

Virtual Patient Question Classification Robustness Analysis

Han Lie

December 2018

Introduction

The Question Classification Robustness analysis focusses on giving insight in to what degree questions were correctly classified by the system compared to hand annotated by a coder.

Question data consisted of all of the questions that were asked by users and to what question subject they were classified by the system and how they were classified by a human coder.

The comparison consisted of the percentage of questions that were classified the same by the system and coder, the percentage of questions that were misclassified by the system compared to a coder and the percentage of questions that were misclassified by the system compared to a coder and were thought to be unclassifiable by the system according to a coder.

Read in the question data:

```
questionData <- read.csv(file=paste("/Users/fooyonghan/Downloads/final-analysis/",
"question-analysis/question-analysis.csv", sep=""), header=TRUE, sep=";")
```

Create a list that contains per user and session the percentage of questions that:

- Matched between the system and human coder (matchPerc).
- Did not match between the system and human coder (missPerc).
- Did not match between the system and human coder and was thought to be unclassifiable by the system according to the human coder (unknownPerc).

```
finalList <- data.frame(userSession=character(), matchCount=integer(),
  unknownCount=integer(), missCount=integer())
names(finalList)<-c("userSession", "matchCount", "unknownCount", "missCount")
```

```
splitQData <- split(questionData, questionData$userSession)
```

```
totalMatch = 0
totalMiss = 0
totalUnknown = 0
```

```
for(i in splitQData){
  matchCount = 0
  unknownCount = 0
  missCount = 0

  typed <- i[4]
  interpreted <- i[6]

  for(j in 1:nrow(typed)){
    if(as.character(typed$questionSubject..System.[j]) ==
      as.character(interpreted$questionSubject..Human.[j])){
      matchCount = matchCount + 1
    }
  }
}
```

```

    }else if(
      (as.character(interpreted$questionSubject..Human.[j]) == "unknown") &&
      (as.character(typed$questionSubject..System.[j]) !=
       as.character(interpreted$questionSubject..Human.[j]))){
        unknownCount = unknownCount + 1
        missCount = missCount + 1
      }else {
        missCount = missCount + 1
      }
    }

    matchPerc <- matchCount/nrow(typed)*100
    missPerc <- missCount/nrow(typed)*100
    unknownPerc <- unknownCount/nrow(typed)*100

    rowFinal <- data.frame(i[1,1], matchPerc, missPerc, unknownPerc)
    names(rowFinal) <-c("userSession", "matchPerc", "missPerc", "unknownPerc")

    finalList <- rbind(finalList, rowFinal)

    totalMatch = totalMatch + matchCount
    totalMiss = totalMiss + missCount
    totalUnknown = totalUnknown + unknownCount
  }

```

Percentage of the missed questions that the system should not be able to answer (according to a coder), as they were not implemented:

```
totalUnknown/totalMiss*100
```

```
## [1] 80.64516
```

Calculate the overall mean and standard deviations of match, mismatch and unclassifiable mismatches:

```
mean(finalList$matchPerc)
```

```
## [1] 71.89604
```

```
mean(finalList$missPerc)
```

```
## [1] 28.10396
```

```
mean(finalList$unknownPerc)
```

```
## [1] 22.8388
```

```
sd(finalList$matchPerc)
```

```
## [1] 20.13263
```

```
sd(finalList$missPerc)
```

```
## [1] 20.13263
```

```
sd(finalList$unknownPerc)
```

```
## [1] 18.9715
```