Plan for Data Analysis

ML Team

Data

- The key point of our project should be the data we generate through the game and how we analyze that data to improve our game and provide users with meaningful achievements.
- There could be two different groups of data: user input data and user-generated data.
  - User input data: user's biography users type in when they sign up
  - User generated data: data generated by users during the game
- There will three analyzed outcomes.
  - 1) User input data
  - 2) User generated data
  - 3) User input data + User generated data
- The types of data we generate and our methods of analysis should be the focal points of our poster and other documents.

Data Type: Input Data

- There could be many types of input data. We should consider our goal carefully. The data should be useful for our analytics. Let me describe data we can ask the users:

1) Height 2) Weight 3) Age 4) Gender 5) Exercise Experience

Please suggest any other data we can ask the users.

Data Type: Generated Data

- There could be many types of generated data. We should consider our goal carefully. The data should be useful for our analytics. Let me describe data we can generate (although it may not be feasible):

1) Play Time, 2) Level of Stage 3) Number of Plays 4) Change in the angle of bent arm 5) Number of times arm is bent .. etec

Please suggest any other data we can generate.

Analyze: Example (Input + Generated)

- Let me first describe the example of the saved data

  - Pose: arm
  - Age: [25 28 29 19 20 18 17 29 29 19 45 48 49 39 40 38 37 49 49 39]
  - play time: [20 30 40 50 10 30 40 100 2 400 587 123 896 309 701 45 890 234 652 12]
  - stage level: [4 5 2 3 5 2 3 4 4 9 5 7 2 9 3 1 6 8 4 2]

Analyze: Example (Input + Generated)

- First, let’s divide into two age groups
  - Age 10 ~ 30
  - Age 31 ~ 50

- Result for Age Group 10~30
  - Pose: arm
  - Age: [25 28 29 19 20 18 17 29 29 19]
  - play time: [20 30 40 50 10 30 40 100 2]
  - stage level: [4 5 2 3 5 2 3 4 4 9]

- Result for Age Group 31~50: skip for now

Analyze: Example: Age Group 10~30

- We want to calculate score pose by pose
- To achieve this, we will remove outliers(0~10%, 90~100%), and normalize the data to analysis
- Equation:
  - j: User's name
  - m: Number of plays by the j
  - N: total number of plays
  - k: 10% of N, indicating outliers

$$\text{Norm}(Arm)_j = \frac{\sum_{i=1}^{m} \text{PlayTime}_i \times \text{StageLevel}_i}{\frac{\sum_{i=k+1}^{n-k} \text{PlayTime}_i \times \text{StageLevel}_i}{n-2k}} \times \frac{1}{m}$$

[FYI]
Pose: arm
Age: [25 28 29 19 20 18 17 29 29 19]
play time: [20 30 40 50 10 30 40 100 2]
stage level: [4 5 2 3 5 2 3 4 4 9]