University of Glasgow | School of Computing Science

Assessed Coursework

Course Name	**Computational Social Intelligence**			
Coursework Number	**1 (first part)**			
Deadline	**Time:**	**17.00**	**Date:**	**Nov. 26th, 2018**
% Contribution to final course mark	**20%**			
Solo or Group ✓	**Solo**	**X**	**Group**	
Anticipated Hours	**10 hours**			
Submission Instructions	**The coursework must be submitted via the Moodle page.**			
Please Note: This Coursework cannot be Re-Assessed				

Code of Assessment Rules for Coursework Submission

Deadlines for the submission of coursework which is to be formally assessed will be published in course documentation, and work which is submitted later than the deadline will be subject to penalty as set out below.

The primary grade and secondary band awarded for coursework which is submitted after the published deadline will be calculated as follows:

(i) in respect of work submitted not more than five working days after the deadline
   a. the work will be assessed in the usual way;
   b. the primary grade and secondary band so determined will then be reduced by two secondary bands for each working day (or part of a working day) the work was submitted late.

(ii) work submitted more than five working days after the deadline will be awarded Grade H.

Penalties for late submission of coursework will not be imposed if good cause is established for the late submission. You should submit documents supporting good cause via MyCampus.

Penalty for non-adherence to Submission Instructions is 2 bands

You must complete an "Own Work" form via https://studentltc.dcs.gla.ac.uk/ for all coursework

1.Introduction

The first part of the Assessed Exercise revolves around the study presented in Lecture 05 ("Observing Behaviour"):

- A.Vinciarelli, P.Chatziioannou & A.Esposito, "When the Words are not Everything: The Use of Laughter, Fillers, Back-Channel, Silence and Overlapping Speech in Phone Calls", Frontiers in ICT, 2:4, 2015.

The article above is available on Moodle (see Topic "Lecture 05" on the Moodle page of the course) and it provides a general description of data and research issues.

This is the first part of the Assessed Exercise, the second part will be handed out on November 8th, 2018. The marking will be done over both parts.

The deadline for submission is November 26th, 2018.

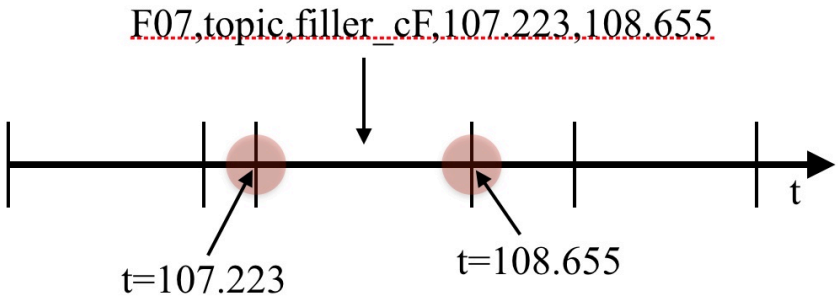

Figure 1 The figure shows a segmentation. Every segment is a time interval in which an event (filler, laughter, etc.) takes place. The record includes the call in which the segment can be found (C07 in the figure), the topic (it can be neglected), the event (filler_cF in the example), the start time of the segment (107.223 in the example), the end time of the segment (108.655 in the example).

2.The Data

The data (file "corpus.csv") is the manual segmentation of 60 phone calls between unacquainted individuals (120 in total). The segmentation is the task of splitting a recording into segments (time intervals) characterized by a property of interest. In the case of the data under exam, every segment corresponds to one action that a person can perform during a conversation (see Figure 1).

In total, the data file contains 29,859 records. Each of them corresponds to a segment and it includes the following fields separated by a comma (see Figure 1):

1. Code of the call: "Fnn", where nn is an integer number ranging between 1 and 60 (for the numbers lower than 10 there is a leading "0");
2. Topic: this field can be neglected in this Assessed Exercise;
3. Event: "XX_yZ", where
    a. XX corresponds to "laughter", "filler" or "bc";
    b. y is the role code and corresponds to "r" (receiver) or "c" (caller);
    c. Z is the gender code and corresponds to "M" (male) or "F" (female);

    In a large fraction of the records, the field Event does not fit the format above, such records must be neglected.
4. Start time (in milliseconds);
5. End time (in milliseconds).

Two examples of record are as follows:

- F01,pistol,filler_cF,327.666, 328.213
  The segment comes from call "01", corresponds to a filler uttered by a female subject (the gender code is "F"), starts at second 327.666 and ends at second 328.213;

- F37,other,laughter_rM,221.639,222.666
  The segment comes from call "37", corresponds to a laughter event displayed by a male subject (the gender code is "M"), starts at second 221.639 and ends at second 222.666.

Please consider that, in some cases, the field Event includes multiple items separated by a space:

- F01,pistol,filler_cF rM,354.735,355.011
  The Event field includes both filler_cF (to be included in the analysis) and rM (to be discarded in the analysis).

3.The Assessed Exercise

You have to address the following research questions:

- Do female subjects laugh more frequently than male ones?
- Do female subjects use fillers more frequently than male ones?
- Is there a laughter length difference between male and female subjects?
- Is there a filler length difference between male and female subjects?

For each of the above questions you are expected to perform the following tasks:

- Formulate a clear research hypothesis and, correspondingly, a suitable null hypothesis;
- Select an appropriate statistical test among those that have been presented during the course;
- Explain whether you reject the null hypothesis and, if yes, what is the confidence level. Furthermore, explain whether you use a two-tailed or a one-tailed test and why.

You are free to use the programming language you prefer and the use of statistical libraries that perform statistical tests is allowed. However, you have to show that you know how to calculate the statistic you use for the test (chi Square or Student's t).

4.The report

The report must include a general description of the problem and, for each of the research questions in Section 3, the following elements:

- Formulation of research hypothesis and corresponding null hypothesis;
- Description of the data, e.g., number of laughter occurrences for female and male subjects, average and variance of laughter length for female and male subjects, etc.;
- Description of the approach, e.g., which test? What are the parameters (degrees of freedom, expectations, etc.) and what is the motivation behind every choice (why a certain test? Why a certain value of the parameters?);
- Results of the hypothesis testing (can the null hypothesis be rejected? What is the confidence level? Etc.);
- Explanation of the results (e.g., in case the test shows that the number of laughter events is higher for female subjects to a statistically significant extent, you should mention that your conclusion is that female subjects tend to laugh more than male ones);
- Analysis software you have written (the code must be added in appendix);

The report should not include more than one page per each of the questions to be addressed (see Section 3).

5.Marking Scheme

The Assessed Exercise (including both first and second part) accounts for 20% of the final mark. The weight across the different components is as follows:

- Analysis of the data: 60%;
- Editorial quality: 20%;
- Quality of the code: 20%.

The two parts of the Assessed Exercise have the same weight (each accounts for 10% of the final mark).