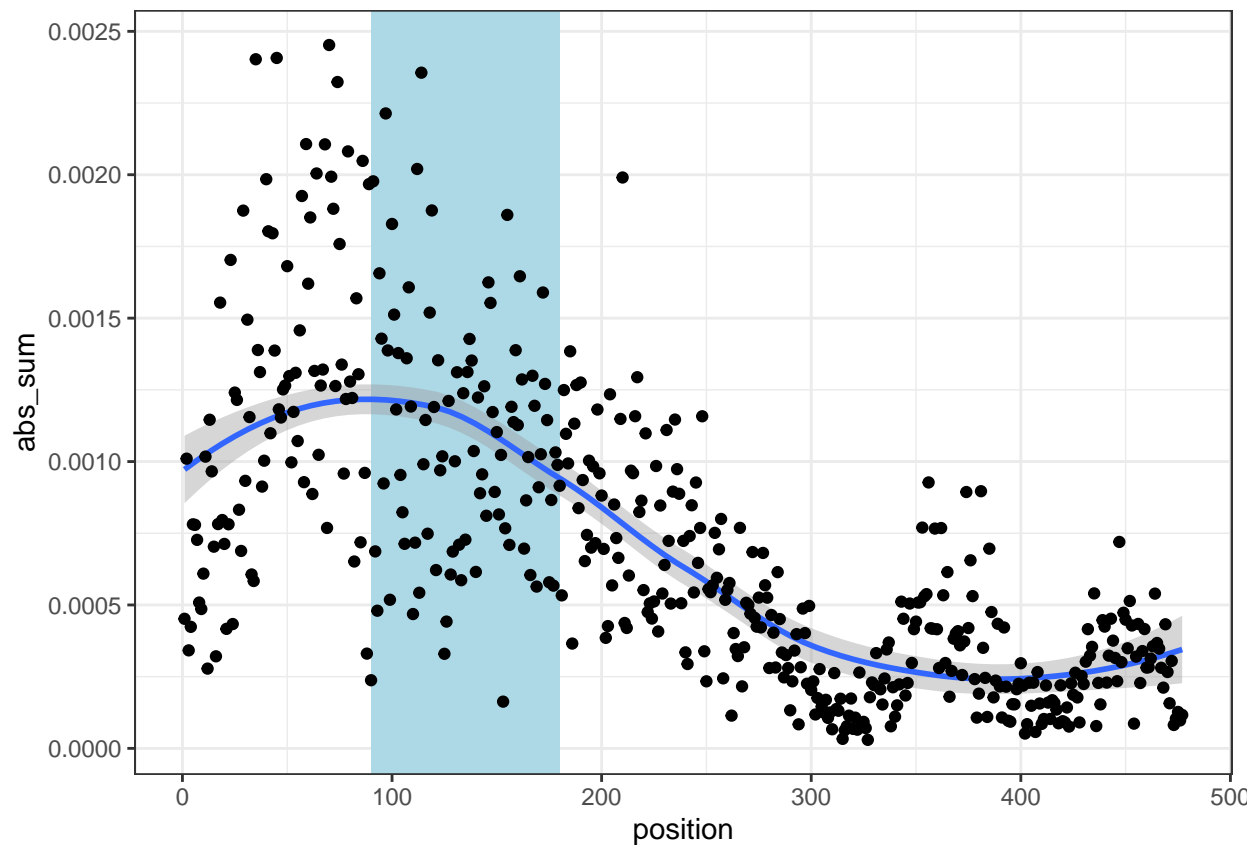
## [[1]]

```



```
abs_sum <- rowSums(abs(as.array(ig)))
df <- data.frame(abs_sum = abs_sum, position = 1 : 477)
ggplot(df, aes(x = position, y = abs_sum)) + geom_rect(aes(xmin = 90, xmax = 180, ymin = -Inf, ymax = Inf))

## 'geom_smooth()' using method = 'loess' and formula = 'y ~ x'
```



```
# abs_sum <- rowSums(abs(as.array(ig)))
#df <- data.frame(abs_sum = abs_sum, position = 1 : 477)
#ggplot(df, aes(x = position, y = abs_sum)) + geom_rect(aes(xmin = 90, xmax = 180, ymin = -Inf, ymax = Inf))

# modification: df should be summed every 3 rows, position reassigned accordingly.
df_mod <- df %>%
  mutate(group = rep(1:(nrow(df) / 3), each = 3)) %>%
  group_by(group) %>%
  summarise(
    abs_sum_sum = sum(abs_sum),
    abs_sum_median = median(abs_sum),
    abs_sum_mean = mean(abs_sum),
    abs_sum_max = max(abs_sum)
  )

# Split special_seq (a string) into triplets (every third character)
chars <- strsplit(special_seq, "")[[1]]

# Create triplets
trip_spec <- sapply(seq(1, length(chars), by = 3), function(i) {
  paste(chars[i:min(i+2, length(chars))], collapse = "")
})

# Get the index of sequences at which trip_spec != triplets
index <- which(trip_spec != triplets)
```

```

# Plot the results with different smooth lines for sum, median, mean, and max
ggplot(df_mod, aes(x = group)) +
  geom_rect(aes(xmin = 30, xmax = 60, ymin = -Inf, ymax = Inf), fill = "lightblue", alpha = 0.2) +
  # draw points of max
  geom_point(aes(x = group, y = abs_sum_max), color = "purple") +
  geom_smooth(aes(y = abs_sum_sum, color = "Sum"), method = "auto", se = FALSE) +
  geom_smooth(aes(y = abs_sum_median, color = "Median"), method = "auto", se = FALSE) +
  geom_smooth(aes(y = abs_sum_mean, color = "Mean"), method = "auto", se = FALSE) +
  geom_smooth(aes(y = abs_sum_max, color = "Max"), method = "auto", se = FALSE) +
  scale_color_manual(values = c("Sum" = "blue", "Median" = "green", "Mean" = "orange", "Max" = "purple")) +
  geom_point(data = data.frame(group = index), aes(x = group, y = 0), shape = 4, size = 3, color = "red") +
  theme_bw() +
  labs(y = "Absolute Sum of IG", color = "Statistic", x = "AA Index")

```

```

## 'geom_smooth()' using method = 'loess' and formula = 'y ~ x'
## 'geom_smooth()' using method = 'loess' and formula = 'y ~ x'
## 'geom_smooth()' using method = 'loess' and formula = 'y ~ x'
## 'geom_smooth()' using method = 'loess' and formula = 'y ~ x'

```

