University College Dublin
An Coláiste Ollscoile, Baile Átha Cliath

AUTUMN TRIMESTER EXAMINATIONS

ACADEMIC YEAR 2019/2020

COMP47600

Text Analytics

Prof. Simon Thompson

Dr. Chris Bleakley

Prof. M. Keane *

Time Allowed: 2 Hours

Instructions for Candidates

Answer any FOUR questions.
All questions carry equal marks.
Total marks available 100.
Use of calculators is prohibited

Student Number

Seat Number

Instructions for Invigilators
Use of calculators is prohibited

1. There are three main approaches to "similarity" that group the techniques used in text analytics. Describe each approach and illustrate each with an example, identifying when it is appropriate to use one approach rather than another, and any issues that can arise in their use [8.33 marks for each part].

[25 marks]

2. In machine learning, a fundamental distinction is often made between supervised and unsupervised methods. Describe the main differences between these two broad classes of methods [5 marks].

Then, give detailed accounts of one example of each class (i.e., provide two specific technique descriptions, one that is supervised and one that is unsupervised) [10 marks].

Finally, illustrate each of these techniques with an example from the text analytics literature [10 marks].

[25 marks]

3. Describe three ways in which Log-Likelihood Ratios (LLRs) have been used to find patterns of significant words in comparisons between different texts or text-corpora [15 marks]. Provide a numeric example showing how LLRs are computed [5 marks]. Describe one study from the text analytics literature that used LLRs and critically evaluate this study [5 marks].

[25 marks]

4. Several different approaches have been used to find temporal regularities in text data. Describe three of the approaches that have been used, illustrating each with an example from the literature, with a critical evaluation of each [8.33 marks for each part].

[25 marks]

5. Text analytics typically begins with the pre-processing of each text-item, in some selected corpus, to prepare it for subsequent processing. Describe five of the typical pre-processing steps that are carried out during this pre-processing stage and show how each step modifies a given text fragment. In describing each pre-processing step, explain why it is used and the benefits that follow from its use [5 marks for each part].

[25 marks]